- **Protocol number: 331-12-283**
- **Document title: A Phase 3, 12-week, Multicenter, Randomized, Double-blind, Placebo-controlled Trial to Evaluate the Efficacy, Safety, and Tolerability of 2 Fixed Doses of Brexpiprazole (OPC-34712) in the Treatment of Subjects with Agitation Associated with Dementia of the Alzheimer's Type**
- **Version number: 3.0**
- **Date of the document: 14 March 2017**
- **NCT number: NCT01862640**

Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational Medicinal Product
Brexpiprazole (OPC-34712)

A Phase 3, 12-week, Multicenter, Randomized, Double-blind, Placebo-controlled Trial to Evaluate the Efficacy, Safety, and Tolerability of 2 Fixed Doses of Brexpiprazole (OPC-34712) in the Treatment of Subjects with Agitation Associated with Dementia of the Alzheimer's Type

Protocol No. 331-12-283
IND No. 115,960
EudraCT No. 2013-000504-41

Statistical Analysis Plan

Version: Final

Date: 14 March, 2017

CONFIDENTIAL - PROPRIETARY INFORMATION

Table of Contents

1 Introduction

This statistical analysis plan (SAP) documents the statistical methodology and data analysis algorithms and conventions to be applied for statistical analysis and reporting of efficacy and safety data of study 331-12-283. All amendments to the protocol and Addendum to protocol amendment are taken into consideration in developing this SAP.

2 Trial Objectives

Primary: To compare the efficacy of 2 fixed doses (1 mg/day and 2 mg/day) of brexpiprazole with placebo in subjects with agitation associated with dementia of the Alzheimer's type, as assessed by the CMAI after 12 weeks of treatment.

Secondary: To evaluate the safety and tolerability of 2 fixed doses of brexpiprazole (1 mg/day and 2 mg/day) compared with placebo in subjects with agitation associated with dementia of the Alzheimer's type after 12 weeks of treatment.

3 Trial Design

This is a phase 3, 12-week, multicenter, randomized, double-blind, placebo-controlled, 3-arm, fixed-dose trial designed to assess the efficacy, safety, and tolerability of brexpiprazole (1 mg/day and 2 mg/day) in the treatment of subjects with agitation associated with dementia of the Alzheimer's type. The trial population will include male and female subjects between 55 and 90 years of age (inclusive), who are living in either an institutionalized setting or in a non-institutionalized setting where the subject is not living alone. In both the institutionalized and non-institutionalized settings, the subject must have a caregiver who can spend a minimum of 2 hours per day for 4 days per week with the subject in order to assess changes in the subject's condition. All subjects must have a diagnosis of probable Alzheimer's disease according to the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria.

The trial comprises a 2- to 42-day screening period, a 12-week double-blind treatment period, and a 30-day post-treatment follow-up period. In addition, for all subjects who terminate early from the trial, all attempts will be made to collect mortality data by telephone contact with the subject's caregiver at Week 16.

This trial will be monitored under the supervision of an independent Data Monitoring Committee (DMC). The DMC will monitor safety periodically, based on a predetermined

schedule. The details of the DMC structure and its roles and responsibilities will be documented in a DMC charter.

The trial is organized as follows:

Screening Period

The screening period will range from 2 days to 42 days and will begin when the informed consent form (ICF) is signed, prior to the initiation of any procedures. The screening period may be extended after discussion with and approval by the medical monitor. An interactive voice response system (IVRS) or interactive web response system (IWRS) will be used to obtain the subject study identification number for each subject with a signed ICF.

12-week, Double-blind Treatment Period

Based on a randomization scheme, eligible CCI [redacted] subjects will be allocated in a 1:1:1 ratio at randomization to 1 of the following 3 treatment groups:

- Brexpiprazole 1 mg/day
- Brexpiprazole 2 mg/day
- Placebo

CCI [redacted]

Subjects will follow a titration schedule, depending upon their assigned treatment group, to gradually increase their dose to their assigned target dose.

Subjects unable to tolerate their assigned dose of brexpiprazole (or matching placebo) will be withdrawn from the trial. Down-titration is not allowed at any time during the study. If a subject is withdrawn, every effort will be made to complete all of the Week 12/Early Termination (ET) evaluations prior to administering any additional medications for the treatment of agitation or other prohibited medications.

Subjects will be evaluated at Baseline, Day 3, and Weeks 2, 4, 6, 8, 10, and 12 during the double-blind treatment period.

All attempts should be made to maintain the subjects' normal routine with regard to physician appointments. Individual circumstances that fall outside this general convention should be discussed with the medical monitor in order to determine appropriateness to

proceed. In addition, the subject's identified caregiver will be contacted by telephone at Weeks 3, 5, and 7 to assess compliance with IMP, confirm any changes to concomitant medications, and assure the subject's well-being. Trial-related efficacy and safety assessments will be performed as outlined in the study protocol.

Follow-up Period

All subjects, whether they complete the trial or are withdrawn prematurely for any reason, will be followed up for a safety evaluation 30 (+ 2) days after the last dose of the IMP during a clinic visit at either the investigator's site or residential facility, if applicable. If a clinic visit is not possible, the subject should be assessed by telephone contact with the subject and a caregiver. For all subjects who terminate early from the trial, all attempts will be made to collect mortality data by telephone contact with the subject's caregiver at Week 16.

Subjects who complete both the 12-week double-blind treatment period and the 30-day safety follow-up visit are eligible to enroll into Trial 331-13-211, which is a 2-month, observational, rollover trial to evaluate the safety of subjects with agitation associated with Alzheimer's disease who previously participated in Trial 331-12-283. For those subjects who plan to enroll into Trial 331-13-211, the 30-day safety follow-up visit for Trial 331-12-283 will occur as a clinic visit at either the investigator's site or residential facility, if applicable.

4 Sample Size and Power Justification

The sample size was calculated based on the treatment effect of 6.5 points with a standard deviation (SD) of 16.5 in the change from baseline (Day 0 visit) to the end of the double-blind treatment period (Week 12 visit) in the CMAI total score, to achieve 85% power at a 2-sided alpha level of 0.05. This results in 117 subjects in each of the groups (i.e., brexpiprazole 1 mg/day, brexpiprazole 2 mg/day, and placebo). After allowance of 10% non-evaluable subjects, the total number of subjects to be randomized is 132 per treatment arm. The sample size was estimated based on a 1:1:1 randomization ratio (brexpiprazole 1 mg/day, brexpiprazole 2 mg/day, placebo).

CCI [redacted] the total number of subjects to be randomized will be approximately 420 CCI [redacted].

5 Data Sets for Analysis and Missing Data

5.1 Data Sets for analysis

The following analysis samples are defined for this trial:

Randomized Sample: consists of all subjects who were randomized into this trial. Subjects are considered randomized when they are assigned a treatment number by IVRS at the end of Screening Period. A subject receiving trial treatment outside of the IVRS will not be considered randomized, but safety will be reported.

Safety Sample: consists of all subjects who were administered at least one dose of IMP. Subjects will be excluded from this population only if there is documented evidence (ie, drug dispensed = drug returned or no trial drug dispensed) that the subject did not take trial drug. If a subject is dispensed trial medication and is lost to follow up, he/she will be considered exposed.

Efficacy Sample: The intent-to-treat (ITT) population consists of all subjects in the randomized sample, who took at least 1 dose of IMP (excluding subjects randomized to brexpiprazole 0.5 mg/day) and have a baseline and at least one postbaseline evaluation for the CMAI total score.

The core dataset for all efficacy analyses is based on the modified ITT population, which is defined in the efficacy sample above. However, as described below, in order to handle missing data and restrictions imposed by different types of analyses (eg, change from baseline analysis), other datasets derived from the ITT population will be used for the efficacy analyses.

5.2 Handling of Missing Data

The CMAI is utilized as the primary efficacy assessment of a subject's level of agitated behaviors. The CMAI consists of 29 items all rated on a 1 to 7 scale with 1 being the "best" rating and 7 being the "worst" rating. The CMAI Total Score is the sum of ratings for all 29 items. The possible total scores are from 29 to 203. The CMAI Total Score will be unevaluable if less than 24 of the 29 items are recorded. If 24 to 28 of the 29 items are recorded, the total score will be the mean of the recorded items multiplied by 29 and then rounded to the first decimal place.

CCI [redacted]

The primary efficacy variable is the change from baseline to Week 12 in the CMAI total score.

In general, missing data will be handled by analysis of Mixed effect Model Repeat Measurement (MMRM) methodology based on observed-case (OC) data from protocol-specified visits in the ITT population under the assumption of missing at random (MAR).

The OC dataset consists of actual observations recorded at each visit during the double-blind acute treatment phase, and no missing data will be imputed.

CCI [redacted]

6 Study Conduct

6.1 Subject Disposition and Reasons for Discontinuations

Subject disposition is summarized for the randomized sample. Disposition is summarized by treatment group and by subgroup of gender, age (<65; >=65 and <75; or >=75), race, and region (North America or Other).

Reasons for discontinuation will be summarized for the randomized sample by treatment group and by subgroup of gender, age, race, and region.

6.2 Treatment Compliance

Based on the Study Medication panel of the CRF, compliance in taking study medication is calculated by dividing the number of tablets taken by the total number of tablets the subjects were scheduled to take during the study period. For lost-to-follow up subjects, last study medication end date record will be used as the treatment end date.

6.3 Protocol Deviation

Protocol deviations are summarized by center and type of deviation for randomized subjects by treatment group. A listing of protocol deviations will also be generated.

7 Baseline Characteristics

7.1 Baseline Definition

For analyses of the double blind treatment period data, the baseline is the Baseline measurement (expected to be at Day 0). Baseline measurement is defined as the last available measurement prior to the start of double-blind study medication.

7.2 Demographic Characteristics

Baseline demographic characteristics including age, gender, race, ethnicity, height, weight, waist circumference, and body mass index (BMI) will be tabulated by treatment group for all randomized subjects. Additional summaries by the following subgroups will be also generated: by gender, by age group (<65; >=65 and <75; or >=75), by race, and by region (North America or Other).

Mean, range and standard deviation will be used to describe continuous variables such as age. Frequency distributions will be tabulated for categorical variables such as race.

7.3 Disease History

A summary of medical, psychiatric, neurological (excluding Alzheimer's), and Alzheimer's disease history will be presented for the Randomized Sample (by treatment group and overall).

7.4 Baseline Disease Characteristics

For the Randomized Sample, baseline and baseline disease characteristics will be summarized by treatment group and overall. The following baseline characteristics will be summarized at baseline: number (%) of institutionalized / non institutionalized subjects; CMAI total score; CCI [redacted] CGI-S score, NPI total score; NPI-AA item score; MMSE score; CCI [redacted] CCI [redacted]

Number of subjects with presence of psychotic symptoms will be summarized at baseline using NPI Delusion and Hallucination score. The counts will be provided for the following categories: NPI item score >=4, >=5, >=6 on either hallucination or delusion score.

All patients randomized under original protocol, amendment 1, and amendment 2 is assumed to be institutionalized and will be included in institutionalized subgroup.

8 Efficacy Analysis

All efficacy analyses pertaining to the double blind treatment period will be performed on the Efficacy Sample, and subjects will be included in the treatment group as randomized. CCI [redacted]

Statistical comparisons are based on 2-sided, 0.05 significance levels.

NPI endpoints are collected using NPI-NH for institutionalized subjects and NPI/NPI-NH for non-institutionalized subjects. Since each score derivation of the NPI items is based on the frequency and severity scores only, no distinction will be made due to the instrument used. All analysis of NPI based endpoints will be based on totality of data, unless specified otherwise.

8.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from the baseline (Day 0 visit) to the end of the double blind treatment period (Week 12 visit) in CMAI Total Score.

8.1.1 Primary Efficacy Analysis

The primary endpoint will be analyzed using a mixed-effect model repeated measure (MMRM) model. The primary efficacy outcome measure is the mean change from baseline (Day 0 visit) to the end of the double-blind treatment period (Week 12 visit) in the CMAI total score. The primary statistical comparisons of interest are brexpiprazole 2 mg/day versus placebo and brexpiprazole 1 mg/day versus placebo. The null hypotheses of these comparisons are that there is no difference between the brexpiprazole treatment groups and placebo in change from baseline to endpoint in CMAI total score. To protect the experiment-wise 2-sided alpha level at 0.05 when making 2 comparisons of brexpiprazole doses versus placebo, the statistical testing will be carried out using a hierarchical testing procedure in the order of 1) comparison of 2 mg/day brexpiprazole versus placebo and 2) comparison of 1 mg/day brexpiprazole versus placebo. Thus, if the test yields a statistically significant result at 0.05 (2-sided) for the comparison of 2 mg/day brexpiprazole versus placebo, then the comparison of 1 mg/day brexpiprazole versus placebo will be tested at an alpha level of 0.05 (2-sided).

The primary analysis will be performed on the Efficacy Sample which includes all randomized subjects who took at least one dose of IMP in the double blind treatment period and who have both a baseline and at least one post-randomization CMAI Total Score during the double blind treatment period. The primary efficacy analysis will be performed by fitting a MMRM analysis with an unstructured (UN) variance covariance structure in which the change from the baseline in CMAI Total Score (Week 2, 4, 6, 8, 10, 12) will be the dependent variable based on the OC data set. The model will include fixed class-effect terms for treatment (1 mg/day, 2 mg/day, and placebo), trial center, visit week, and an interaction term of treatment by visit week and include the interaction term of baseline values of CMAI Total Score by visit week as covariates. The primary comparison between a brexpiprazole and the placebo arm at Week 12 will be estimated as the difference between Least Squares (LS) means from the interaction term of treatment by visit week utilizing the computing software SAS procedure PROC MIXED.

CCI [redacted]

CCI

[redacted]

CCI

[redacted]

8.1.2 Technical Computation Details for Primary Efficacy Analysis

CCI

[redacted]

CCI

CCI

8.1.3 Sensitivity Analyses for Primary Efficacy Endpoint

CCI

CCI

3) CCI

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

8.1.4 Subgroup Analyses for Primary Efficacy Endpoint

Subgroup analyses of change from baseline in CMAI Total Score to every study week in the double blind treatment period will be performed by the following factors:

- Gender
- Race (White and All Other Races)

- Age group (<65; >=65 and <75; or >=75)
- Region (North America and Other)

CCI [redacted]

8.2 Key Secondary Analysis

The key secondary efficacy variable is the change from baseline to endpoint in the CGI-S score, as related to agitation. It will be analyzed by the same statistical methodology specified for the analysis of the primary efficacy variable, based on the Efficacy Sample. In order to control the overall type I error rate for this key secondary efficacy analysis, a hierarchical testing procedure will be used so that the overall experiment-wise type I error rate is maintained at 0.05. Thus, if the primary efficacy analysis for the CMAI total score yields a statistically significant result at 0.05 (2-sided) for both of the comparisons of brexpiprazole 1 mg/day and 2 mg/day versus placebo, then the corresponding comparison for this key secondary efficacy variable (CGI-S score) will be tested at an alpha level of 0.05 (2-sided) using another hierarchical testing procedure in the order of brexpiprazole 2 mg/day versus placebo and brexpiprazole 1 mg/day versus placebo. Thus, brexpiprazole 1 mg/day versus placebo will be tested only if brexpiprazole 2 mg/day versus placebo reaches significance at 0.05 (2-sided) for this key secondary efficacy variable.

8.3 CCI [redacted]

[redacted]

- CCI [redacted] CCI [redacted]

[redacted]

8.4 CCI [redacted]

[redacted]

- CCI

- CCI

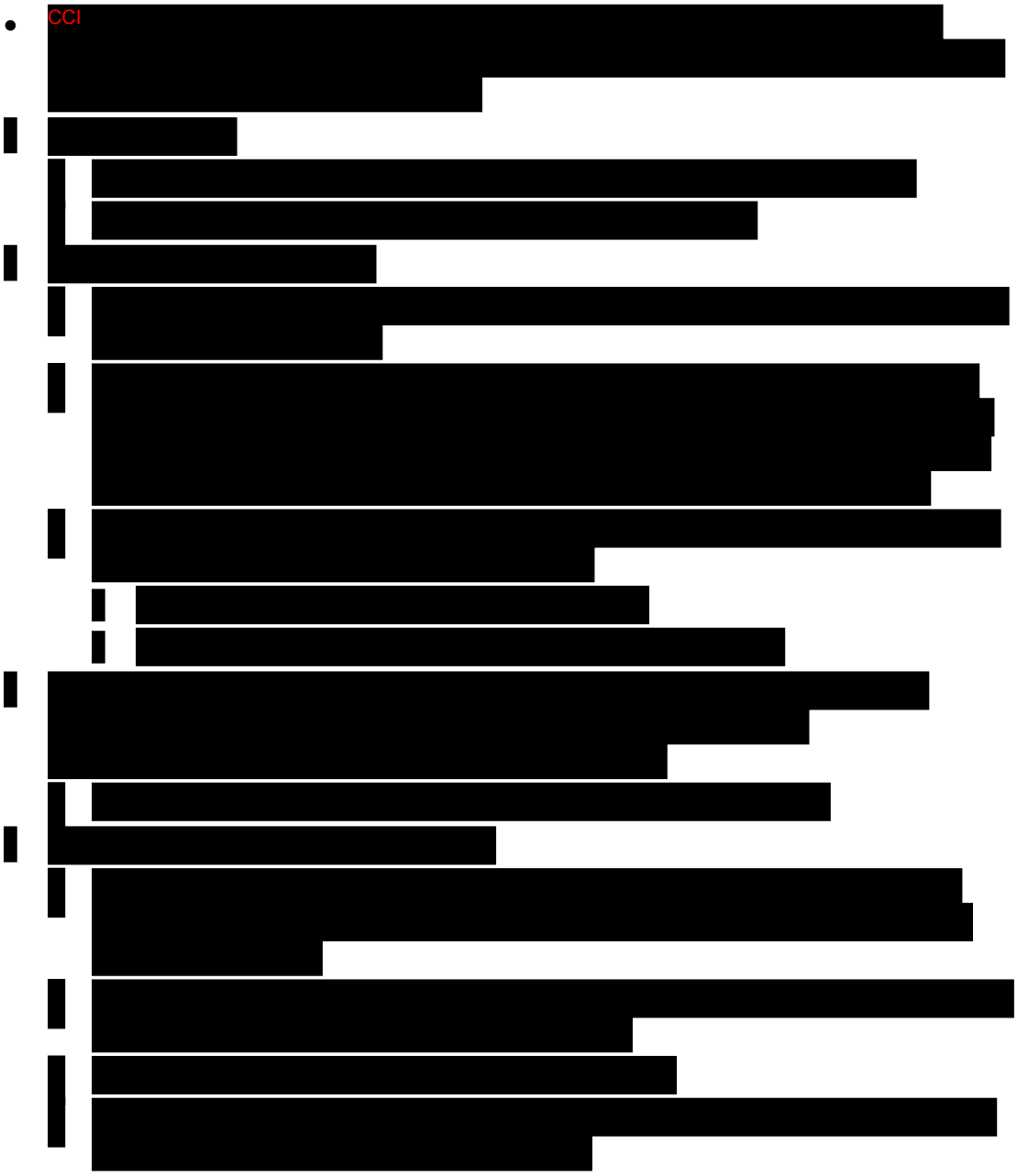
- CCI [redacted]

Summary statistics will be provided by treatment groups. Statistical comparisons will be evaluated using the same MMRM model described in the primary analysis at a nominal 0.05 level (two-sided).

8.5 CCI [redacted]

[redacted]

- CCI



8.6 CCI

8.7 Actigraphy and eDiary Data

Since actigraphy and eDiary are tools to assist the CST in monitoring CMAI rater training, actigraphy and eDiary information will not be made available to site personnel, will not be included in the clinical database, and will not be statistically analyzed.

9 Safety Analysis

Standard safety variables to be analyzed include AEs, clinical laboratory tests, vital signs, ECGs, body weight, waist circumference and physical examination. In addition, data from the following safety scales will be evaluated: MMSE score, CCI [redacted] Prospectively defined criteria will be used to identify potentially clinically relevant abnormal values for clinical laboratory tests, vital signs, ECGs, and body weight.

Safety analysis will be conducted based on the Safety Sample. In general, baseline of a safety variable is defined as the last observation of the variable before taking the first dose of IMP, unless specified otherwise.

9.1 Adverse Events

All adverse events will be coded by primary system organ class (SOC) and Medical Dictionary for Regulatory Activities (MedDRA version 19.0) preferred term (PT). AEs that are gender-specific, e.g., ovarian cancer, will have their incidence rates evaluated for the specific gender. The incidence of the following events will be summarized by treatment group:

a) Treatment-emergent AEs (TEAE) by SOC and PT
b) TEAEs at least 5% in the brexpiprazole group and twice greater than placebo
c) TEAEs at least 2% in the brexpiprazole group and greater than placebo
d) TEAEs by severity
e) Potentially drug-related TEAEs
f) TEAEs with an outcome of death
g) Serious TEAEs
h) TEAEs leading to discontinuations
i) EPS-related AEs
j) EPS-related AEs by time

k) AEs of special interest
l) TEAEs by gender, race, age and region

EPS-related AEs will be grouped into five categories.

1) Dystonic Events, which include cervical spasm, dystonia, emprosthotonos, muscle contractions involuntary, muscle rigidity, muscle spasms, muscle spasticity, myotonia, nuchal rigidity, oculogyration, opisthotonos, pleurothotonus, risus sardonicus, torticollis, and trismus;
2) Parkinsonian Events, which include akinesia, asterixis, athetosis, bradykinesia, cogwheel rigidity, essential tremor, extrapyramidal disorder, freezing phenomenon, gait festinating, hypertonia, hypokinesia, hypokinesia neonatal, intention tremor, masked facies, parkinson's disease, parkinsonian crisis, parkinsonian gait, parkinsonian rest tremor, parkinsonism, tremor, and tremor neonatal;
3) Akathisia Events, which include akathisia, hyperkinesia, and psychomotor hyperactivity;
4) Dyskinetic Events, which include ballismus, buccoglossal syndrome, choreoathetosis, clumsiness, dyskinesia, dyskinesia neonatal, dyskinesia oesophageal, fumbling, nodding of head, on and off phenomenon, and tardive dyskinesia;
5) Residual Events, which include chorea, huntington's chorea, muscle twitching, and myoclonus.

CCI [redacted]

The following AEs of special interest will be summarized by PT terms only:

1) Fall
2) Orthostatic hypotension
3) Syncope
4) Dizziness

Treatment-emergent AEs are defined as AEs with an onset date on or after the start of double-blind treatment. In more detail, TEAEs are all adverse events which started after start of double blind study drug treatment; or if the event was continuous from baseline and was worsening, serious, study drug related, or resulted in death, discontinuation, interruption or reduction of study therapy. Adverse Events occurring up to 30 days after the last day of double-blind dosing will be included in the summary tables.

9.2 Clinical Laboratory Tests

9.2.1 Change from Baseline in Lab Tests

Summary statistics for mean and mean change from baseline in the routine clinical laboratory measurements, prolactin concentrations, coagulation parameters (PT, aPTT, and INR), HbA1c, cortisol, ACTH, and TSH will be provided by treatment and by visit.

9.2.2 Potentially Clinically Relevant Values

In addition, the incidence of treatment-emergent potentially clinically relevant (PCR) values identified using prospectively defined criteria in Appendix 1 for laboratory tests will be summarized by treatment group. A listing of PCRs by subject and by test will be provided.

9.2.3 Potentially Liver Injury Related Laboratory Test

Total bilirubin level will be checked for any subjects with increased ALT or AST levels greater or equal to three times the upper normal limits (or baseline).

Liver injury related laboratory test results be will summarized for subjects who met following criteria in the Short-term Controlled Trials and Long-term Open-label Trials groups. The corresponding listing will be provided as well.

- AST or ALT $\geq$ 3 x ULN and
- T_Bili $\geq$ 2 x ULN

9.2.4 Metabolic Change

In addition to mean change from baseline, the incidence of treatment emergent significant changes in fasting lipids, fasting glucose, and metabolic syndrome will be summarized by treatment group using the following criteria.

Criteria for Treatment-Emergent Significant Change in Lipids and Glucose		
LAB PARAMETER	BASELINE*	ANYTIME POST BASELINE
LDL Direct, Fasting (MG/DL)	Borderline 100-<160 Normal/Borderline <160 Normal <100 Any Value	High >=160 High >=160 Borderline/High >=100 Increased >=30
HDL Cholesterol, Fasting (MG/DL)	Normal >=40 Any Value	Low <40 Decreased >=20
Triglycerides, Fasting (MG/DL)	Normal <150 Borderline 150-<200 Normal/Borderline <200 Normal <150	High 200-<500 High 200-<500 High 200-<500 Borderline/High/Very High >=150

Criteria for Treatment-Emergent Significant Change in Lipids and Glucose		
LAB PARAMETER	BASELINE*	ANYTIME POST BASELINE
	Any Value	Increased >=50
Glucose Fasting, Serum (MG/DL)	Normal <100 Impaired 100-<126 Normal/Impaired <126 Any Value	High >=126 High >=126 High >=126 Increased >=10

* Baseline is calculated from day 0; if day 0 is unavailable screen visit will be used

Criteria for Treatment-Emergent Metabolic Syndrome	
DESCRIPTION	ANYTIME POST BASELINE*
Central Obesity	Waist Circumference >=102cm(MALE), >=88cm (FEMALE)
Dyslipidemia	Triglycerides >= 150mg/dl
Dyslipidemia	HDL < 40mg/dl (MALE), <50mg/dl (FEMALE)
Supine Blood Pressure	Systolic>=130mmHg and Diastolic>=85mmHg
Glucose Fasting, Serum	>=100mg/dl
Metabolic Syndrome	Met 3 Or More of the Above Criteria at a Visit

* Baseline is calculated from day 0; if day 0 is unavailable screen visit will be used

9.3 Physical and Neurological Examination and Vital Signs

Physical and neurological examination findings will be listed by subject.

Summary statistics for change from baseline in vital signs, body weight, and waist circumference will be provided by treatment group.

In addition, the incidence of treatment-emergent PCR values identified using prospectively defined criteria in Appendix 2 for vital signs and body weight will be summarized by treatment group. Listing of PCRs by subject and by test will be provided.

9.4 12-Lead ECG

Summary statistics for change from baseline in ECG parameters will be provided by treatment and by visit.

In addition, the incidence of treatment-emergent potentially clinically relevant (PCR) values identified using prospectively defined criteria for ECG will be summarized by treatment group. Listing of PCRs by subject and by test will be provided.

For the analysis of QT and QTc, data from three consecutive complexes (representing three consecutive heart beats) will be measured to determine average values. The following QT corrections will be used for reporting purposes in the clinical study report:

1) QTcB is the length of the QT interval corrected for heart rate by the Bazett formula: QTcB=QT/(RR)0.5 and
2) QTcF is the length of the QT interval corrected for heart rate by the Fridericia formula: QTcF=QT/(RR)0.33
3) QTcN is the length of the QT interval corrected for heart rate by the FDA Neuropharm Division formula: $QTcN=QT/(RR)^{0.37}$

Categorical changes in ECG parameters during the double blind treatment period will be summarized based on the following criteria:

Categorical Change Criteria in QT/QTc Parameters		
Classification	**Category**	**Criteria**
QT	New Onset (≥ 450 Msec for men or ≥ 470 Msec for women)	New onset in QT means a subject who attains a cut off value during treatment period but not at baseline.
QTc *	New Onset (≥ 450 Msec for men or ≥ 470 Msec for women)	New onset in QTc means a subject who attains a cut-off value during treatment period but not at baseline.
	New Onset (≥ 450 Msec for men or ≥ 470 Msec for women) And > 10% Increase	New onset and > 10% increase in QTc means a subject who attains a cut off value and > 10% increase during treatment period but not at baseline
	New Onset (> 500 Msec)	New onset (> 500 msec) in QTc means a subject who attains a value > 500 msec during treatment period but not at baseline.
	Increase 30 - 60 Msec	Increase from baseline value > 30 and ≤ 60 msec in QTc
	Increase > 60 Msec	Increase from baseline value > 60 msec in QTc

* QTc categorical change criteria apply to QTcB, QTcF and QTcN.

9.5 Body Weight, Waist Circumference and Body Mass Index (BMI)

Analyses of body weight, waist circumference and BMI will be performed for the Safety Sample. Body mass index is defined as weight in kilograms divided by the square of height in meters. The mean change from baseline to Week 12 (OC) and the last visit in the double blind treatment period in body weight will be tabulated and analyzed using ANCOVA. The ANCOVA models will include the baseline as a covariate and the treatment group as fixed effect.

Percentages of subjects showing significant weight gain (≥ 7 % increase in weight), as well as percentages of subjects showing significant weight loss (≥ 7 % decrease in weight)

from baseline to Week 12 (OC and LOCF) will be analyzed using Cochran-Mantel-Haenszel (CMH) General Association Test.

9.6 CCI [redacted]

[redacted]

9.7 Mini-Mental State Examination (MMSE)

The mean changes from baseline to Week 12 (OC) and the last visit in the double blind treatment period in MMSE will be tabulated and analyzed by treatment group using ANCOVA. The ANCOVA model for the OC data set will include the baseline item score as covariate and treatment group as main effect. The ANCOVA model for last visit will include the baseline item score as covariate, and study center, treatment group as main effects. The analyses will be performed for Safety Sample.

9.8 CCI [redacted]

[redacted]

9.9 Concomitant Medications

Number and proportion of subjects taking concomitant medications prior to study therapy, during the double blind treatment period, and after study therapy are tabulated by drug classification using the WHO drug dictionary.

9.10 Extent of Exposure

The start date of double-blind study therapy - brexpiprazole or placebo - will be the first day of double-blind dosing. The number and percentage of subjects, who receive double-blind study medication, will be presented by week and by treatment group. Each dosing week will be based on the actual week; i.e., Day 1-7 in Week 1, Day 8-14 in Week 2, etc. This summary will be performed on the Safety Sample.

The number and percentage of completers will be presented by week and by treatment group.

The mean daily dosage will be summarized by week and treatment group using descriptive statistics. The mean daily dosage per subject per week will be determined for each week of the study. This will be calculated by dividing the sum of individual total doses by the number of days in the week interval. The summary will contain for each treatment group the number of subjects receiving double-blind study medication, and the mean and range of the mean daily dose for each week.

10 Conventions

10.1 Study Visit Windows

Study visit windows will be used to map visits using study day intervals. This visit window convention applies to tables and listings for all efficacy and safety scales. This derived study window variable will be named as WEEK and will be footnoted. In listings it will be listed along with the CRF study visit.

Table below shows classifications for study day intervals in the double-blind period. The variable “target day” is defined using the number of days since the start of double-blind dosing. The first day of double-blind dosing is defined as “Day 1”. If more than one observation falls within a particular study day interval, then the last observation within that interval is used. CCI [redacted]

CCI [redacted]

[redacted]

Study Day and Visit Window

Week	Target Day [a]	Study Day Interval [a]
1	7	2-10
2	14	11-17
3	21	18-24
4	28	25-31
5	35	32-38
6	42	39-45
7	49	46-52
8	56	53-59
9	63	60-66
10	70	67-73
11	77	74-80
12	84	81-91 [b]

[a] Relative to the first day of IMP in the double blind treatment period.

[b] CCI [redacted] s.

10.2 Pooling of small centers

Primary efficacy analysis will be performed on the ITT Efficacy Sample which comprises those subjects in the Randomized Sample who have a baseline value and at least one post-randomization value for CMAI Total Score in the double blind treatment period.

CCI [redacted]

CCI [redacted]

10.3 Scales: Rules for Scoring and Handling of Missing Data

10.3.1 Cohen-Mansfield Agitation Inventory (CMAI)

The CMAI consists of 29 items all rated on a 1 to 7 scale with 1 being the "best" rating and 7 being the "worst" rating. The CMAI Total Score is the sum of ratings for all 29 items. The possible total scores are from 29 to 203. The CMAI Total Score will be unevaluable if less than 24 of the 29 items are recorded. If 24 to 28 of the 29 items are recorded, the total score will be the mean of the recorded items multiplied by 29 and then rounded to the first decimal place.

CCI [redacted]

.

10.3.2 Clinical Global Impression-Severity of Illness Scale (CGI-S)

The severity of agitation for each subject will be rated using the CGI-S. To perform this assessment, the investigator (or designee) will answer the following question: "Considering your total clinical experience with this particular population, how mentally ill (as related to agitation) is the subject at this time?" Response choices are 0 = not assessed; 1 = normal, not at all ill; 2 = borderline mentally ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; and 7 = among the most extremely ill subjects. The score 0 (= not assessed) will be set to missing. The CGI-S is therefore a 7-point scale from 1 through 7.

10.3.3 CCI

[redacted]

10.3.4 CCI

[redacted]

10.3.5 Neuro-psychiatric Inventory

The NPI consists of 12 items. For each item there is a screening question to determine if the behavioral change is present (rated 1) or absent (rated 0). For each item there are three scores: frequency, severity, and caregiver distress (NPI/NPI-NH) or occupational disruptiveness (NPI-NH). Frequency is rated on a 1 to 4 scale, severity is rated on a 1 to 3 scale and the caregiver distress is rated on a 0 to 5 scale. The individual item score is calculated as presence x frequency x severity and has a range from 0 to 12. If presence is zero, the individual item score and caregiver distress score will be set to zero. For all items, low scores are 'better' than high scores.

The NPI Total Score is calculated by adding the individual item scores of the 12 items together. The possible total scores are from 0 to 144. The NPI Total Score will be

unevaluable if less than 10 of the 12 items are recorded. If 10 or 11 of the 12 items are available then the total score is the mean of the available scores times 12. All imputed scores are rounded to the first decimal place.

CCI

[redacted]

10.3.6 CCI

[redacted]

10.3.7 CCI

[redacted]

CCI [redacted]

[redacted]

10.3.8 CCI [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

10.3.9 National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA)

The NINCDS-ADRDA, which has shown good reliability and validity, provides criteria for the possible and probable diagnosis of Alzheimer's disease. These criteria require that cognitive impairment and a suspected dementia syndrome be confirmed by neuropsychological testing for a clinical diagnosis of Alzheimer's disease. The NINCDS-ADRDA criteria specify 8 cognitive domains that may be impaired in Alzheimer's disease: memory, language, perceptual skills, attention, constructive abilities, orientation, problem solving and functional abilities.

10.3.10 Hachinski Ischemic Scale (Rosen Modification)

The Rosen-modified Hachinski Ischemic Scale assesses whether a subject's dementia is likely due to vascular causes by the response to 8 questions: abrupt onset, stepwise deterioration, somatic complaints, emotional incontinence, history of hypertension, history of stroke, focal neurologic signs, and focal neurologic symptoms. The Rosen modified Hachinski Ischemic Scale will be completed to assess eligibility for the trial by the same neurologist who performs the neurological examination.

10.3.11 CCI [redacted]

[redacted]

CCI

[redacted]

10.3.15 Mini-Mental State Examination (MMSE)

The MMSE is a brief practical test for assessing cognitive dysfunction. The test consists of 5 sections (orientation, registration, attention and calculation, recall, and language) and has a total possible score of 30. The MMSE is used for screening subjects and is also to be completed at Week 12/ET.

The MMSE is a 19 item scale. Items 1 to 10, 15, and 17 to 19 are rated on a scale from 0 to 1, item 14 is rated on a 0 to 2 scale, items 11, 13 and 16 are rated on a scale from 0 to 3, and Item 12 is rated on a scale from 0 to 5. Low scores are the worst, high scores are the best. The MMSE Total Score is calculated by adding the individual item scores. The possible range for the MMSE Total Score is from 0 to 30. If the maximum total of the

missing items could contribute more than 6 points to the total score then the total score will be set to missing. Otherwise a mean non-missing items score will be calculated by summing the non-missing items and dividing them by the maximum score possible from the non-missing items. For missing items with possible scores from 0 to 1, the mean score will be imputed for each missing item. If item 14 is missing, two times this mean will be imputed for item 14. If items 11, 13, or 16 are missing three times this mean for will be imputed each missing item. If item 12 is missing 5 times then this mean will be imputed for item 12. After all by-item imputation has been done, the individual item scores will be added and this sum will be rounded to the first decimal place to arrive at an imputed total score. In other terms, the MMSE Total Score is simply the mean non-missing items score multiplied by 30, and then rounded to the first decimal place.

11 References

1 Siddiqui O, Hung JHM, O'Neill R. MMRM vs. LOCF: A comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharmaceutical Stats. 2009; 19(2):227-46.

2 Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. Applied Statistics. 1994;43:49-93.

3 Little RJA. Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc. 1993;88:125-34.

4 Little RJA. Modeling the drop-out mechanism in repeated measures studies. J Am Stat Assoc. 1995;90:1112-21.

5 Hedeker D, Gibbons RD. Application of random effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods. 1997;2:64-78.

6 Ali MW, Siddiqui O. Multiple imputation compared with some information dropout procedures in the estimation and comparison of rates of change in longitudinal clinical trials with dropouts. J Biopharmaceutical Stats. 2000;10(2):165-81.

7 Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. Biometrics. 1989; 45:939-55.

8 Rabinowitz J, Davidson M, Paul De Deyn P, Katz I, Brodaty H, Cohen-Mansfield J. Factor Analysis of the Cohen-Mansfield Agitation Inventory in Three Large Samples of Nursing Home Patients With Dementia and Behavioral Disturbance. The American Journal of Geriatric Psychiatry. 2005; 13(11): 991-998.

9 Brunner C, Spiegel R. Eine Validierungsstudie mit der NOSGER (Nurses' Observation Scale for Geriatric Patients), einem neuen Beurteilungsinstrument für die Psychogeriatrie. Zeitschrift für Klinische Psychologie. 1990; 19:211-229.

12 Potential Clinical Relevance Criteria from Protocol

Appendix 1 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	≤ 15%
Absolute neutrophil count	≤ 1,000/ mm^3
Platelet count	≤ 75,000/ mm^3 or ≥ 700,000/ mm^3
Urinalysis	
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	≤ 2.5 mEq/L or ≥ 6.5 mEq/L
Sodium	≤ 126 mEq/L or ≥ 156 mEq/L
Calcium	≤ 8.2 mg/dL or ≥ 12 mg/dL
Glucose	
Fasting	≥ 100 mg/dL
Non-Fasting	≥ 200 mg/dL
Total Cholesterol, Fasting	≥ 240 mg/dL
LDL Cholesterol, Fasting	≥ 160 mg/dL
HDL Cholesterol, Fasting	
Men	< 40 mg/dL
Women	< 50 mg/dL

Laboratory Tests	Criteria
Triglycerides, Fasting	≥ 150 mg/dL

Appendix 2 Criteria for Identifying Vital Signs 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 (baseline status not considered)
Weight	-	≥ 7% increase ≥ 7% decrease

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).

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 ≥ 200 msec	increase of ≥ 50 msec
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 ≥ 120 msec	increase of ≥ 20 msec
Infarction		
Acute or subacute	all	not present → present
Old	all	not present → present ≥ 12 weeks post study entry
ST/T Morphological		
Myocardial Ischemia	all	not present → present
Symmetrical T-wave inversion	all	not present → present
Increase in QTc	QTcF ≥ 450 msec (men) QTcF ≥ 470 msec (women)	

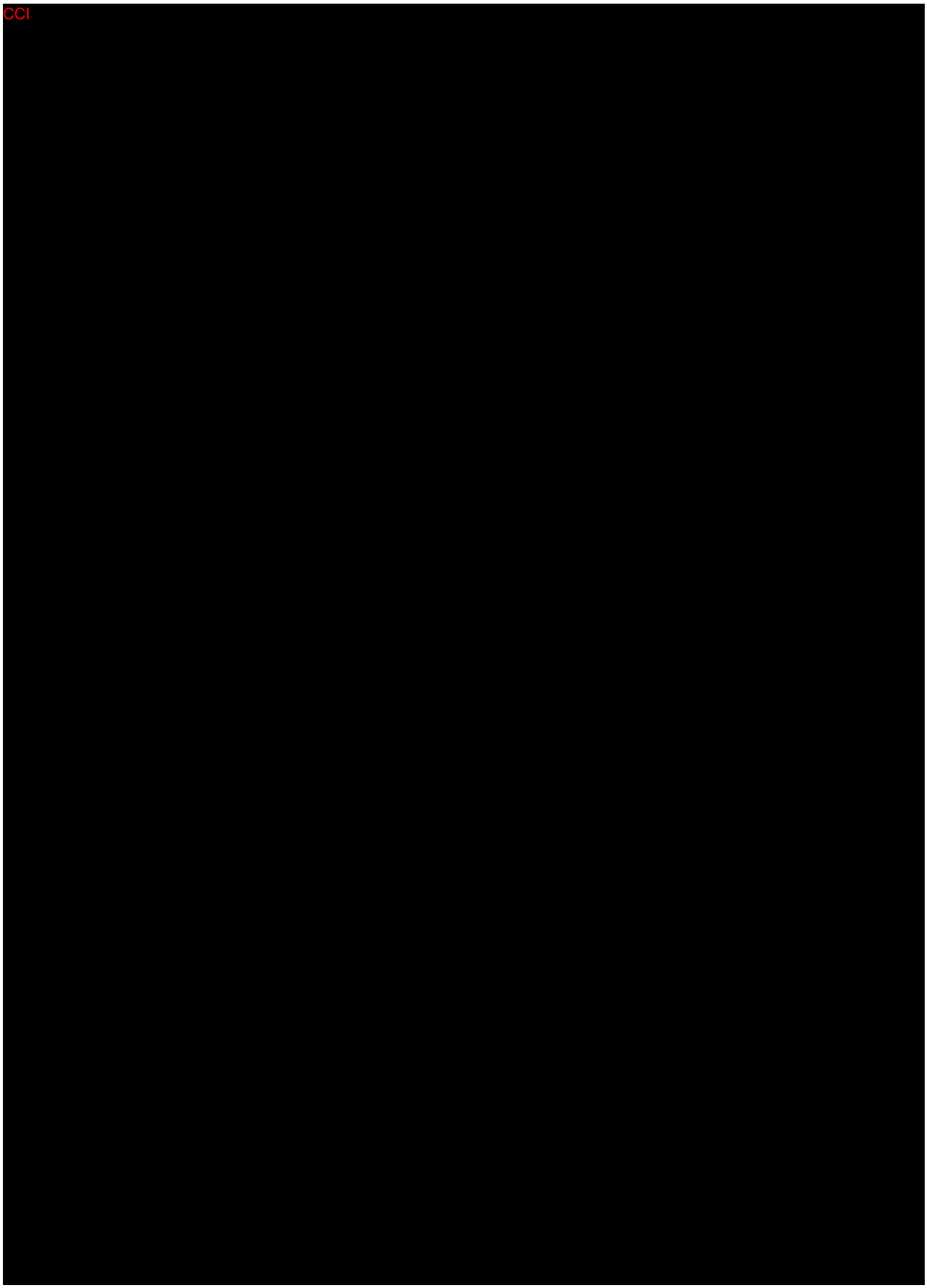
[a] In order to be identified as potentially clinically relevant, an on-treatment value must meet the "Criterion Value" and also represent a change from the subject's baseline value of at least the magnitude shown in the "Change Relative to Baseline" column.

[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.

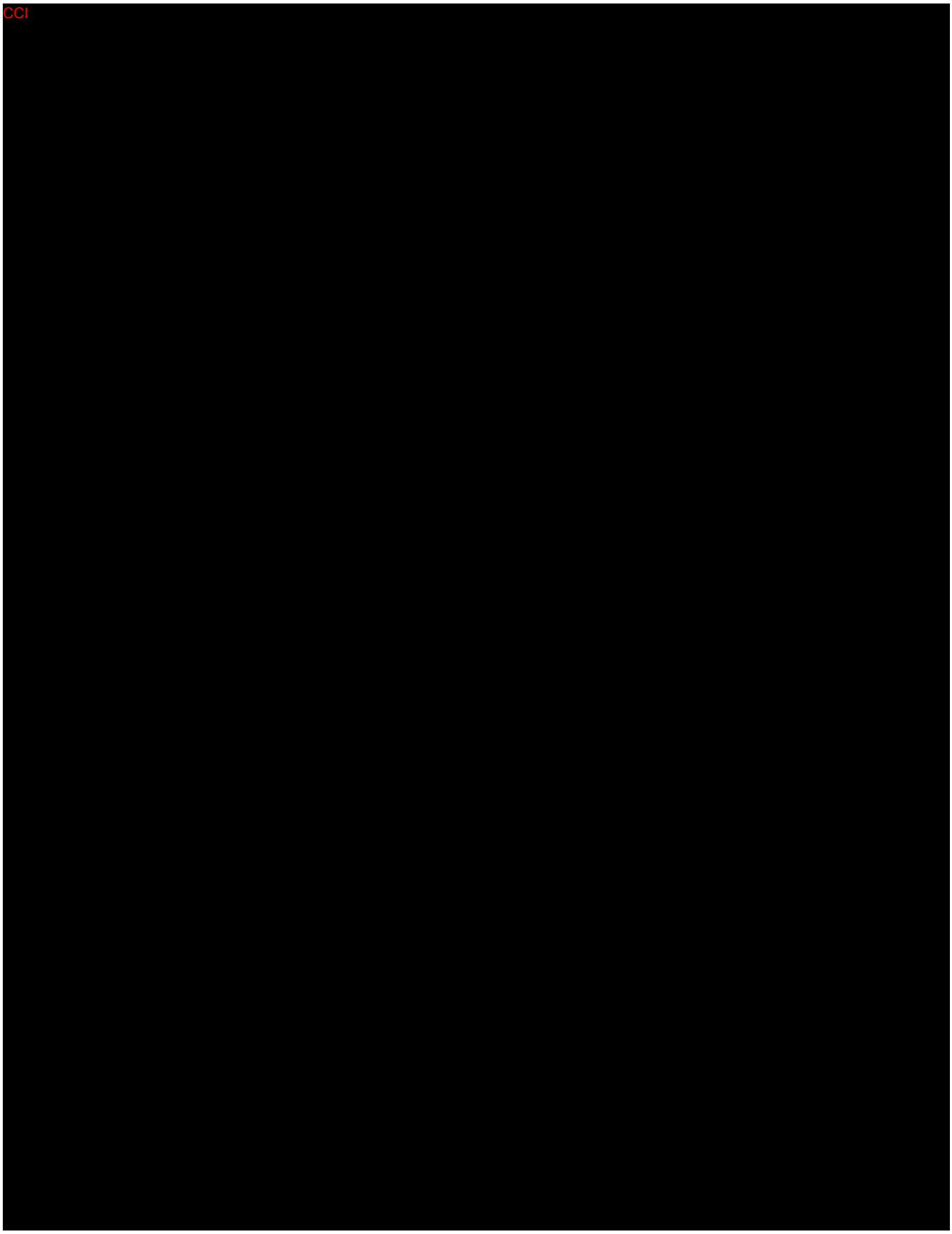
CCI



CCI

[redacted]

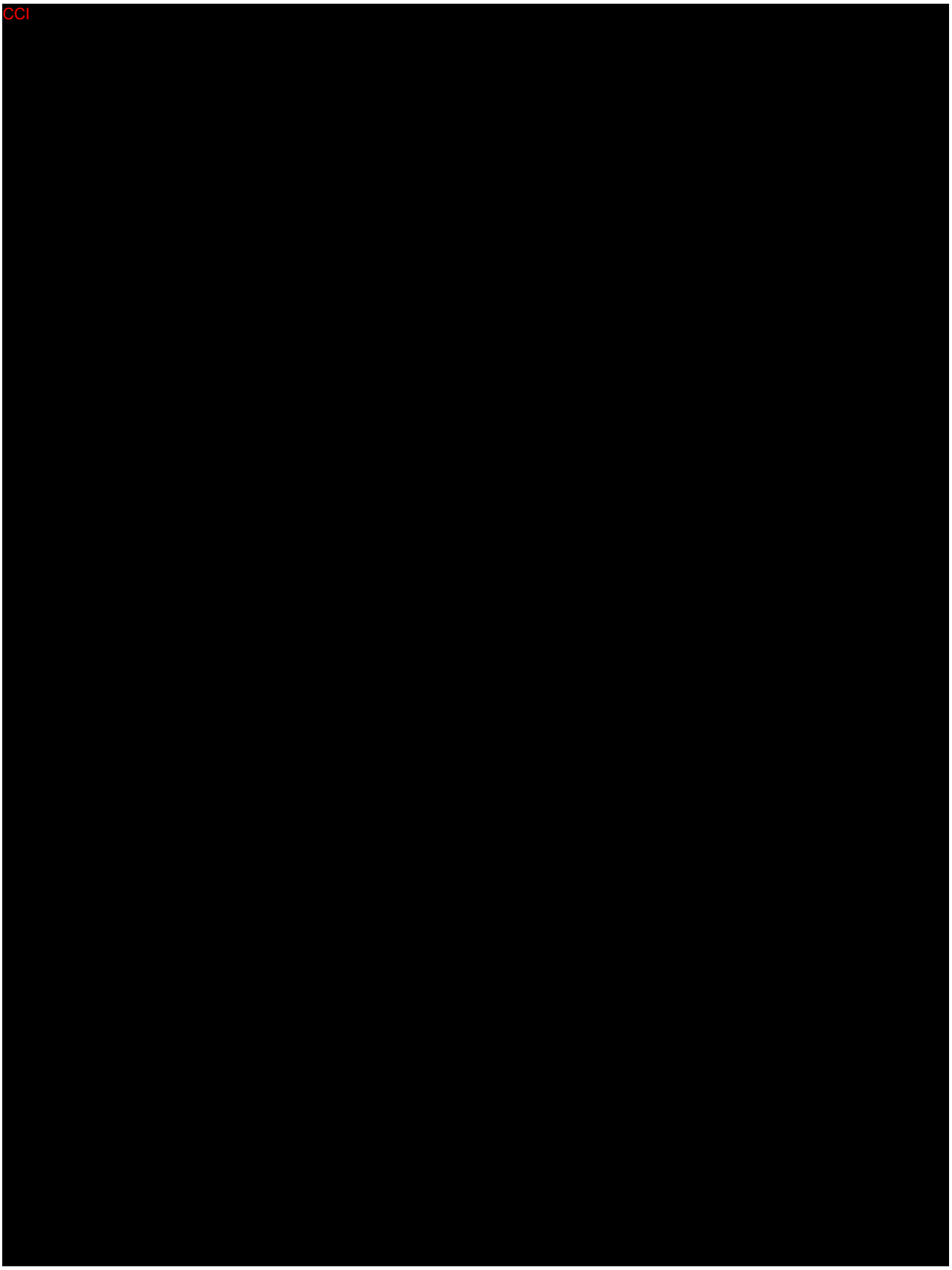
CCI



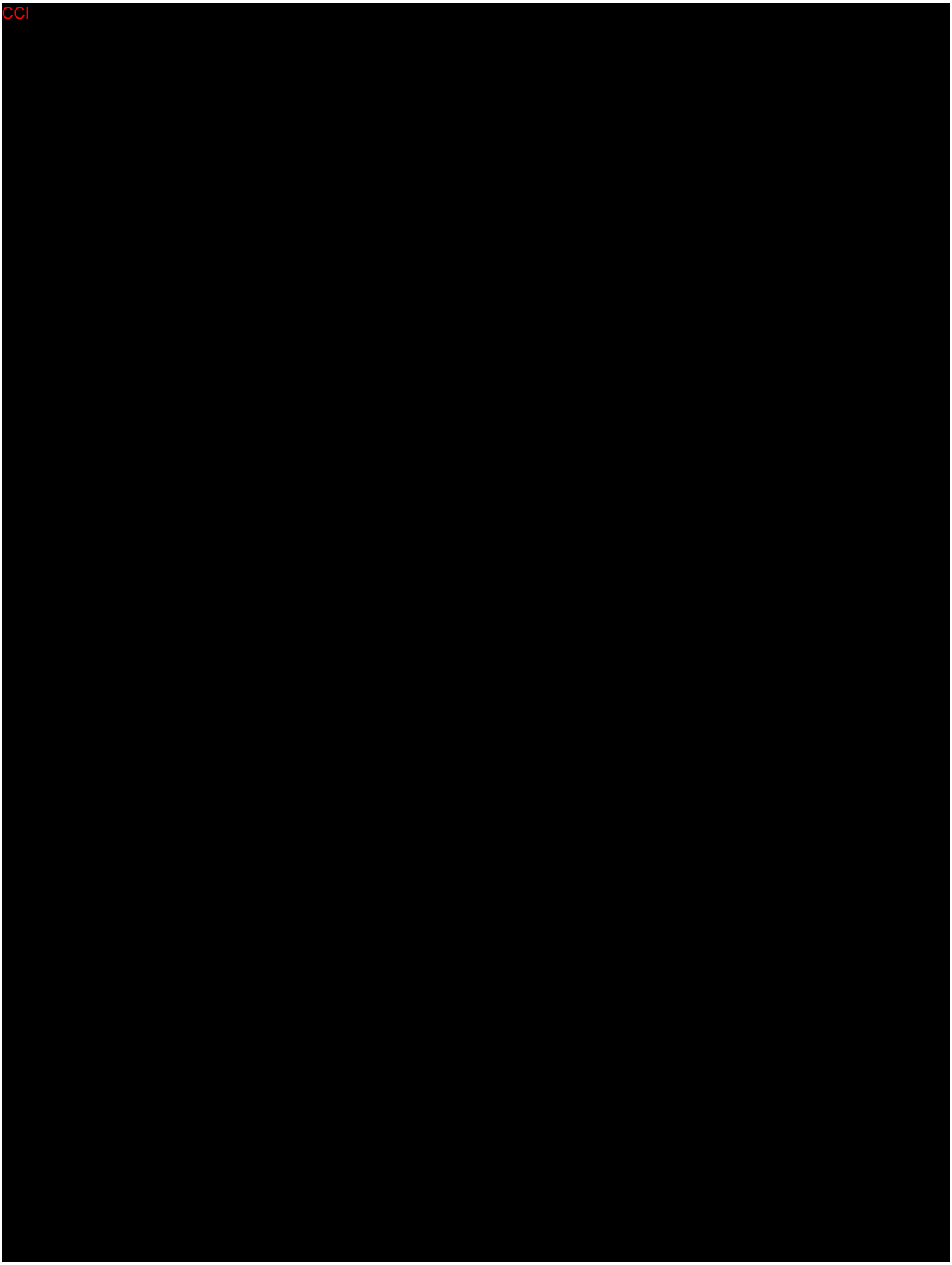
CCI

[redacted]

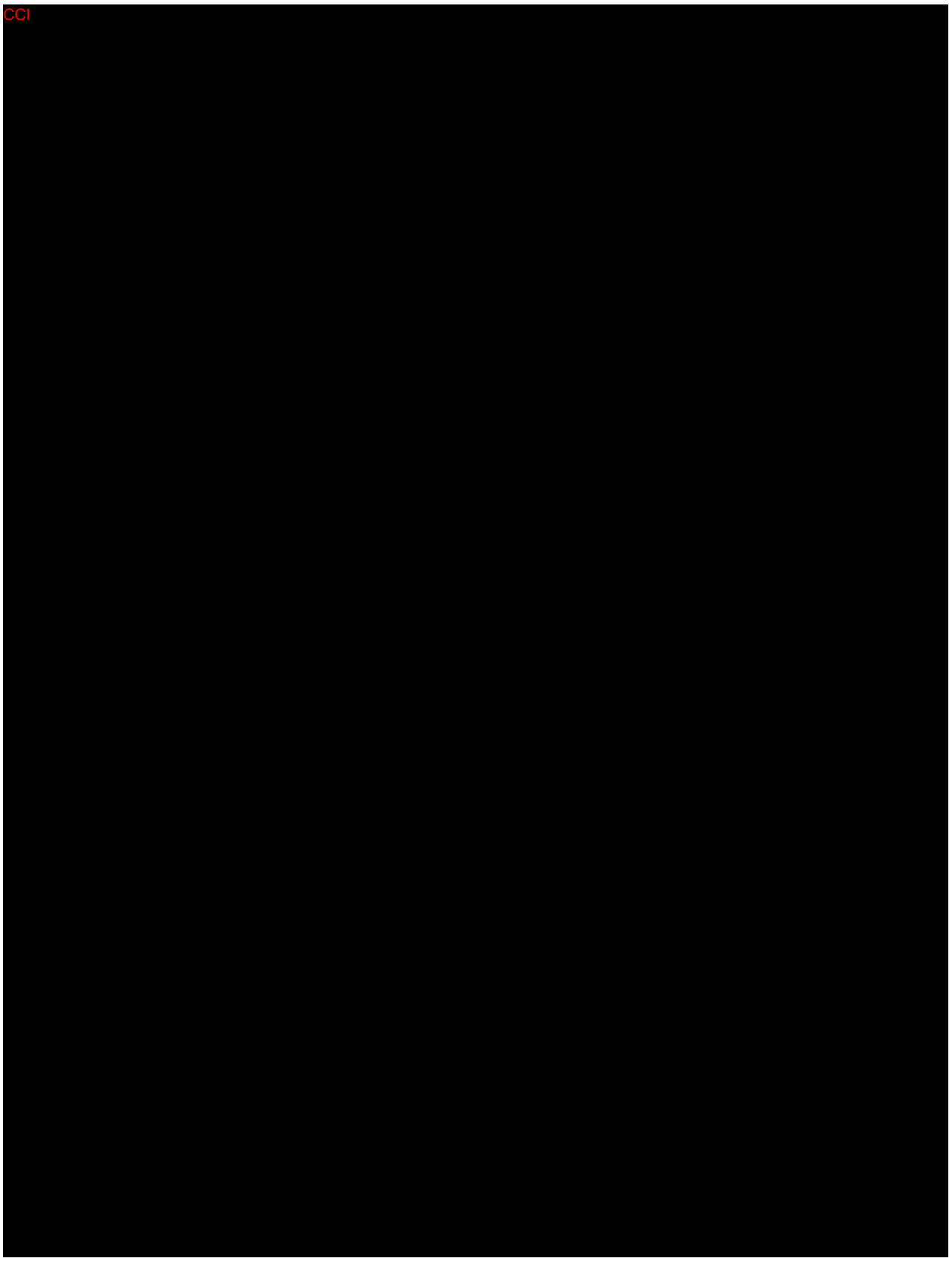
CCI



CCI



CCI



CCI

[redacted]

Otsuka Pharmaceutical Development & Commercialization, Inc.
This page is a manifestation of an electronically captured signature

OPC-34712

SIGNATURE PAGE

Document Name: P331-10-283_SAP

Document Number: 0001026892

Document Version: 3.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'GMT'Z)
PPD	Biostatistics Approval	13-Mar-2017 16:12 GMT+00
	Clinical Approval	13-Mar-2017 18:13 GMT+00