Official Title: **A Phase IIIb Multi-Center, Open-Label, Mirror-Image, Trial in Adult Subjects With Schizophrenia Treated Prospectively for 6-months With Abilify MyCite®**

NCT Number: NCT03892889

Document Date: SAP Version 2.0 : 28 October 2020

Table of Contents

Otsuka Pharmaceutical
Development & Commercialization, Inc.

Investigational New Drug
Aripiprazole (OPC-14597)

STATISTICAL ANALYSIS PLAN
for
Protocol 031-201-00301
IND No. 115927

A Phase IIIb Multi-Center, Open-Label, Mirror-Image, Trial in Adult Subjects with Schizophrenia Treated Prospectively for 6-months with Abilify MyCite

Version 2.0 Final

Date: October 28, 2020

Protocol 31-201-301 Interim Analysis Plan

Table of Contents

Protocol 31-201-301 Interim Analysis Plan

List of In-text Tables

List of In-text Figures

List of Appendices

List of Appendices Abbreviations and Definitions of Terms

Abbreviation	Definition
AE	Adverse event
app	Application
[redacted]	[redacted]
CRF	Case report form
[redacted]	[redacted]
CYP	Cytochrome P450
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ECG	Electrocardiogram
eCRF	Electronic case report form
eICF	Electronic informed consent form
[redacted]	[redacted]
ET	Early termination
FDA	Food and Drug Administration
GCP	Good Clinical Practice
IB	Investigator's Brochure
ICF	Informed consent form
ICH	International Council for Harmonisation
ICMJE	International Committee of Medical Journal Editors
ID	Identification
IEM	Ingestible event marker
IND	Investigational new drug
IRB	Institutional review board
IRE	Immediately reportable event
ITT	Intent-to-Treat
LAI	Long-acting injectable
MDD	Major depressive disorder
N or n	Number of subjects
[redacted]	[redacted]
PDC	Proportion of days covered
PQC	Product Quality Complaint
[redacted]	[redacted]
SAE	Serious adverse event
SD	Standard deviation
TEAE	Treatment-emergent adverse event
ULN	Upper limit of normal

Protocol 31-201-301 Interim Analysis Plan

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 31-201-00301. All amendments to the protocol are taken into consideration in developing this SAP.

2 Study Objectives

The objective of this study is to compare inpatient psychiatric hospitalization rates (proportion of subjects with ≥ 1 inpatient psychiatric hospitalizations) between the retrospective period (Months 1 to 3) while subjects are on oral standard-of-care antipsychotic treatment and the prospective period (Months 1 to 3) after the subjects switch to Abilify MyCite.

3 Study Design

This is a phase 3b, open-label, prospective, clinical trial designed to assess the difference between inpatient psychiatric hospitalization rates in subjects on oral standard-of-care antipsychotic treatment(s) for a period of 6 months followed by a switch to Abilify MyCite for a period of 3 months (Months 1 to 3). At the Month 3 visit, the investigator should decide if subjects will continue on Abilify MyCite for an additional 3 months (Months 4 to 6) or switch to a standard-of-care treatment (eg, oral atypical antipsychotics or a long acting injectable [LAI]) for the duration of treatment.

This trial will include male and female subjects who are 18 to 65 years of age, inclusive, with a diagnosis of schizophrenia according to Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria. Subjects must have had at least 1 inpatient psychiatric hospitalization within 4 years (48 months) and must have been prescribed oral antipsychotics for at least 6 months or longer prior to screening.

Subjects will enter a screening period (up to 45 days). If deemed eligible to participate, subjects will enter an open-label Abilify MyCite treatment prospective phase for up to 6 months. A schematic of the study design is provided in Figure 3.1-1.

3.1 Type/Design of the Trial

Oral Standard-of-Care Antipsychotic Treatment(s) (Retrospective Phase)

To be eligible for trial participation, subjects must have been prescribed oral antipsychotic standard-of-care pharmacotherapy for at least 6 months or longer, with evidence of prescription (i.e., pharmacy records) during the retrospective phase and prior

to screening. Subjects administered antipsychotics other than aripiprazole must cross-titrate onto oral aripiprazole during the screening period at the discretion of the investigator to be eligible for trial participation.

Screening

After providing informed consent, subjects will enter the 45-day (Days -45 to -1; approximately 6 weeks) screening period to determine if they are eligible to enter the Abilify MyCite treatment (prospective) phase. The subject's past psychiatric history for the 6 months prior to screening must be available, inclusive of a comprehensive review of all psychiatric hospitalizations and interventions (psychiatric only).

As safety of the subject dictates, 2 options are available to the investigator as subjects transition from the screening period to the Abilify MyCite treatment (prospective) phase:

1) Antipsychotic medications other than oral aripiprazole may be tapered off and/or discontinued during the screening period prior to the subject receiving the first treatment with Abilify MyCite.

2) Antipsychotic medications other than oral aripiprazole may be continued during the screening period and cross-titration with oral aripiprazole trial drug may occur. Some investigators may choose this option for the subject depending as local practice patterns dictate and on the subject's safety.

Abilify MyCite treatment (prospective) phase

For subjects enrolled in the trial, those not on aripiprazole at screening and who need to cross-titrate will do so during the screening period for conversion to aripiprazole from other antipsychotics according to the approved labels. At the baseline visit (Day 1), Abilify MyCite onboarding will be provided with informational materials and additional call center support. Initiation of Abilify MyCite treatment will commence at the baseline (Day 1) visit, with subjects ingesting the Abilify MyCite tablet embedded with an IEM sensor product, wearing the patch, and using the smartphone app for up to 6 months, including a required 3 months (Months 1 to 3) of Abilify MyCite treatment. This will be followed, at investigator discretion, to either change to oral standard-of-care antipsychotic treatment (eg, oral or LAI) or remain on Abilify MyCite for the second 3-month period (Months 4 to 6), during the prospective phase. The first dose of Abilify MyCite is to be taken at the baseline (Day 1) visit (in the clinic) during onboarding, unless they have already taken their daily dose. Subjects should not take their oral standard-of-care medication that morning before the baseline (Day 1) visit.

During the assessment period, subjects will visit the investigator for clinical evaluations at baseline (Day 1), Month 3, and Month 6/early termination (ET). Monthly assessments

will follow a 28-day schedule, and at each visit and at any unscheduled visits, the subject's clinical status will be evaluated. All hospitalizations and all interventions (psychiatric only as well as pharmacy cost of all medications) will be recorded, in addition to all other assessments outlined in this protocol. The primary endpoint will be assessed at Month 3. The final trial visit will be on Month 6/ET and will be followed by a 30-day (- 3/ + 2 day) safety follow-up period for subjects continuing on Abilify MyCite. For subjects completing Months 4 to 6 on standard of care, a 30-day safety follow-up period is not required. Subjects will also visit the clinic for medication dispensing and accountability at the scheduled visits. Any treatment changes, including reason for the change, will be recorded in the electronic case report form (eCRF).

Evaluation of safety and tolerability of Abilify MyCite as assessed by frequency and severity of the AEs to be collected in this trial, i.e., serious AEs (SAEs), nonserious AEs, potential hepatotoxicity cases, and pregnancies.

Subject use of the Abilify MyCite system can be monitored by their respective investigators, who are suggested to review the physician dashboard data at a minimum of every 2 weeks and make changes to current treatment plan and therapy at their discretion. The investigators may request that a subject return to the site for unscheduled visits as deemed appropriate. In the event of an unscheduled visit, review of the investigator dashboard and safety events will be collected.

Safety follow-up

All subjects who complete or withdraw from the trial while on Abilify MyCite (unless they have withdrawn their consent for participation in the trial) will receive a telephone call for safety follow-up at 30 days (- 3/ + 2 days) after the last trial visit.

This telephone contact will be made to assess if there were any AEs experienced since the last trial visit. In addition, the subject will be asked about any new medications or changes in existing medications and the information will be documented. The trial design is presented in Figure 3.1 1.

Protocol 31-201-301 Interim Analysis Plan

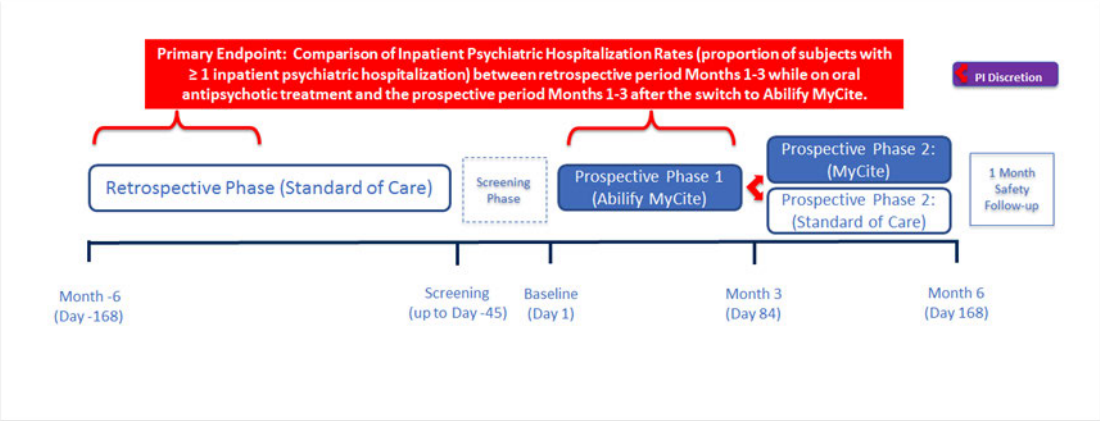

Figure 0-1 Trial Design Schematic

Note: Monthly assessments will follow a 28-day schedule.

3.2 Trial Treatment

The subjects will receive Abilify MyCite as prescribed by their investigators. The subject and the investigator will initiate the system at the baseline (Day 1) visit and use it for up to 6 months, including a required 3 months (Months 1 to 3) of Abilify MyCite treatment. This will be followed, at investigator discretion, to either change to standard-of-care antipsychotic treatment (eg, oral or LAI) or remain on Abilify MyCite for the second 3-month period (Months 4 to 6) during the prospective phase.

3.3 Study Population

The study population will include male and female subjects, 18 to 65 years of age, inclusive, with a diagnosis of schizophrenia according to DSM-5 criteria. Subjects must have had at least 1 inpatient psychiatric hospitalization within the 4 years (48 months) prior to screening but must not have been hospitalized for any psychiatric reason during the 4 weeks prior to signing the ICF or at any time during the screening period. Subjects must have been on oral antipsychotics treatment in the last 7 months prior to screening.

4 Sample Size and Power Justification

The primary endpoint of this trial compares the paired proportions of inpatient psychiatric hospitalization during the retrospective period (Month 1 to 3) and the prospective period (Month 1 to 3), which will be analyzed using a McNemar test.

Based on published results of other similar studies for the particular time periods

(retrospective Month 1 to 3 and prospective Month 1 to 3) we are evaluating, it is reasonable to assume that 15.5% of the subjects are hospitalized during retrospective Month 1 to 3 and 7% are hospitalized during prospective Month 1 to 3; about 5% of subjects who are not hospitalized retrospective will be hospitalized prospectively, while 13.5% of the subjects who are hospitalized retrospectively will not be hospitalized. With this assumption, 200 completed subjects will provide 80% power (alpha = 0.05 two-sided) to detect a difference of 8.5% in the paired hospitalization proportions between pre-switch and post-switch periods while the proportion of total discordant pairs is 18.5%. In addition, this sample size will provide 80% power for the comparison of retrospective Month 4 to 6 vs prospective Month 4 to 6, assuming 30% of the subjects are hospitalized pre-switch (12.5% subjects will be hospitalized post-switch and another 17.5% won't have hospitalization post-switch) and, overall, 20% are hospitalized post-switch (among them, 12.5% subjects are hospitalized both pre-switch and post-switch and 7.5% have no hospitalization pre-switch).

Two interim analyses are planned at when 50% and 70% of subjects have completed the prospective period (Month 1 to 3); in order to conserve the overall Type I error at 0.05 level, 224 to 300 evaluable subjects are needed to complete the 3-month treatment to maintain 80% power. This sample size assumes the 3 sequential tests are made using LanDeMets Spending Function with O'Brien-Fleming boundary, a futility conditional power of 15% is built in the sample size calculation to allow the trial to stop for futility. Based on the conditional power and the alpha level at the interim analysis, the trial could stop for efficacy or futility, continue with the initially planned 224 sample size or increase to up to 300 if the conditional power fall between 50% to 80% (Section 4.7). The 2-sided alpha levels for these 2 interim analyses are 0.00312 and 0.0139 respectively, and the alpha left for the final analysis will be 0.04528.

5 Statistical Analysis Dataset

5.1 Data/Data Sets Specifications

5.1.1 Data Sets Analyzed

The following analysis samples are defined for this study:

Enrolled Sample: comprises all subjects who sign an eICF for the trial and entered screening.

Safety Sample: comprises all subjects who receive at least one dose of study medication in the prospective phase.

Intent-to-Treat Sample: consists of all Enrolled Sample who met the study inclusion/exclusion criteria and who are intended to receive study medication in the prospective phase.

Modified Intent-to-Treat (mITT): dataset which will consist all ITT subjects entering the prospective phase who have taken at least 80% of the study medication for the specified analysis period, and had complete information on hospitalization during the specified analysis period; that is for the Month 1-3 mITT: the specified period will be Month 1-3 prospective/respective; for the Month 4-6 mITT and Month 1-6 mITT, the specified period will be Month 4-6 prospective/respective and Month 1-6 prospective/respective, respectively.

5.1.2 Definition of Baseline and Last Visit

The following definitions of baseline values are applicable to all efficacy analysis and safety analysis unless otherwise specified.

Baseline values are defined as the last available evaluations prior to or on the date of the first MyCite treatment.

Last visit is defined as the last scheduled visit with available data.

5.1.3 Study Week/Month Windows

For all the efficacy analysis with two assessment at month 3 and month 6 including [redacted] healthcare utilization record evaluation, [redacted], analysis windows for Month 3 LOCF and Month 6 LOCF are defined as the last assessment within in the timeframe defined in Table 5.1.3-1. [redacted] Healthcare utilization record evaluation, [redacted] will be performed for timepoints at Month 3 LOCF, Moth 6 LOCF and Last visit.

Table 5.1.3-1 Mapping of Trial Windows A

Visit	Number of Trial Days in Each Phase Window (days)
Month 3	**68- 97**
Month 3 LOCF	**1 – 97**
Month 6	**135 - 182**
Month 6 LOCF	**98 - 182**

[redacted]

[redacted]

5.1.4 Handling of Missing Data

5.1.4.1 Criteria for Potential Values for Scale Assessments

In order to assess sensitivity of results due to missing data, 2 types of data sets will be used to perform analyses by visit of mean change from baseline endpoints: last observations carried forwarded (LOCF) and observed cases (OC). The LOCF data set will include data recorded after baseline, if no observation is recorded at that visit, data carried forward from the immediately previous visit (scheduled or unscheduled) will be used to impute for the missing observations. Baseline data will not be carried forward to impute missing values for the LOCF data set. The OC data sets will also be used to perform efficacy analyses and as the primary data sets for safety analyses. The OC data set corresponding to a visit will consist of data from all patients who were evaluated at that visit, i.e., patients with missing data due to dropout or other reasons will not be included in the OC data set.

5.1.4.2 Partial Dates for Hospitalization

In the event that only partial dates (missing day or month, but not year) are available for the hospitalization and medical intervention record during the retrospective or prospective periods (including the admission and discharge date of inpatient hospitalizations, start and end date of psychiatric non-inpatient treatment visits, and the emergency room and/or outpatient treatment visit date), the following rules will be used to impute missing month or day:

a) <u>For partial admission/start/visit dates:</u>

1. If only the month is missing:

   If the admission/start/visit year equals to the corresponding (retrospective or prospective) period Month 1-3 start year, then the earliest corresponding period for Month 1-3 start month that will cover this event in this period will

be imputed for the missing month. If the admission/start/visit year equals to the corresponding (retrospective or prospective) period Month 4-6 start year, then the earliest corresponding period for Month 4-6 start month that will cover this event in this period will be imputed for the missing month. Otherwise, default to 'Jan'.

2. If only the day is missing:

   If the admission/start/visit year equals to the corresponding (retrospective or prospective) period Month 1-3 start year and admission/start/visit month equals to the corresponding period Month 1-3 start month, the corresponding period Month 1-3 start day or '01' will be imputed for the missing day so that the final date will be within the Month 1-3 period. If the admission/start/visit year equals to the corresponding (retrospective or prospective) period Month 4-6 start year and admission/start/visit month equals to the corresponding period Month 4-6 start month, the corresponding period Month 4-6 start day or '01' will be imputed for the missing day so that the final date will be within the Month 4-6 period. Otherwise, default to '01'.

3. If both month and day are missing:

   If the admission/start/visit year equals to the corresponding (retrospective or prospective) period Month 1-3 start year, the day will be imputed as 15th and then the month will be imputed following Scenario 1. If the admission/start/visit year equals to the corresponding (retrospective or prospective) period Month 4-6 start year, the day will be imputed as 15th and then the month will be imputed following Scenario 1. Otherwise default to Jan. 01.

b) <u>For partial discharge/end dates:</u>

1. If only the month is missing:

   Default to the same month or the following month of the admission/start date, whichever imputes an earlier date that is after the admission/start date. However, if a patient died and the discharge/end year equals to the year the subject died, default to the month the patient died, or the same month or the following month of the admission/start date, whichever imputes an earlier date that is after the admission/start date.

2. If only the day is missing:

   If a patient died, and if the discharge/end year equals to the year the patient died and discharge/end month equals to the month the patient died, the day the subject died will be imputed for the missing day. Otherwise, default to the last day of the discharge/end month.

3. If both month and day are missing:

   The day will be imputed as 15th and then the month will be imputed following Scenario 1.

6 Disposition and Demographic Analysis

6.1 Disposition of Subjects

Disposition of subjects will be summarized for overall and by center. Subject study duration and reason for discontinuation will be tabulated with the count and percent of enrolled subjects in each category of study duration or discontinuation reason.

6.2 Demographic and Baseline Characteristics

Summary statistics for demographic and baseline disease characteristics will be calculated (by gender for demographic characteristics) for all subjects enrolled. Demographic characteristics include age, race, ethnicity, body weight, height and body mass index. Baseline disease characteristics include time to diagnosis of Schizophrenia disease, [redacted] Summary statistics will consist of mean, median, range, and standard deviation (SD) for continuous variables and tabulations of frequency distributions for categorical variables.

7 Efficacy Analyses

7.1 Primary Efficacy Analysis Endpoint

The primary endpoint of this trial is the comparison of inpatient psychiatric hospitalization rates (proportion of subjects with ≥ 1 inpatient psychiatric hospitalization[s]) between the retrospective period (Months 1 to 3) while on oral standard-of-care antipsychotic treatment and the prospective period (Months 1 to 3) after the switch to Abilify MyCite. Primary analysis will use the mITT for Month 1 to 3.

7.1.1 Statistical Methods of Primary Efficacy Analysis

The primary analysis will be achieved by testing for statistically significant (at 2-sided overall alpha = 0.05). The analysis will be performed using a McNemar test on those who have hospitalization data during Months 1 to 3 prior to the screening period and during Months 1 to 3 after switch to Abilify MyCite.

Protocol 31-201-301 Interim Analysis Plan

7.1.2 Technical Details

Inpatient psychiatric hospitalization is defined as closed (locked) ward or open ward inpatient psychiatric hospitalization, where the hospitalization is not primarily for psychosocial reasons.

The start date and end date of the retrospective period Months 1-3 are calculated in the following ways: start date = (Date signing informed consent - 28*6 Days) and end date = (Date of signing informed consent – 28*3-1 Days). The start date and end date of the prospective period Months are calculated as: start date = (Mycite start date) and end date = (Mycite start date + 28*3 - 1). Similarly, the start date and end date of the retrospective period Months 4-6 are calculated in the following ways: start date = (Date signing informed consent - 28*3 Days) and end date = (Date of signing informed consent -1 Days). The start date and end date of the prospective period Months are calculated as: start date = (Mycite start date + 28*3) and end date = (Mycite start date + 28*6 - 1).

Inpatient psychiatric hospitalization is deemed to occur during the retrospective (or prospective) Months 1-3 if:

- The hospitalization starts before the start date of retrospective (or prospective) Months 1-3 and continues into Months 1-3
- The hospitalization starts during the retrospective (or prospective) period.

Similarly, Inpatient psychiatric hospitalization is deemed to occur during the retrospective (or prospective) Months 4-6 if:

- The hospitalization starts before the start date of retrospective (or prospective) Months 4-6 and continues into Months 4-6
- The hospitalization starts during the retrospective (or prospective) period

All the above hospitalization in Months 1-3 and Months 4-6 will be counted in the Month 1-6 period.

7.1.3 Sensitivity Analyses

As sensitivity analysis, the primary efficacy analysis will be performed on the mITT subjects who have completed the 3-month treatment only that is all subjects who are in study for at least 75 days (12 weeks adjusting for the 9 days allowed for visit window).

7.2 Secondary Endpoint

The secondary endpoint is improved adherence based on overall Proportion of days Covered (PDC) with Abilify MyCite versus retrospective oral atypical antipsychotics.

Protocol 31-201-301 Interim Analysis Plan

PDC is calculated from the [redacted] xxx database for prior data and from xxx for prospective data; PDC is defined as a ratio as the followings:

$$PDC = \left(\frac{\text{Number of days in period "covered"}}{\text{Number of days in period}} \right) \text{ x } 100\%$$

Numerator = denominator:[a]number of days from index to Day 84 or 168 (Month 3 or Month 6) where one or more antipsychotics was available (based on observed pharmacy fill dates and their days' supply or study medication information)

Denominator = 84 days or 168 days

In order to avoid underestimating the total days of therapy, the start date of each new prescription fill (for the same generic medication only) will be adjusted if it overlaps with the preceding fill's days' supply. It is assumed that when a patient refills a prescription before the preceding medication supply was exhausted (i.e., early refill), the new days' supply does not start until the end of the preceding days' supply for the drug (i.e. patient finishes the supply for the preceding fill before starting the new supply).[b]

PDC will be summarized by the descriptive statistics (n, mean, median, standard deviation (SD), standard error (SE), maximum, minimum) for the corresponding prior, prospective, and their difference at prior/prospective Month 1-3 and prior/prospective Month 1-6. The paired t-test will be used based on the prospective phase efficacy sample when applicable.

In addition, patients will be categorized as adherent (≥80% PDC)[c] versus non-adherent (<80% PDC) at Month 3 or Month 6. McNemar test will be used to compare the PDC adherence of prior vs prospective period at Month 3 and Month 6, respectively.

7.3 Subgroup Analyses

In earlier 2020, an outbreak of respiratory disease caused by a novel coronavirus named "Coronavirus Disease 2019" (COVID-19) had widely spread all over the world. On March 13, 2020, the President of the United States declared a national emergency in response to COVID-19. It is after this date, centers enrolled in this study had implemented remote visits

---

[a] Martin BC, Wiley-Exley EK, Richards S, Domino ME, Carey TS, Sleath BL. Contrasting measures of adherence with simple drug use, medication switching, and therapeutic duplication. Ann Pharmacother. 2009 Jan;43(1):36-44.

[b] Nau DP. [Accessed May 30, 2018]; Proportion of days covered (PDC) as a preferred method of measuring medication adherence. 2011 Available at: http://www.pqaalliance.org/images/uploads/files/PQA%20PDC%20vs%20%20MPR.pdf.

[c] Karve S, Cleves MA, Helm M, Hudson TJ, West DS, Martin BC. Good and poor adherence: optimal cut-point for adherence measures using administrative claims data. Curr Med Res Opin. 2009 Sep;25(9):2303-10.

1.1 Secondary Endpoint

in response to this pandemic. In order to assess the effects of this change on the primary efficacy endpoint and secondary endpoint, subgroup analysis is planned.
The COVID subgroup is defined based on the time the subject started the study treatment: all subjects started Mycite treatment after March 13th will start the remote access or remote changes for all their visits and will be classified as COVID=Yes; subjects started the Mycite on or before March 13th will have COVID=No.

7.4 Exploratory Efficacy Analyses

Protocol 31-201-301 Interim Analysis Plan

[redacted]

7.4.1.1 Statistical Methods

The variables relating to hospitalization data will be compared between the 3 months of the retrospective versus the corresponding prospective periods using McNemar test; similar comparisons will be done for Month 1-6 as well.

The number and cumulative or mean duration of psychiatric hospital stays (or treatment visits) will be summarized using descriptive statistics and compared between the SOC by paired t-test between the retrospective and corresponding prospective periods.

Continuous variables in the form of change from baseline will be summarized by the descriptive statistics (n, mean, median, standard deviation (SD), standard error (SE), maximum, minimum) and tested using the paired t-test for Efficacy Sample and Completer Sample separately. The frequency and percentage will be provided for the variables of discontinuation rate [redacted] based on the Efficacy Sample and Completer Sample.

The Kaplan-Meier estimate of the median discontinuous time and the overall Kaplan-Meier curve for efficacy sample will be provided for the time-to-discontinuation due to all causes.

Protocol 31-201-301 Interim Analysis Plan

7.4.1.2 Technical Details

Number of inpatient psychiatric hospital stays per subject during the 3 months (Months 1-3) of the retrospective (or prospective) period is determined as the total number of inpatient psychiatric hospitalization, closed (locked) ward or open ward, but not primarily for psychosocial reasons for each subject during the respective/retrospective periods. Cumulative duration of inpatient psychiatric hospitalization per subject is the sum of duration of each reported inpatient psychiatric hospitalization for each subject, which is defined as min(Date of Discharge, End date of the 3-months of the respective period) – max(Start Date of Admission, Start date of the 3-months of the respective period) + 1. Mean duration of inpatient psychiatric hospitalization per subject is derived as the cumulative duration divided by the number of inpatient psychiatric hospitalization for each subject. If a subject did not experience any inpatient psychiatric hospitalization during the l3-months, zero will be assigned to his/her number of inpatient psychiatric hospital stays, and to his/her cumulative and mean duration of inpatient psychiatric hospitalization. Similarly, method will be used to analyze the cumulative duration of inpatient psychiatric hospitalization and mean duration of inpatient psychiatric hospitalization during Month 1-6.

Inpatient hospitalizations for psychosocial reasons are defined as inpatient psychiatric hospitalization, closed (locked) ward or open ward, where hospitalization is primarily psychosocial reasons. For the last 3 months (Months 4-6) of the retrospective (or prospective) period, number and cumulative or mean duration of inpatient hospitalizations for psychosocial reasons are derived in similar ways to those of inpatient psychiatric hospitalization.

Number and cumulative or mean duration of inpatient non-psychiatric hospitalizations during the 3 months (Months 1-3) of the retrospective (or prospective) period are derived in similar ways to those of inpatient psychiatric hospitalization.

Psychiatric partial hospitalization program, psychiatric intensive outpatient program or psychiatric assertive community treatment program are reported in their corresponding category in the CRF panel of psychiatric non-inpatient treatment visits. For each category of the program, number of programs during the 3 months (Months 1-3) of the retrospective (or prospective) period can be determined by their number of occurrences during the respective period. Duration (i.e., the total number of stays/sessions) of each program is derived as the product of the mean number of stays/sessions attended per week (as recorded in CRF) and the overall length of the program in weeks, rounded to the smallest following integer. For both retrospective and prospective periods, the overall length of each program defined as min(Program End Date, End date of the 3 months of

the respective period) – max(Program Start Date, Start date of the 3-months of the respective period) + 1. For each subject, cumulative duration for each category of the programs is the sum of duration of each program in that category and mean duration is the average over all reported programs in that category.

Number of psychiatric emergency visit, psychiatric outpatient treatment visit, nonpsychiatric emergency room visit, nonpsychiatric urgent care facility visit, or nonpsychiatric outpatient visit can be determined from their reported occurrences in the category of emergency room and/or outpatient treatment visits from CRF. Psychiatric outpatient treatment visit will include such visits with a psychiatrist (MD) provided in a private doctor's office, with a psychiatrist (MD) provided in a hospital outpatient clinic, or of other type. Nonpsychiatric outpatient visit will include such visits with a medical doctor provided in a private doctor's office, with a medical doctor provided in a hospital outpatient clinic, or of other type.

For endpoints in the form of change from baseline, baseline definitions are provided in Section 5.1.2.

Protocol 31-201-301 Interim Analysis Plan

7.4.2 Other Analyses

The dashboard data is collected every 2 weeks by physicians or caregivers. The summary statistics of dashboard access by month, at Month 3, and Month 6 will be provided. The features on decision/discussion making will be summarized by the following categories: pill ingestion data, Subject reported reason code for missed doses, missed dose alert, activity, rest, mood, and subject reported rest quality rating. The count and percentage of general medical/treatment decisions made as a result of MyCite data will be summarized as: no change to treatment, adherence counseling, provided patient education, recommendations on lifestyle changes, medication changes, referred to other service/specialist, other. Further the subgroup of medication changes (increase dose, decrease dose, added medication, stopped medication, alter medication) will also be summarized.

8 Safety Analyses

In general, safety variables will be summarized by incidence and listed for Safety Samples.

Safety variables to be analyzed include AEs, clinical laboratory tests, vital signs, ECGs, physical examination findings, weight, height, BMI, extrapyramidal symptoms (EPS), suicidality using the C-SSRS, investigator rating of injection site reaction, subject reported visual analogue scale (VAS) score of injection site pain.

In general, summary statistics of changes from baseline will be provided for safety variables based on all available data for each post-baseline visit and for the last visit as well. Baseline and last visit measurements of safety variables are defined as in Section 5.1.2. Data from the unscheduled visits are included in the mean change from baseline calculation and in the incidence's calculation.

8.1.1 Adverse Events

All adverse events (AEs) are coded by system organ class and Medical Dictionary for Regulatory Activities (MedDRA version xx) preferred term. A treatment-emergent AE (TEAE) is defined as an AE which starts after start of study medication (MyCite), or an AE continues from baseline and is serious, study drug-related, or results in death,

discontinuation, interruption or reduction of study medication. If the AE increases severity or frequency during this study, a new AE should be recorded.

Adverse events will be analysis by the following groups:

- TEAEs by severity

Protocol 31-201-301 Interim Analysis Plan

- Potentially drug-related TEAEs
- TEAEs with an outcome of death
- Serious TEAEs
- Discontinuations due to TEAEs
- TEAEs with at least 2% incidence
- TEAEs with at least 5% incidence

In addition, deaths, SAEs, and AEs leading to discontinuation from study or study treatment will be listed.

For the purpose of this trial, device-related SAEs include any event that is related to the IEM, patch, or the phone and app (reported to Otsuka). Device-related AE/SAE will be summarized. If an AE is related to the patch, then the skin irritation scoring system will be completed by the investigator. Patch-related AEs Grade 2 or above will be considered medically significant for purposes of this trial. Scores from the skin irritation will be summarized by irrigation grade.

8.1.2 Clinical Laboratory Tests

All available clinical laboratory tests results will be listed.

8.1.3 Vital Signs

Vital signs are taken at Screening, baseline, Month 3, and Month 6/End of Treatment of. Assessments will include orthostatic (supine and standing) blood pressure and heart rate. In addition, body weight and waist circumference are measured at baseline, Week 24/End of Treatment of , and every 48 weeks. BMI is calculated for each visit at which body weight is measured. The potential clinically relevant vital sign abnormalities are listed by subject. Criteria for the potential clinically relevant vital sign abnormalities are provided in Appendix 1. Incidences of clinically relevant vital signs abnormalities based on the observation from the scheduled and unscheduled post-baseline visits.

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

Protocol 31-201-301 Interim Analysis Plan

8.1.4 Electrocardiogram Results

Twelve (12)-lead ECGs are recorded at Screening only; results will be provided in a listing.

8.1.5 Physical Examination

Physical examination data are listed by subject.

8.1.6 Columbia Suicide Severity Rating Scales

Columbia Suicide Severity Rating Scales (C-SSRS) are assessed at baselines, and at Month 3 and Month 6. C-SSRS data at baseline are summarized for incidence of reporting:

- Suicidality
- Suicidal behavior (and its 4 types)
- Suicidal Ideation (and its 5 types)

Suicidality is defined as reporting at least one occurrence of any suicidal behavior or suicidal ideation. Suicidal behavior is defined as reporting any type of suicidal behaviors (actual attempt, interrupted attempt, aborted attempt and preparatory acts or behavior). Suicidal ideation is defined as reporting any type of suicidal ideation. In addition to the above three incidences, C-SSRS data collected from post-baseline are also summarized for incidence of reporting:

- Complete suicidality
- Emergence of suicidal ideation
- Emergence of serious suicidal ideation
- Worsening of suicidal ideation
- Emergence suicidal of behavior

Any completed suicide is considered as complete suicidality. 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 any type of suicidal ideation with ideation severity rating of 4 or 5 during treatment. Worsening of suicidal ideation is defined to occur when the most severe suicidal ideation rating during treatment is more severe than its rating at baseline. Emergence of suicidal behavior is defined as having no suicidal behavior at baseline and reporting any type of suicidal behavior during treatment.

The suicidal ideation intensity total score is the sum of intensity scores of 5 items (frequency, duration, controllability, deterrents and reasons for ideation). The score of each intensity item ranges from 0 (none) to 5 (worst) which leads to the range of the total score from 0 to 25. A missing score of any item for will result in a missing total score. If no suicidal ideation is reported, a score of 0 is given to the intensity scale.

The incidence of C-SSRS events and change from baseline in suicidal ideation intensity total scores are presented by nominal study visits as recorded in CRF without mapping to time window of the study week/month.

8.2 Other Data Analysis

8.2.1 Exposure to Study Medication

Exposure to study mediation will be summarized for all subjects entered and for completers of Month 1-3, respectively. Descriptive statistics will be presented for days of exposure. Number (percent) of subjects staying on MyCite at Month 4-6 will also be provided.

8.2.2 Compliance to Study Medication

The number of subjects who missed study doses (excluding missed study doses due to premature discontinuation from the study) will be summarized by number of study dose missed. This summary will be performed with the ITT analysis set. A corresponding listing of subjects with missed doses during the study periods will be provided.

8.2.3 Concomitant Medication

Non-CNS concomitant medications (taken 30 days prior to the study entry through end of study) and CNS-active concomitant medications (taken 7 months prior to the study entry through end of study) will be summarized. Number and proportion of patients taking concomitant medications will be tabulated for Safety Sample of each drug classification using the B2 Enhanced version of WHODrug B3G (March 2020), for prior to, during and the post the study medication treatment period (including oral aripiprazole tablets).

8.2.4 Protocol Deviation

Protocol deviations will be listed and tabulated by protocol deviation type and centers.

In addition, protocol deviations affected by the COVID-19 will be summarized. Listing of subjects with protocol deviations affected by the COVID-19 will also be provided.

Protocol 31-201-301 Interim Analysis Plan

9 Interim Analysis

Two interim analysis are planned when 50% and 70% of subjects have completed the prospective phase (Months 1-3). At the end of the interim analysis, a decision will be made to continue with/without sample size increase, stop for efficacy or stop for fertility based on the p-value and conditional power from the interim data (as described in **Table 9-1**). Subjects will continue to be enrolled into the trial during the interim analysis; if the trial is stopped due to futility, all randomized subjects will be terminated; however, if the interim analysis stops the trial for efficacy, then all randomized subjects will continue until the last observation. Subjects in screening will be discontinued and no further randomization will occur. The 2-sided alpha levels for these 2 interim analyses are 0.00312 and 0.0139 respectively, and the alpha left for the final analysis will be 0.04528. The sample size will be re-estimated only based on the conditional power determined at the interim analysis. The adaptive designs methodology published by Chen, Dements and Lan (2004) will be used to increase the sample size based on the estimate of the treatment effect size of the primary efficacy endpoint at interim, possibly combined with other external information, without inflating the Type I error. Should the timing or frequency of the interim analysis be different from the previously specified, stopping boundary would be re-calculated using the alpha spending function.

Based on the interim analysis results, the trial sample size will be adjusted according to Table 9-1. While the interim analysis follows the boundaries listed in Table 9.-1, "final" analysis must be performed after a recommendation based on the interim analysis. This "final" analysis will include any subjects that met the criteria of analysis between the data cutoff for the interim analysis and the final database lock, and the alpha level for the "final" analysis will be derived based on the total number of subjects enrolled using O'Brien-Fleming spending function.

Table 9-1 Sample Size Re-estimation Schemes

Interim Results	Interim Analysis Recommendation	Sample Size (completers)
P – value less than specified alpha level at interim	Stop for efficacy	About 112 or 157
Conditional power ≥ 80%	No adaption to the planned sample size	224
Conditional power ≥ 50% and < 80%	Increase sample size up to 300	300

Protocol 31-201-301 Interim Analysis Plan

Table 9-1 Sample Size Re-estimation Schemes		
Interim Results	Interim Analysis Recommendation	Sample Size (completers)
Conditional power ≥ 15% and < 50%	No adaption to the planned sample size	224
Conditional power < 15%	Stop the trial for futility	About 112 or 157

10 References

1. Chue P, Llorca PM, Duchesne I, Leal A, Rosillon D, Mehnert A. (2005). hospitalization Rates in Patients During Long-Term Treatment With Long-Acting Risperidone Injection. *J Applied Research*, 5(2):266-74.

2. Olivares JM, Rodriguez-Martinez A, Bur´on JA, Alonso-Escolano D, Rodriguez-Morales A. (2008). Cost-Effectiveness Analysis of Switching Antipsychotic Medication to Long-Acting Injectable Risperidone in Patients with Schizophrenia: A 12- and 24-Month Follow-Up from the e-STAR Database in Spain. *Appl Health Econ Health Policy*, 6(1): 41-53.

3. YHJ Chen, DL DeMets and KKG Lan Increasing the sample size when the unblinded interim result is promising. *Statist. Med.2004,* 23 1023—1038

Appendix 1 Criteria for Potential Clinical Relevance Vital Signs Abnormalities

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 2 Table of Contents for Proposed Tables and Listings

Summary Tables:

CT-1.1 Subject Disposition by Site (Enrolled Sample)
CT-2.1 Reasons for Discontinuation (ITT Sample)
CT-2.2 Discontinuation from treatment / study due to COVID
CT-3.1.1 Demographic Characteristics by Gender and Site and Overall (ITT Sample)
CT-3.1.2 Demographic Characteristics by Gender and Site and Overall (Month 1-3 mITT)
CT-3.2.1.1 Baseline Disease Characteristics by Gender and Site and Overall (ITT Sample)
[redacted]
CT-3.2.2.1 Baseline Disease Characteristics by Gender and Site and Overall (Month 1-3 mITT)
[redacted]

CT-4.1.1 Medications Taken Prior to Start of Trial (Safety Sample)
CT-4.2 Medications Taken During Trial (Safety Sample)
CT-4.3 Medications Taken Post Trial (Safety Sample)

CT-4.4.1 Listing of ConMeds due to COVID
CT-4.4.2 Summary of Protocol Deviations (ITT Sample)

Protocol 31-201-301 Statistical Analysis Plan

CT-5.1.1 Analysis of Inpatient Psychiatric Hospitalization Rate Between the Retrospective Period Month 1-3 and the Prospective Period Month 1 - 3 (Month 1-3 mITT)
CT-5.1.2 Analysis of Inpatient Psychiatric Hospitalization Rate Between the Retrospective Period Month 1-3 and the Prospective Period Month 1 - 3 (ITT Sample)
CT-5.1.3 Analysis of Inpatient Psychiatric Hospitalization Rate Between the Retrospective Period Month 1-3 and the Prospective Period Month 1 - 3 (Month 1-3 mITT and Completer)

CT-5.1.2.1 Number of Inpatient Psychiatric Hospital Stays Per Subject Between Retrospective Period and Prospective Period Month 1-3 (Month 1-3 mITT)
CT-5.1.2.2 Number of Inpatient Psychiatric Hospital Stays Per Subject Between Retrospective Period and Prospective Period Month 1-3 (ITT Sample)
CT-5.1.2.3 Cumulative Duration of Inpatient Psychiatric Hospitalization Per Subject Between Retrospective Period and Prospective Period Month 1-3 (Month 1-3 mITT)
CT-5.2.3.4 Cumulative Duration of Inpatient Psychiatric Hospitalization Per Subject Between Retrospective Period and Prospective Period Month 1-3 (ITT Sample)
CT-5.1.2.5 Mean Duration of Inpatient Psychiatric Hospitalization Per Subject Between Retrospective Period and Prospective Period Month 1-3 (Month 1-3 mITT)
CT-5.1.2.6 Mean Duration of Inpatient Psychiatric Hospitalization Per Subject Between Retrospective Period and Prospective Period Month 1-3 (ITT Sample)
CT-5.2.1 Summary of PDC for Month 1-3 Retrospective/Prospective Period (ITT Sample)
CT-5.2.2 Comparison of PDC Adherence for Month 1-3 Retrospective/Prospective Period (ITT Sample)

CT-6.2.1 Summary of number of dashboard access by month and overall (ITT)
CT-6.2.2 Summary of features the physician/caregiver focus on discussion/decision making by month and overall (ITT)
CT-6.2.3 Summary of medical/treatment decisions made as a result of MyCite data by month and overall (ITT)

CT-6.6.1 Summary of features the physician/caregiver focus on in discussion/decision making
CT-6.6.2 Summary of medical/treatment decisions made as a result of MyCite data
CT-6.6.3 Summary of number of dashboard access
CT-6.4 Number (%) visits done remotely due to COVID

CT-7.1 Extent of Exposure to Study Medication - (Safety Sample)
CT-7.2 Summary statistics of Exposure to Study Medication (Safety Sample)

CT-8.1 Adverse Events (All Causalities) By Treatment-Safety Sample
CT-8.2.1 Incidence of Treatment-emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.2.2 Incidence of Treatment-emergent Adverse Events by System Organ Class, MedDRA Preferred Term and Severity (Safety Sample)
CT-8.3.1 Incidence of Treatment-emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.3.2 Incidence of Treatment-emergent Adverse Events by System Organ Class, MedDRA Preferred Term and Severity (Safety Sample)
CT-8.4.1 Incidence of Serious Treatment-emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.4.2 Incidence of Serious Treatment-emergent Adverse Events by System Organ Class, MedDRA Preferred Term and Severity (Safety Sample)
CT-8.5.1 Incidence of Discontinuation of Treatment Due to Treatment-Emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.5.2 Incidence of Discontinuation of Treatment Due to Treatment-Emergent Adverse Events by System Organ Class, MedDRA Preferred Term and Severity (Safety Sample)
CT-8.6.1 Incidence of Non-serious Treatment-Emergent Adverse Events Greater Than or Equal to 5% in Any Group by System Organ Class and MedDRA Preferred Term (Safety Sample)

Protocol 31-201-301 Statistical Analysis Plan

CT-8.6.2 Occurrence (Number of Events) of Non-serious Treatment-emergent Adverse Events with Incidence Greater Than or Equal to 5% in Any Group by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.7.1 Incidence of Serious Treatment-emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.7.2 Occurrence (Number of Events) of Serious Treatment-emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.8 Occurrence (Number of Events) of Potentially Medication or Serious Treatment-emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.9 Incidence of Device-related Treatment-Emergent Adverse Events by System Organ Class, MedDRA Preferred Term and Severity (Safety Sample)
CT-8.9 Incidence of Device-related Treatment-Emergent Serious Adverse Events by System Organ Class, MedDRA Preferred Term and Severity (Safety Sample)
CT-8.11 Frequency of Subjects with Skin Irritation Scores Related to Patch (Safety Sample)
CT-8.12 Incidence of Patch-related Treatment-Emergent Adverse Events with Skin Irritation Score >= 2 by System Organ Class, MedDRA Preferred Term and Severity (Safety Sample)
CT-8.13 Columbia-Suicide Severity Rating Scale(C-SSRS), Suicidality
(Safety Sample)
CT-8.14 Columbia-Suicide Severity Rating Scale(C-SSRS) Suicidality Change during Post-baseline (Safety Sample)
CT-8.15 Summary of the suicidal ideation intensity total score (Safety Sample)

CT-11.1 Criteria for Potentially Clinically Relevant Abnormalities in Vital Signs
CT-11.2.1 Listing of Potentially Clinically Relevant Abnormalities in Vital Signs During the Prospect Treatment Period (Safety Sample)
CT-11.2.2 Incidence of Potentially Clinically Relevant Abnormalities in Vital Signs During the Prospect Treatment Period (Safety Sample)
CT-11.2.3 Mean Change from Baseline of the Double-Blind Treatment Period in Vital Signs During the Prospect Treatment Period (Safety Sample)

Listings:
CT-9.1 Listing of Deaths (Safety Sample)
CT-9.2 Listing of Serious Adverse Events (Safety Sample)
CT-9.3 Listing of Discontinuation of Treatment Due to Adverse Events (Safety Sample)
CT-9.4 Listing of Patch-related Treatment-emergent Adverse Events (Safety Sample)
CT-9.5 AE listing for subjects with confirmed or presumed diagnosis of COVID

AE-1.1 Adverse Events for All Enrolled Subjects
DREAS-1 Discontinued Subjects and Reason for Discontinuation
DREAS-2 Adverse Events for Subjects Who Withdrew Consent
EFF-1.1 Prior and Post Inpatient Psychiatric Hospitalization for mITT – Month 1-3
EFF-1.2 Prior and Post Inpatient Psychiatric Hospitalization for mITT – Month 4-6
EFF-1.3 Prior and Post Inpatient Psychiatric Hospitalization for ITT Sample
EFF-1.4 Prior and Post All Other (non-inpatient) Psychiatric Treatment Visits Excluding Inpatient Psychiatric Hospitalization for ITT Sample

EFF-1.5 PDC by Subject (ITT Sample)

Protocol 31-201-301 Statistical Analysis Plan

[redacted]

PDATA-1.1.1 Consent and Eligibility
PDATA-1.1.2 Inclusion Criteria and Exclusion Criteria
PDATA-2 Subject Enrollment List
PDATA-3.1 Subject Disposition
PDATA-3.2 Study Completion Status and Reasons for Discontinuation
PDATA-4 Demographics Characteristics
PDATA-5.1 Psychiatric History
PDATA-5.2 Diagnosis of Schizophrenia
PDATA-6.1 Concomitant Medications: Medications Taken Prior to Start of Study
PDATA-6.2 Concomitant Medications: Medications Taken During Study
PDATA-6.3 Concomitant Medications: Medications Taken Post Study
PDATA-7.1 Study Medication - MyCite
PDATA-7.5 Study Medication Accountability – Abilify MyCite
PDATA-8 Medical History
PDATA-9 Physical Examination
PDATA-10 Adverse Events
PDATA-11 Vital Signs
PDATA-12.1 Electrocardiogram Results
PDATA-12.2 Electrocardiogram Results - Raw 3 Beat Data
PDATA-13.1 Hospitalization and Medical Interventions - 6 Months Retrospective Period
PDATA-13.2 Hospitalization and Medical Interventions - Prospective Period
[redacted]

PDATA-17.1 Visual Analog Scale for Subject-reported Rating of Injection Site Pain
PDATA-17.2 Investigator Assessment of Most Recent Injection Site
[redacted]
PDATA-18 Birth Control Status
PDATA-19 Screening Failures
PDATA-20 Follow-up
PDATA-21 Protocol Deviations - CRF
LAB-1 Laboratory Test Results: Urine Pregnancy Test
PDEV-1 Protocol Deviations by Center and Type of Deviation
PDEV-2 Protocol Deviations by Subject
PDEV-3 Protocol Deviations Criteria

Figures:

Figure 5.1 Kaplan-Meier Curve of Time to Discontinuation Due to All Causes

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: P031_201_00301_Statistical Analysis Plan Final

Document Number: 1000105170

Document Version: 2.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
[redacted]	Biostatistics Approval	05-Nov-2020 21:15:23
[redacted]	Clinical Approval	05-Nov-2020 13:56:51

Otsuka Pharmaceutical
Development & Commercialization, Inc.

Investigational Medicinal Product

Aripiprazole (OPC-14597)

Protocol No. 031-201-00301

A Phase IIIb Multi-Center, Open-Label, Mirror-Image, Trial in Adult Subjects with Schizophrenia Treated Prospectively for 6-months with Abilify MyCite

INTERIM ANALYSIS PLAN

Date: July 31, 2020

Protocol 31-201-301 Interim Analysis Plan

Table of Contents

List of In-text Tables

List of Appendices

Protocol 31-201-301 Interim Analysis Plan

1 Introduction

This interim analysis plan (IAP) documents the statistical methodology, data analysis algorithms, and conventions to be applied in the interim analysis and reporting of Trial 31-201-301.

2 Statistical Methods

2.1 Sample Size and Power Justification

The primary endpoint of this trial compares the paired proportions of inpatient psychiatric hospitalization during the retrospective period (Month 1 to 3) and the prospective period (Month 1 to 3), which will be analyzed using a McNemar test.

Based on published results of other similar studies for the particular time periods (retrospective Month 1 to 3 and prospective Month 1 to 3) we are evaluating, it is reasonable to assume that 15.5% of the subjects are hospitalized during retrospective Month 1 to 3 and 7% are hospitalized during prospective Month 1 to 3; about 5% of subjects who are not hospitalized retrospective will be hospitalized prospectively, while 13.5% of the subjects who are hospitalized retrospectively will not be hospitalized. With this assumption, 200 completed subjects will provide 80% power (alpha = 0.05 two-sided) to detect a difference of 8.5% in the paired hospitalization proportions between pre-switch and post-switch periods while the proportion of total discordant pairs is 18.5%. In addition, this sample size will provide 80% power for the comparison of retrospective Month 4 to 6 vs prospective Month 4 to 6, assuming 30% of the subjects are hospitalized pre-switch (12.5% subjects will be hospitalized post-switch and another 17.5% won't have hospitalization post-switch) and, overall, 20% are hospitalized post-switch (among them, 12.5% subjects are hospitalized both pre-switch and post-switch and 7.5% have no hospitalization pre-switch).

Two interim analyses are planned at when 50% and 70% of subjects have completed the prospective period (Month 1 to 3); in order to conserve the overall Type I error at 0.05 level, 224 evaluable subjects are needed to complete the 3-month treatment to maintain 80% power. This sample size assumes the 3 sequential tests are made using LanDeMets Spending Function with O'Brien-Fleming boundary, a futility conditional power of 15% is built in the sample size calculation to allow the trial to stop for futility. Based on the conditional power and the alpha level at the interim analysis, the trial could stop for efficacy or futility, continue with the initially planned 224 sample size or increase to up to 300 if the conditional power fall between 30% to 80% (Section 2.2). The 2-sided alpha

levels for these 2 interim analyses are 0.00312 and 0.0139 respectively, and the alpha left for the final analysis will be 0.04528.

2.2 Interim Efficacy Analysis

The purpose of the interim analyses is to decide whether there is sufficient evidence that the study will be stopped to declare efficacy or futility, or addition subjects are needed to have sufficient power to achieve the primary objective.

The primary endpoint of this trial is the comparison of inpatient psychiatric hospitalization rates (proportion of subjects with ≥ 1 inpatient psychiatric hospitalization[s]) between the retrospective period (Months 1 to 3) while on oral standard-of-care antipsychotic treatment and the prospective period (Months 1 to 3) after the switch to Abilify MyCite.

The analysis on the primary endpoint will be performed using a McNemar test, those who have hospitalization data during Months 1 to 3 prior to the screening period and during Months 1 to 3 after switch to Abilify MyCite for the modified Intent-to-Treat (mITT) sample; this will be a 2-sided test.

Two interim analysis are planned when 50% and 70% of subjects have completed the prospective phase (3 months). At the end of the interim analysis, a decision will be made to continue with/without sample size increase, stop for efficacy or stop for futility based on the p-value and conditional power from the interim data (as described in Table 2-1). Subjects will continue to be enrolled into the trial during the interim analysis; if the trial is stopped due to futility, all randomized subjects will be terminated; however, if the interim analysis stops the trial for efficacy, then all randomized subjects will continue until the last observation. Subjects in screening will be discontinued and no further randomization will occur. The 2-sided alpha levels for these 2 interim analyses are 0.00312 and 0.0139 respectively, and the alpha left for the final analysis will be 0.04528. The sample size will be re-estimated only based on the conditional power determined at the interim analysis. The adaptive designs methodology published by Chen, Dements and Lan (2004) will be used to increase the sample size based on the estimate of the treatment effect size of the primary efficacy endpoint at interim, possibly combined with other external information, without inflating the Type I error. Should the timing or frequency of the interim analysis be different from the previously specified, stopping boundary would be re-calculated using the alpha spending function.

Based on the interim analysis results, the trial sample size will be adjusted according to Table 2-1.

Protocol 31-201-301 Interim Analysis Plan

Table 2-1 Sample Size Re-estimation Schemes

Interim Results	Interim Analysis Recommendation	Sample Size (completers)
P – value less than specified alpha level at interim	Stop for efficacy	About 112 or 157
Conditional power ≥ 80%	No adaption to the planned sample size	224
Conditional power ≥ 30% and < 80%	Increase sample size up to 300	300
Conditional power ≥ 15% and < 30%	No adaption to the planned sample size	224
Conditional power < 15%	Stop the trial for futility	About 112 or 157

2.2.1 Data Set Analyzed

The following analysis samples are defined for this study:

Enrolled Sample: comprises all subjects who sign an E-informed consent form (eICF) for the trial.

Safety Sample: comprises all subjects who receive at least one dose of study medication in the prospective phase.

Intent-to-Treat Sample: consists of all Enrolled Sample who met the study inclusion/exclusion criteria and who are intended to receive study medication in the prospective phase.

Modified Intent-to-Treat (mITT): dataset which will consist all ITT subjects entering the prospective phase who have taken at least 80% of the study medication for the specified analysis period, and had complete information on hospitalization during the specified analysis period; that is for the Month 1-3 analysis: the specified period will be Month 1-3 prospective/respective; for the Month 4-6 mITT and Month 1-6 mITT, the specified period will be Month 4-6 prospective/respective and Month 1-6 prospective/respective, respectively.

2.2.2 Sensitivity Analysis

As sensitivity analysis, the primary efficacy analysis will be performed on the mITT subjects who have completed the 3-month treatment only.

3 Data Analysis

The interim analysis will be performed by the Otsuka Pharmaceuticals. The interim analysis plan will be finalized before the first interim analysis.

In addition to the efficacy analysis mentioned in Section 2, subject disposition, reasons for discontinuation, demographics and baseline disease characteristics will be summarized. Summary statistics will consist of mean, median, range, and standard deviation (SD) for continuous variables and tabulations of frequency distributions for categorical variables.

The dashboard data will be summarized by month and Month 3. The summary statistics of dashboard access by month and Month 3 will be provided. The count and percentage of features on decision/discussion making and general medical/treatment decisions made as a result of MyCite data will be summarized per the categories collected on the CRF.

The incidences of the treatment-emergent adverse events (TEAEs) serious adverse event (SAE) are summarized by system organ class, preferred term and by severity and relationship to study treatment. Discontinuation due to TEAEs will also be summarized. For the first interim analysis at about 50% of the subjects complete 3 months treatments, only data up to Month 3 (including AEs and SAEs) will be included in the analysis..

4 Confidentiality of Results

Interim database will be cleaned and frozen for the interim analysis. The decision to continue or stop the trial will be strictly based on the prespecified criteria in Table 2-1.

An internal IARC will be set up from clinical and statistics functions to make the decision based on the interim analysis results and the prespecified criteria.

5 References

YHJ Chen, DL DeMets and KKG Lan Increasing the sample size when the unblinded interim result is promising. *Statist. Med.2004,* 23 1023—1038.

Appendix 1 Table of Contents for Tables and Listings in the Interim Analysis Report

Interim Analysis Summary Tables

CT-1.1 Subject Disposition by Site (Enrolled Sample)

CT-2 Reasons for Discontinuation (ITT Sample)

CT-3.1.1 Demographic Characteristics by Gender and Site (ITT Sample)

CT-3.1.2 Demographic Characteristics by Gender and Site (Month 1-3 mITT)

CT-3.2.1 Baseline Disease Characteristics by Gender and Site (ITT Sample)

CT-3.2.2 Baseline Disease Characteristics by Gender and Site (Month 1-3 mITT)

CT-5.1.1 Analysis of Inpatient Psychiatric Hospitalization Rate Between the Retrospective Period Month 1-3 and the Prospective Period Month 1 - 3 (Month 1-3 mITT)

CT-5.1.2 Analysis of Inpatient Psychiatric Hospitalization Rate Between the Retrospective Period Month 1-3 and the Prospective Period Month 1 - 3 (Month 1-3 mITT and 3 Month Completers)

CT-5.1.3.1 Number of Inpatient Psychiatric Hospital Stays Per Subject Between Retrospective Period and Prospective Period (Month 1-3 mITT)

CT-5.1.3.2 Cumulative Duration of Inpatient Psychiatric Hospitalization Per Subject Between Retrospective Period and Prospective Period (Month 1-3 mITT)

CT-5.1.3.3 Mean Duration of Inpatient Psychiatric Hospitalization Per Subject Between Retrospective Period and Prospective Period (Month 1-3 mITT)

CT-5.2.1 Summary of number of dashboard by month and overall (Month 1-3 ITT)

CT-5.2.2 Summary of features the physician/caregiver focus on discussion/decision making by month and overall (Month 1-3 ITT)

CT-5.2.3 Summary of medical/treatment decisions made as a result of MyCite data by month and overall (ITT)

CT-8.1 Adverse Events Till Month 3 (All Causalities) (Safety Sample)

CT-8.2.1 Incidence of Treatment-emergent Adverse Events by Drug, System Organ Class and MedDRA Preferred Term Till Month 3 (Safety Sample)
CT-8.2.2 Incidence of Treatment-emergent Adverse Events by Drug, System Organ Class, MedDRA Preferred Term and Severity Till Month 3 (Safety Sample)
CT-8.3.1 Incidence of device related Treatment-emergent Adverse Events by Drug, System Organ Class and MedDRA Preferred Term Till Month 3 (Safety Sample)
CT-8.3.2 Incidence of device related Treatment-emergent Adverse Events by Drug, System Organ Class, MedDRA Preferred Term and Severity Till Month 3 (Safety Sample)

Ct-8.3.3 Incidence of Discontinuation of Treatment Due to Treatment-Emergent Adverse Events by Drug, System Organ Class and MedDRA Preferred Term Till Month 3 (Safety Sample)

CT-8.4.1 Incidence of Serious Treatment-Emergent Adverse Events by System Organ Class up Till Month 3.

CT-8.5.1 Incidence of Discontinuation of Treatment Due to Treatment Emergent Adverse Events Till Month 3

by System Organ Class, MedDRA Preferred Term (Safety Sample)

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: P031_201_00301_IAP

Document Number: 1000092444

Document Version: 2.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
[redacted]	Clinical Approval	31-Jul-2020 17:46:04
[redacted]	Biostatistics Approval	31-Jul-2020 16:08:32