Table of Contents

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

Otsuka Pharmaceutical Development & Commercialization, Inc.

Digital Therapeutic

CT-152

A Multi-center, Randomized, Controlled Trial to Evaluate the Effectiveness of a Digital Therapeutic (CT-152) as Adjunctive Therapy in Adult Subjects Diagnosed with Major Depressive Disorder

Protocol No. 345-201-00002

Statistical Analysis Plan

Version: 1.0

Date: Feb 11, 2022

CONFIDENTIAL – PROPRIETARY INFORMATION

Confidential
May not be used, divulged, published or otherwise disclosed
without the consent of Otsuka Pharmaceutical Development & Commercialization, Inc.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

Table of Contents

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

CCI [redacted]

CCI [redacted]

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

List of In-text Tables

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

List of In-text Figures

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

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 345-201-00002. All amendments to the protocol are taken into consideration in developing this SAP.

2 Study Objectives

Primary objective: To compare the effectiveness of CT-152 with sham, in adult subjects diagnosed with MDD who are on ADT monotherapy.

Safety objective: To evaluate the safety of CT-152 in adult subjects diagnosed with MDD who are on ADT monotherapy.

3 Study Design

This is a multi-center, randomized, controlled trial to evaluate the effectiveness of CT-152 in adult subjects diagnosed with MDD who are on ADT monotherapy for the treatment of depression. The trial population will include male or female subjects aged 22 to 64 years old at the time of informed consent, with a current primary diagnosis of MDD based on the criteria in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), single or recurrent episode, without psychotic features and do not meet criteria for MDD with mixed features subtype, and who are on ADT monotherapy.

Subjects will participate in the trial for up to 13 weeks. The trial will include a screening period of up to 3 weeks, a treatment period for 6 weeks, and an extension period for 4 weeks. See Figure 3-1 for a schematic of the trial design.

The screening period begins after informed consent has been obtained. Subjects who fulfill entry criteria at the screening visit will download the digital mobile application on their smartphone and receive access to an onboarding software module. A call center can assist with the downloading of the digital mobile application. During the screening period, subjects will become familiar with the software. The subject's understanding of, and interest in, the trial will be demonstrated through adequate adherence to onboarding requirements. This will be assessed by the investigator via confirmation with the subject and completion of tasks by the subject CCI [redacted] during the allotted 3-week screening window.

Following the screening visit, subjects will be considered eligible based upon the following:

- Adherence and performance on the onboarding software module by the subjec CCI [redacted]

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

- Continuing to meet all inclusion and no exclusion criteria based on investigator assessment.

On Day 1, the eligible subjects will be randomized 1:1 (CT-152 or sham) across approximately 50 trial sites in the US. The sample size at any single trial site will be capped at approximately 15% of the total subjects randomized into the trial. Randomization will be stratified by trial site.

During the treatment period (Day 1 [baseline] to Week 6), subjects will have a remote visit at Weeks 2, 4, and 6 and will be contacted by telephone at Weeks 1, 3, and 5. Subjects will be expected to be adherent with the recommended schedule of their digital mobile application exercises during the treatment period.

After Week 6, subjects will continue participation in the trial during the extension period (Weeks 7 to 10). In the extension period, the digital mobile applications will remain installed for each group, with the Emotional Faces Memory Task (EFMT) and Shapes Memory Task (SMT) no longer available. Psychotherapy content provided previously will remain available for optional reference in the CT-152 group but no new therapeutic content will be introduced and no required treatment schedule is in place. The 2 groups will each receive brief short message service (SMS) messages in the extension period reminding subjects of the previously completed CT-152 or sham treatment courses, and will continue their ADT. Subjects will have a remote visit at Weeks 8 and 10 and will be contacted by telephone at Weeks 7 and 9. The end of the trial will be Week 10.

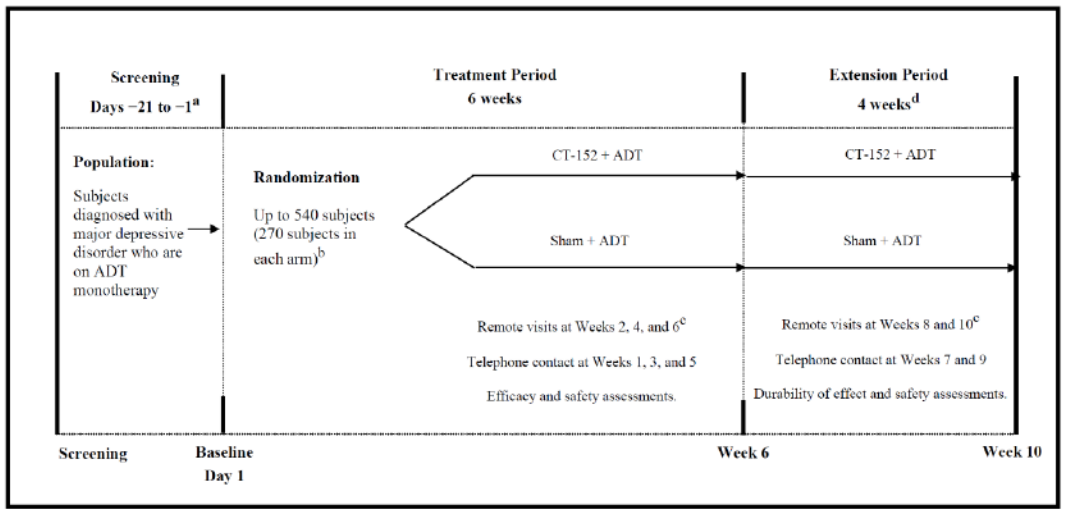


Figure 3-1 Trial Design Schematic

ADT = antidepressant therapy.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

Note: All visits will be performed remotely. The screening visit may be performed in person at the discretion of the investigator.

[a] The subject's understanding of, and interest in, the trial will be demonstrated through adequate adherence to onboarding requirements CCI [redacted]

[b] The initial sample size is 360 subjects randomized in total. The final sample size could be increased to 540 subjects (270 subjects in each arm) as per recommendation of the Data Monitoring Committee.

[c] If deemed necessary by the investigator, additional evaluations may also be performed at an unscheduled visit either remotely or in person.

[d] After Week 6, subjects will continue participation in the trial during the extension period (Weeks 7 to 10). In the extension period, the digital mobile applications will remain installed for each group. Subjects will receive brief SMS messages reminding subjects of the previously completed CT-152 or sham treatment courses, and will continue their ADT.

4 Sample Size and Power Justification

The primary efficacy endpoint is the change from baseline at Week6 in Montgomery-Asberg Depression Rating Scale (MADRS) total Score. The trial will compare CT-152 to sham, randomized at a ratio of 1:1, with an overall alpha of 0.05 for the primary efficacy endpoint.

The initial sample size is calculated to detect a 3-point difference between CT-152+ADT and sham+ADT in the change from baseline in MADRS total score with 85% power at a 2-sided $\alpha = 0.05$ level, assuming a common standard deviation of 9. The resulting sample size is 324 evaluable subjects in total (162 subjects in each arm). To compensate for subjects that fail to have evaluable assessments of MADRS total score in the full analysis set (FAS) Sample (estimated at up to 10% of all subjects), a total of 360 subjects (180 subjects in each arm) will be randomized in this trial.

Due to the limitations of applying assumptions on the treatment effect size, and in order to ensure adequate power of the trial, an unblinded interim analysis (IA) will be conducted by a Data Monitoring Committee (DMC) on when approximately the first 180 randomized subjects had an opportunity to complete the Week 6 visit. The final sample size could be increased to 540 subjects (270 subjects in each arm) as per recommendation of the DMC. Using the O'Brien-Fleming spending function, an alpha level of 0.003 (2-sided) is allocated to this interim analysis. The alpha level left for the final analysis is 0.049 (2-sided). The power and sample size were obtained using the PASS 14 (2015) statistical computing software.

5 Data Sets for Analysis and Missing Data

5.1 Data Sets for Analysis

The following analysis datasets are defined for this trial:

- Enrolled Sample – all subjects enrolled in the trial.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

- Randomized Sample – all subjects allocated based on the randomization process and recorded in the database. Subjects treated without being randomized will not be considered as randomized and therefore will not be included in any efficacy or safety population analyses.
- Safety Sample – all randomized subjects who receive at least 1 occurrence of either CT-152 or sham use. All safety assessments will be performed on the safety analysis set. If any subjects receive a treatment that differs from the assigned treatment, then the safety analyses will be conducted on the treatment actually received.
- FAS Sample – all randomized subjects who receive at least 1 occurrence of either CT-152 or sham use, and have baseline and at least 1 postbaseline assessment of MADRS total score. Subjects included in FAS Sample are defined as evaluable subjects.
- Per-protocol (PP) Analysis Sample – all subjects in the FAS Sample who were treated for 6 weeks, did not have a major protocol deviation, and were adherent to CT-152 or sham treatment in this trial. The classification of protocol deviations will be established in a blinded review by the sponsor in a meeting prior to breaking the blind. Specifically, the following criteria will be used to exclude subject from PP analysis sample.

  1. Subjects in the FAS Sample but not treated for 6 weeks

  CCI [redacted]

  3. Subjects who reported concomitant medication use that will impact the primary efficacy and key secondary endpoints

  4. Subjects who had major protocol deviation that will impact primary efficacy and key secondary endpoint

  5. Subject who received the wrong study digital therapeutic

  The PP Analysis Sample will only be used for sensitivity analysis of the primary efficacy and key secondary variables.

The core dataset for all efficacy analyses is the FAS Sample, which is created based on the intent-to-treat (ITT) principle. However, as will be described below, in order to handle missing data and restrictions imposed by different types of analyses (e.g., change from baseline analysis), other datasets derived from the FAS dataset will be used for the efficacy analyses.

5.2 Handling of Missing Data

The MADRS is utilized as the primary efficacy assessment of a subject's level of depression. The MADRS consists of 10 items, all rated on a 0 to 6 scale with zero being the "best" rating and 6 being the "worst" rating. The MADRS total Score is the sum of ratings for all 10 items; therefore, possible total scores range from 0 to 60. The MADRS total Score will be unevaluable if less than 8 of the 10 items are recorded. If 8 or 9 of the 10 items are recorded,

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

the MADRS total Score will be the mean of the recorded items multiplied by 10 and then rounded to the first decimal place.

The mixed-effect model repeated measure (MMRM) assumes data are missing at random (MAR), which is a reasonable assumption in longitudinal clinical trials in MDD[1]. However, the possibility of "missing not at random" (MNAR) data can never be ruled out. As sensitivity analyses, selection model[2], pattern-mixture model[3,4,5,6], and/or shared parameter model[7] will be used to explore data missing mechanisms of MNAR and investigate the response profile of dropout patients by last dropout reason under MNAR mechanism for the following 3 scenarios: 1) Dropout reasons due to either adverse event (AE) or lack of efficacy (LOE) as MNAR, 2) Dropout reasons due to either AE or LOE or subject withdrew consent as MNAR, 3) All dropouts as MNAR using both 1) Delta adjustment imputation method which is to departure from MAR assumption by progressively increasing the delta until conclusion from the primary analysis is overturned, and 2) Placebo based imputation methods in which missing data for both sham and CT-152 group are imputed based on the imputation model derived from sham data. If CT-152 improved outcomes prior to dropout, this benefit is carried into subsequent imputed values, but will diminish over time in accordance with the correlation structure. Details are provided in Section 8.1.3 Sensitivity Analysis.

The observed-cases (OC) data set will consist of actual observations recorded at each scheduled visit during the treatment period and/or extension period and no missing data will be imputed. MMRM, Wu-Bailey, and pattern-mixture model will be performed on the OC dataset.

The last-observation-carried-forward (LOCF) data set will include data recorded at a scheduled visit during or, if no observation is recorded at that visit, data carried forward from the previous scheduled visit. For the treatment period, baseline data will not be carried forward to impute missing values. For the extension period, the only imputation will be applied to Week 10 data. If no observation is recorded at Week10, available observations from Week 8 will be carried forward. Week 6 data will not be carried forward to impute missing values at Week 8. The analysis of covariance (ANCOVA) analysis will be primarily performed on the LOCF data. The model will include treatment and study center as main effects, and baseline value as a covariate. The ANCOVA analysis will also be applied on the OC dataset, where study center will not be included in the model.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

6 Study Conduct

6.1 Subject Disposition, Completion Rate, and Reasons for Discontinuation

Subject disposition will be summarized on the Randomized Sample by treatment group and by center.

Subject completion rate and reasons for discontinuation will be summarized on the Randomized Sample by treatment group.

6.2 Adherence to Digital Therapeutic

Adherence data to digital therapeutic use in the treatment period will be analyzed on Safety Sample by treatment group. Number and percentage of subjects will be summarized based on the number of treatment sessions they complete[redacted] Mean, range, and standard deviation of the number of treatment sessions completed will also be provided. The listing of protocol adherence by subject will be provided.

6.3 Protocol Deviation

Protocol deviations including the types of deviations will be summarized by center and treatment group on Randomized Sample. Listing of protocol deviations will be provided.

7 Baseline Characteristics

7.1 Baseline Definition

For the variables which are collected at baseline, the baseline is the Baseline measurement (expected to be at Day 1). Baseline measurement is defined as the last available measurement prior to the start of intervention with double-blind digital therapeutic (CT-152 or sham) in the treatment period.

7.2 Demographic Characteristics

Baseline demographic characteristics include age, sex, race, and ethnicity. For the Randomized Sample, demographic characteristics will be summarized by treatment group using the descriptive statistics. 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 Medical and Psychiatric History

A summary of medical and psychiatric history will be presented on the Randomized Sample by treatment group and overall. Listings of medical and psychiatric history will be provided.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

7.4 Neuropsychiatric Diagnosis

A summary of MINI International Neuropsychiatric Interview (M.I.N.I.) will be presented on the Randomized Sample by treatment group and overall. The number and percentage of patients who meet each diagnosis criteria, and the number and percentage of patients with each primary diagnosis will be summarized.

7.5 Baseline Psychiatric Evaluation

For the Randomized Sample, baseline psychiatric scale evaluation will be summarized by treatment group and overall. The mean, median, range and standard deviation will be used to summarize the assessments of: MADRS total score, GAD-7 total score, CGI-S score, and WHODAS 2.0 total score. Descriptive statistics of PHQ-9 total score, and CCI at screening visit will also be provided.

8 Efficacy Analysis

All efficacy analyses pertaining to the treatment period and extension period will be performed on the FAS Sample, and patients will be included in the treatment group as randomized.

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

8.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from the baseline defined in Section 7.1 to Week 6 in MADRS total score.

8.1.1 Primary Efficacy Analysis

The null hypothesis of the statistical test comparing CT-152+ADT and sham+ADT, based on the primary efficacy endpoint, is that the change in MADRS total score using CT-152+ADT is equal to the change in MADRS using the sham+ADT.

$$H_0: \Delta\mu_{\text{MADRS CT-152+ADT}} = \Delta\mu_{\text{MADRS sham+ADT}}$$

$$H_A: \Delta\mu_{\text{MADRS CT-152+ADT}} \neq \Delta\mu_{\text{MADRS sham+ADT}}$$

The primary analysis will be conducted on the change from baseline in MADRS total score to the final on-therapy evaluation (Week 6) based on the FAS Sample adjusted for the baseline MADRS total score.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

CCI [redacted]. This trial is considered positive if the trial will be stopped at the interim analysis for efficacy, or if the p-value of the statistical comparison based on the primary efficacy endpoint at final is < 0.049.

8.1.1.1 Primary Estimand

The primary estimand defining the treatment effect of interest in the protocol uses the hypothetical strategy specified in the International Council for Harmonisation (ICH) E9 Addendum. The objective of the primary analysis is to compare the efficacy of CT-152 and sham. The estimand, or target of estimation, following the hypothetical strategy is the treatment effect seen, had no withdrawals occurred. Subjects who withdraw from a CT-152 or sham either could have lost their treatment effect, had the subjects not taken any other treatment after withdrawal, or could have their treatment effect been masked, had the subjects taken other treatment after withdrawal. This means that any observations taken after subjects stop CT-152 or sham will most likely not contribute relevant information about the treatment effect. Due to this strategy, the last collected efficacy assessment after premature trial discontinuation will be done only once at the early termination (ET) visit. Every effort will be made to complete all of the ET evaluations prior to administering any other treatment. In the case of terminal or lost to follow-up events no ET evaluations would be expected, and only scheduled assessments performed before such an event has occurred.

The primary estimand for this trial is defined by the following components:

- Target population: FAS Sample
- Endpoint: change from baseline to Week 6 in the MADRS total score
- Intercurrent events: premature treatment discontinuation
- Measure of intervention effect: difference in endpoint means between CT-152+ADT and sham+ADT

In this hypothetical strategy, the event of withdrawing CT-152+ADT or sham+ADT is considered missing at random (MAR), and the primary endpoint of the trial could be considered as a combination of the measurements of on-treatment completers at Week 6 and the imputation of the endpoint to Week 6 following the trend in each treatment group using the mixed model repeated measurements (MMRM) method to impute missing data for subjects who withdraw CT-152+ADT or sham+ADT during the trial. All data collected during the trial treatment period will be used for statistical analysis. For the primary efficacy analysis, the treatment effect will be estimated using the MMRM method described below.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

Under the MAR assumption, MMRM provides an unbiased estimate of treatment effect for the treatment period. Analyses with missing values imputed by multiple imputation under missing not at random (MNAR), and other methods will be performed as sensitivity analyses.

8.1.1.2 Primary Analysis Method

The primary analysis will utilize a MMRM with an unstructured (UN) variance covariance matrix in which the change from baseline in MADRS total score at scheduled treatment period visits (Weeks 2, 4, and 6) will be the dependent variable based on the OC data set on FAS Sample. The model will include fixed class effect terms for treatment, visit, treatment by visit interaction, and site to assess heterogeneity of treatment effects. This model will also include an interaction term of visit by baseline MADRS total score as a covariate. The primary comparison between the CT-152+ADT or sham+ADT at Week 6 will be estimated as the difference between least squares means utilizing the computing software SAS procedure PROC MIXED.

The mathematical formula for the MMRM model is as follows:

$$\boldsymbol{y} = \boldsymbol{X}\beta + Zu + \epsilon$$

Where

$\boldsymbol{y}$ denotes the observed vector of change from baseline to Week 6 in MADRS total score;

$\boldsymbol{X}$ is the known matrix for the fixed effects and covariate;

β is an unknown fixed-effect parameter vector;

Z is the known matrix for the random effect of subject;

u is an unknown vector of random effect of subject with mean **0** and variance-covariance matrix **G** (unstructured covariance matrix);

ϵ is an unknown vector of random errors with mean **0** and variance **R.**

In case there is a convergence problem with the MMRM model with the unstructured variance-covariance matrix, the following structures other than unstructured will be used in order of 1) heterogeneous toeplitz (TOEPH), 2) heterogeneous autoregressive of order 1 (ARH1), and 3) heterogeneous compound symmetry (CSH). The first (co)variance structure converging to the best fit will be used as the primary analysis. If a structured covariance has to be used, the "sandwich" estimator of the standard error of the fixed effects parameters will be used in order to deal with possible model misspecification of the covariance matrix.

Small trial sites will be defined as trial sites that do not have at least one evaluable subject (evaluable with regard to the primary efficacy variable) in each treatment arm. All small trial sites will be pooled to form "pseudo sites" CCI [redacted]

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

CCI

[redacted]

8.1.2 Technical Computation Details for Primary Efficacy Analysis

The SAS code for the PROC MIXED procedure to carry out the above MMRM analysis with an unstructured variance covariance structure is illustrated as follows:

```
proc mixed;
    class treatment site visit subjid;
   model change=treatment center visit treatment*visit baseline*visit  / s cl
ddfm=kenwardroger;
   repeated visit /type=un subject=subjid r rcorr;
   lsmeans treatment*visit / pdiff cl alpha=0.05 slice=visit;
run;
```

where baseline is the last available measurement prior to the start of intervention with double-blind digital therapeutic in the treatment period.

8.1.3 Sensitivity Analyses

8.1.3.1 Sensitivity Analyses for Missing at Random (MAR) Assumption

Traditionally the dropout mechanisms are divided into three types (Little, 1995): (1) Missing Completely at Random (MCAR), in which the probability of dropout doesn't depend on the observed data and the missing data; (2) Missing at Random (MAR), in which the probability of dropout depends on the observed data, and (3) Missing Not at Random (MNAR), where the probability of dropout depends on the missing data and possibly the observed data.

Most of MNAR methods (Diggle P, Kenward MG, 1994) have treated all observations with dropout as if they fall within the same dropout type. In practice, we would find that different dropout reasons may be related to the outcomes in different ways, for example, detailed dropout reasons for this study are: adverse events (AE), lack of efficacy (LOE), lost to follow-up, protocol deviation, sponsor discontinued study, subject met (protocol specified) withdrawal criteria, subject was withdrawn from participation by the investigator, and subject withdrew consent to participate. Dropout due to an AE and LOE may lead to MNAR dropout. Subject withdrew consent may also lead to MNAR dropout. However, it is debatable whether

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

a dropout caused by subjects withdrew consent is MAR or MNAR. Except AE, LOE, and subject withdrew consent, all the other dropout reasons may be assumed as either MCAR or MAR dropout. Dropout due to COVID-19 will also be assumed as MAR.

As sensitivity analyses for missing at random (MAR) assumption, analyses for missing not at random (MNAR) will be carried out. Pattern Mixture Models (PMM) based on Multiple Imputation (MI) with mixed missing data mechanisms will be used to investigate the response profile of dropout patients by last dropout reason under MNAR mechanism for the following three scenarios:

1) Dropout reasons due to either AE or LOE as MNAR
2) Dropout reasons due to either AE or LOE or subject withdrew consent as MNAR
3) All dropouts as MNAR

Delta Adjustment Imputation Methods

This MNAR sensitivity analysis is to departure from MAR assumption by progressively increasing the delta until conclusion from the primary analysis is overturned. The delta is 0%, 10%, 20%, 30%, ..., 100% of the expected treatment difference of 3 points and/or the observed treatment difference between CT-152 and sham from the primary analysis of MMRM model until conclusion of the primary analysis is overturned. When delta=0 it is MAR. When delta > 0 it is MNAR.

1) Using Monte Carlo Markov Chain (MCMC) methodology from PROC MI to impute the intermittent missing data to a monotone missing pattern
2) Using a standard MAR-based multiple imputation approach from PROC MI to impute the monotone missingness data
3) For patients in the treated group and with a dropout reason of AE or LOE or subject withdrew consent, a delta will be added for all the values after the dropout time
4) Using analysis of covariance (ANCOVA) model to analyze the completed data through multiple imputation.
5) Obtaining the overall results using PROC MIANALYZE.

Placebo Based Imputation Methods

Similar to "Standard" multiple imputations, except parameters for imputation model obtained from only the placebo (sham) group. Missing data for both sham and CT-152 group are imputed based on the imputation model derived from sham data. If CT-152 improved outcomes prior to dropout, this benefit is carried into subsequent imputed values, but will diminish over time in accordance with the correlation structure.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

In addition, model based MNAR methods such as the shared parameter model (Wu and Baily, 1989) and random coefficient pattern mixture model (Hedeker D, Gibbons RD, 1997) will be also performed.

LOCF and OC Analyses

Change from baseline for the MADRS total score will be evaluated using ANCOVA with baseline value as covariate and treatment and, in LOCF analyses, trial site as main effects. For the OC analyses, trial site will not be included in the model.

8.1.3.2 Per Protocol Analyses

Per-protocol analyses will be conducted as sensitivity analyses to the primary efficacy endpoint and key secondary endpoint described in Section 8.2 on the PP analysis set.

8.1.3.3 Sensitivity Analyses for Violation of Normality Assumption

The primary endpoint MMRM analysis is a maximum likelihood method that relies on normality assumption. Residual analyses will be carried out to examine model assumption.

In the case of gross violations of the normality assumptions, nonparametric van Elteren test[10] (van Elteren, 1960) will be performed to compare treatment effect at Week 6 on both LOCF dataset and Multiple Imputation (MI) data. The van Elteren test is a generalized CMH procedure useful for stratified continuous data in non-normal setting. It belongs to a general family of Mantel-Haenszel mean score tests. The test is performed via SAS procedure PROC FREQ, by including CMH2 and SCORES=MODRIDIT options in the TABLE statement. The stratification factor is trial site.

In addition, other methods that are robust to distributional assumption will also be performed to provide different views on the primary efficacy result, these include generalized estimating equations (GEE), weighted GEE (WGEE), and MI-robust regression[11].

For MI-van Elteren test and MI-robust regression, imputation datasets will be generated with SAS MI procedure, each dataset will be analyzed, then an overall estimate is derived with SAS MIANALYZE procedure.

8.1.3.4 Sensitivity Analyses for Randomized Sample

Primary analysis will be performed on Randomized Sample with multiple imputation method to impute the missing MADRS total scores, including subjects who are excluded from FAS Sample, that is, those who do not complete the baseline and any postbaseline MADRS assessment. The detailed steps of the method are described below:

1) Using MCMC methodology from PROC MI to impute the intermittent missing data to a monotone missing pattern, with 100 multiple imputations

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

2) Using a standard MAR-based multiple imputation approach from PROC MI to impute the monotone missingness data
3) Using ANCOVA model to analyze the completed data through multiple imputation for the primary efficacy endpoint of change from baseline to Week 6 in MADRS total score. Fixed variables in the model are treatment, treatment by visit and pooled center, with baseline MADRS by visit as covariate
4) Obtaining the overall results using PROC MIANALYZE.

8.1.4 Subgroup Analyses

Subgroup analyses of change from baseline in MADRS total score to every scheduled visit in the treatment period will be performed by the following factors:

- Sex (Based on the biological status)
- Race (White and All Other Races)
- Age group (Age < 55 and Age ≥ 55)
- Cannabis use (subjects reporting current cannabis use at screening or with positive THC test results at screening)

All subgroup analyses will be conducted using the same MMRM analysis as for the primary efficacy analysis except that the fixed class effect term for trial center will not be included in the model.

Interaction effects of treatment-by-subgroup will be assessed at Week 6 for the subgroups identified in the previous paragraph. MMRM analyses will be performed by adding addition of terms for subgroup-by week and treatment-by-subgroup-by-week. These treatment-by-subgroup interaction analyses will be presented in statistical documentation.

8.2 Key Secondary Efficacy Endpoint

The key secondary efficacy endpoint is the change from baseline to Week 6 in the GAD-7 total score. It will be analyzed using the same method (MMRM) as in the primary analysis with a replacement of the interaction term of visit by baseline GAD-7 total score as a covariate. The null hypothesis of the statistical test comparing CT-152+ADT and sham+ADT, based on the key secondary efficacy endpoint, is that the change from baseline to Week 6 in the GAD-7 total score using CT-152+ADT is equal to the change from baseline to Week 6 in the GAD-7 total score using sham+ADT.

$$H_0: \Delta\mu_{\text{GAD-7 CT-152+ADT}} = \Delta\mu_{\text{GAD-7 sham+ADT}}$$

$$H_A: \Delta\mu_{\text{GAD-7 CT-152+ADT}} \neq \Delta\mu_{\text{GAD-7 sham+ADT}}$$

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

where $\Delta\mu_{\text{GAD-7 CT-152+ADT}}$ and $\Delta\mu_{\text{GAD-7 sham+ADT}}$ are the change from baseline to Week 6 in the GAD-7 total score in CT-152+ADT and sham+ADT treatment groups, separately.

In order to control the overall type I error rate, the statistical tests based on the primary efficacy endpoint and the key secondary efficacy endpoint will be done using a hierarchical approach. If the primary efficacy analysis for the MADRS total score yields a statistically significant result with the pre-specified alpha level at the IA or the final analysis (if the trial does not stop at the IA) for the comparison of CT-152 and sham, then the corresponding comparison for the key secondary efficacy endpoint (GAD-7 total score) will be tested using the same alpha-spending as that of the primary endpoint.

Should the primary outcome not achieve statistical significance in the final analysis, the key secondary endpoint will be examined as an other efficacy endpoint.

8.2.1 Efficacy Endpoints for Assessment for Durability

The durability of effect of CT-152 will be assessed based on MADRS total score and GAD-7 total score.

It will include 3 MADRS assessments at Weeks 6, 8, and 10. In addition to demonstrating an CCI with statistical significance, durability will be demonstrated by a point estimate of the difference in change from baseline at Weeks 8 and 10 CCI when comparing CT-152 and sham. Such point estimates will be provided using the same MMRM model in the primary efficacy analysis with change from baseline in MADRS total score at Week 8 and 10 in the extension period as the dependent variable.

Durability based on GAD-7 will also include 3 assessments, at Weeks 6, 8, and 10, demonstrating a numerically larger improvement on point estimate of the difference in change from baseline in GAD-7 total score at Weeks 8 and 10 in CT-152 compared to sham. Such point estimates will be provided using the same MMRM model in the key secondary efficacy analysis with change from baseline in GAD total score at Week 8 and 10 in the extension period as the dependent variable.

8.3 Other Efficacy Endpoints

Other efficacy analyses are listed below. All other efficacy endpoints will be evaluated at a nominal 0.05 level (2-sided) without adjusting for multiplicity.

1) Change from baseline to Weeks 2 and 4 in the MADRS total score
2) Change from baseline to Weeks 2 and 4 in the GAD-7 total score
3) Change from baseline to Weeks 2, 4, and 6 in the CGI-S score.
4) Change from screening to Weeks 4 and 6 in the PHQ-9 total score.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

5) Change from baseline to Week 6 in the WHODAS 2.0 total score.
6) MADRS response at Weeks 2, 4, and 6, where response is defined as ≥ 50% reduction in MADRS total score from baseline
7) MADRS partial response at Weeks 2, 4, and 6, where partial response is defined as ≥ 30% and < 50% reduction in MADRS total score from baseline
8) MADRS response at Weeks 8 and 10 , where response is defined as full or partial, and defined as ≥ 30% reduction in MADRS total score from baseline.

Endpoints (1) through (4) will be evaluated using the same MMRM model described in the primary analysis. The LOCF and OC analyses using ANCOVA will also be conducted. The LOCF analysis model includes treatment and trial site as main effects, and baseline value as covariate. For the OC analyses, trial site will not be included in the model.

Endpoint (5) will be evaluated using ANCOVA with baseline as covariate and treatment and, in LOCF analyses, trial site as main effects. Endpoints (6) through (7) will be evaluated by Cochran-Mantel-Haenszel (CMH) General Association Test controlling, in LOCF analyses, for trial site. OC analyses will also be conducted for variables (5) through (7) but will not control for trial center.

Endpoint (8) will be analyzed using descriptive statistics on LOCF dataset.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

CCI

[redacted]

8.6 Interim Analysis and Adaptive Design

An unblinded interim analysis of efficacy data will be conducted on approximately the first 180 randomized subjects. The unblinded interim analysis will be carried out when these subjects have either completed the Week 6 visit or discontinued prior to Week 6. The difference between CT-152 and sham based on the primary efficacy endpoint will be tested at the unblinded interim analysis. The sponsor will remain blinded to the IA results, and will only receive DMC's recommendation as per the interim analysis plan (IAP) and the DMC charter.

To protect the overall type I error at the 2-sided alpha level of 0.05 (2-sided) and to limit the estimation bias, the O'Brien-Fleming spending function will be used for this two-stage group sequential design. A significance level of 0.003 (2-sided) will be allocated to this unblinded interim analysis. The corresponding final significance level is 0.049 (2-sided).

Based on the primary analysis result at IA, the DMC will be allowed to make recommendations based on 1 of the 4 outcomes:

- Early stop for efficacy if the unblinded interim analysis p-value is < 0.003 (2-sided; Early Efficacy), and all point estimates of difference in change from baseline to Weeks 6, 8, and 10 in MADRS total score are above 1.6.
- Proceed as planned to 360 randomized subjects if the conditional power observed at the unblinded interim analysis is $\geq 85\%$; or is $> 15\%$ and $\leq 50\%$ (Favorable or Unfavorable Zone).
- Increase sample size from 360 randomized subjects to 540 randomized subjects, if the conditional power observed at the unblinded interim analysis is $> 50\%$ but $< 85\%$ (Promising Zone).
- Early stop for futility if the conditional power observed at the unblinded interim analysis is $\leq 15\%$ (Futility).

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, DeMets, and Lan

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

(2004)[13] will be used to increase the sample size based on an interim estimate of the treatment effect size, possibly combined with other external information, without inflating the type I error.

If the study is terminated after claiming efficacy in the interim analysis, a "final" analysis will be performed. This "final" analysis will include the efficacy data occurring between the data cutoff for the interim analysis and the final database lock (i.e. all efficacy data up to database lock), and the alpha level for the "final" analysis will be derived based on the information fraction using the Lan-DeMets Spending Function with O'Brien-Fleming boundary.

In this SAP, the statistical tests based on the primary efficacy endpoint and the key secondary efficacy endpoint will be done using a hierarchical approach, in order to control the overall type I error rate. In this hierarchical testing, the alpha-spending for the primary endpoint will be applied to the key secondary endpoint. Specifically, if the trial stops at the IA, the key secondary endpoint will be tested on the IA data at a two-sided 0.003 level.

If the trial does not stop at the IA, the above sequence of testing will be repeated for the final analysis with the use of a two-sided 0.049 significance level. That is, at the final analysis, the primary efficacy endpoint will be tested at a two-sided 0.049 significance level, and upon achieving significance on the primary endpoint, the key secondary endpoint will be tested at the same level. In that scenario, the final analysis will be conducted using the data of approximately 360 or 540 (if sample size is adapted) subjects. There will be no data accrual after the database lock for the final analysis.

9 Safety Analysis

Standard safety variables to be analyzed include AEs and clinical laboratory tests. In addition, suicidality will be evaluated through analysis of data from appropriate scales.

Analyses of the treatment period and extension period safety data will be performed on the Safety Sample unless indicated otherwise. In general, baseline of a safety variable is defined as the last observation of the variable prior to the start of double-blind digital therapeutic, unless specified otherwise.

9.1 Adverse Events

All adverse events (AEs) will be coded by system organ class (SOC) and Medical Dictionary for Regulatory Activities (MedDRA) Preferred Term (PT). The incidence of the following events in the treatment period and the extension period will be summarized by treatment group and overall:

a) Treatment-emergent AEs (TEAEs)

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

b) TEAEs by severity

c) TEAEs potentially causally related to the digital therapeutic (CT-152 or sham)

d) TEAEs with an outcome of death

e) Serious TEAEs

f) TEAEs leading to discontinuation of the digital therapeutic (CT-152 or sham)

Deaths, SAEs, and AEs leading to discontinuation from digital therapeutic or trial, will be listed by subject on the Safety Sample.

9.1.1 Adverse Events in the Treatment Period

TEAEs in the treatment period are defined as AEs with an onset date on or after the start of double-blind digital therapeutic. In more detail, TEAEs are all adverse events which started after the first use of double-blind digital therapeutic; or if the event was continuous from baseline and was worsening, serious, study digital therapeutic related, or resulted in death, discontinuation, interruption or reduction of study digital therapeutic. Adverse Events prior to the start of the extension period will be included in the summary tables.

The incidence of AEs in the treatment period will be tabulated by treatment group and overall on the Safety Sample. Incidence of TEAE during the treatment period of at least 5% in CT-152 group and also greater than sham by SOC and MedDRA PT will be provided.

9.1.2 Adverse Events in the Extension Period

TEAEs in the extension period are defined as AEs with an onset date during that period, or if the event was continuous from the treatment period and was worsening, serious, study digital therapeutic related, or resulted in death or discontinuation. AEs occurring through the last subject visit or contact in this period will be included in these summary tables.

9.2 Clinical Laboratory Tests

Listings of urine drugs-of-abuse screen and pregnancy test results by subject will be provided.

9.3 Suicidality Data

Suicidality will be monitored during the study using the C-SSRS and will be summarized as number and percentage of subjects reporting any suicidal behavior, ideation, behavior by type (4 types), ideation by type (5 types) and treatment emergent suicidal behavior and ideation. Summary will be provided for the treatment period and extension period on the Safety Sample.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

Suicidality is defined as report of at least one occurrence of any type of suicidal ideation or at least one occurrence of any type of suicidal behavior during assessment period (count each person only once).

Treatment emergent suicidal behavior and ideation is summarized by four types: Emergence of suicidal ideation, Emergence of serious suicidal ideation, Worsening of suicidal ideation, Emergence of suicidal behavior.

Emergence of suicidal behavior/ideation is defined as report of any type of suicidal behavior/ideation during treatment when there was no baseline suicidal behavior/ideation.

Emergence of serious suicidal ideation is defined as observation of suicidal ideation severity rating of 4 or 5 during treatment when there was no baseline suicidal ideation.

Worsening of suicidal ideation is defined as a suicidal ideation severity rating that is more severe than it was at baseline.

9.3.1 Clinically important suicidality

Clinically important suicidality may include:

1) Suicidal behavior (with or without intent of suicide or serious self-harm)
2) Acute suicidality to such a degree that precaution against suicide must be exercised.
3) Response of "yes" to C-SSRS ideation items 4 or 5 or any of the C-SSRS behavioral items.
4) Score of "4", "5", or "6" on the MADRS suicidal thoughts item.

1) and 2) are referred to suicidality-related AEs, which will be identified using Standardized MedDRA Query (SMQ) for suicidality.

The incidence of emergent clinically important suicidality will be compared between CT-152 and sham groups using Cochran-Mantel-Haenszel (CMH) General Association Test on OC dataset. Rates and 95% confidence intervals will be reported and the estimated differences between groups will be provided.

9.4 Other Safety Analysis

The incidence of worsening depressive symptoms during the treatment period will be compared between CT-152 and sham groups using Cochran-Mantel-Haenszel (CMH) General Association Test on OC dataset. Worsening depressive symptoms are defined as 1) CCI [redacted] 2) Any AE of worsening depression. Rates and their 95% confidence intervals will be reported and the estimated differences between groups will be provided.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

9.5 Concomitant Medications

Number and proportion of patients taking concomitant medications prior to the treatment period, during the treatment period, and during the extension period are tabulated by drug classification using the World Health Organization (WHO) drug dictionary on the Safety Sample.

9.6 Extent of Exposure

Duration of exposure is defined as the last day of intervention with double-blind digital therapeutic - the first day of intervention with double-blind digital therapeutic +1, during the treatment period. The number and percentage of subject who receive double-blind digital therapeutic, will be presented by week and by treatment group. Each treatment week will be based on the actual week, i.e., Day 1-7 in Week 1, Day 8-14 in Week 2, etc. Mean, range, and standard deviation of the duration of exposure will also be provided. This summary will be performed on the Safety Sample.

10 Conventions

10.1 Study Visit Windows

Study visit windows will be used to map visits using study day intervals. Observations at each scheduled visit and Early Termination will be assigned to Week 2, Week 4, Week 6, Week 8 and Week 10 based on their visit windows as shown in Table 10.1-1. This visit window convention applies to tables and listings for all efficacy scales (MADRS, GAD-7, CGI-S, PHQ-9, WHODAS2.0, CCI , CCI , and CCI). This derived study window variable will be named as WEEK and will be footnoted. In listings it will be listed along with the eCRF study visit.

Table 10.1-1 shows classifications for study day intervals in the treatment period and extension period. The variable "target day" is defined using the number of days since the start of intervention with double-blind digital therapeutic in the treatment period. The first day of intervention with double-blind digital therapeutic 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. Assessments done during the treatment period and occurring more than seven days after the last date of the interaction with double-blind digital therapeutic in the treatment period will be excluded from efficacy analyses of the treatment period.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

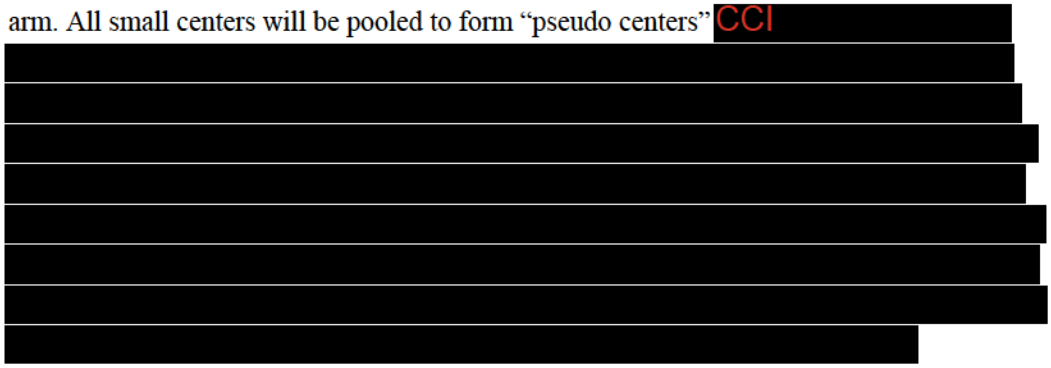
Table 10.1-1 Study Day and Visit Windows in the Trial

Week	Target Day[a]	Study Day Interval [a]
2	14	2~20
4	28	21~34
6	42	35~48
8	56	49-62
10	70	63-77

[a] Relative to the first day of intervention with double-blind digital therapeutic in the treatment period.

10.2 Pooling of small centers

Primary efficacy analysis will be performed on the FAS Sample which comprises those subjects in the Randomized Sample who have baseline and at least 1 postbaseline assessment of MADRS total score. Small centers will be defined as centers that do not have at least one evaluable subject (evaluable with regard to the primary efficacy variable) in each treatment arm. All small centers will be pooled to form "pseudo centers" CCI



10.3 Scales: Rules for Scoring and Handling of Missing Data

10.3.1 Montgomery-Asberg Depression Rating Scale (MADRS)

The MADRS is utilized as the primary efficacy assessment of a subject's level of depression. The MADRS consists of 10 items, all rated on a 0 to 6 scale with zero being the "best" rating and 6 being the "worst" rating. The MADRS total Score is the sum of ratings for all 10 items; therefore, possible total scores range from 0 to 60. The MADRS total score will be unevaluable if less than 8 of the 10 items are recorded. If 8 or 9 of the 10 items are recorded, the MADRS total Score will be the mean of the recorded items multiplied by 10 and then rounded to the first decimal place. MADRS is assessed at each scheduled visit.

10.3.2 Generalized Anxiety Disorder-7 (GAD-7)

The GAD-7 is a self-reported questionnaire designed to assess anxiety in subjects. The scale contains 7 items and each item is rated from 0 (not at all) to 3 (nearly every day). The total score ranges from 0 to 21. A higher score on the GAD-7 represents greater anxiety symptomatology. GAD-7 is assessed at each scheduled visit.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

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

The severity of agitation for each subject will be rated using the Clinical Global Impression-Severity of Illness Scale (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. CGI-S is assessed at each scheduled visit in the treatment period.

10.3.4 Patient Health Questinaire-9 (PHQ-9)

The PHQ-9 is a standardized, self-administered rating scale that assesses the severity of depressive symptoms. The scale consists of 9 items, representing the 9 criteria upon which the diagnosis of DSM-IV depressive disorders is based. Each item is rated from 0 (not at all) to 3 (nearly every day). The total score ranges from 0 to 27. A higher score on the PHQ-9 represents a higher severity of depressive symptoms. PHQ-9 is assessed at screening, and Weeks 4 and 6.

10.3.5 World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0)

The WHODAS 2.0 is a 36-item self-assessment scale to measure a subject's function and disability across 6 domains of life: cognition (understanding and communicating), mobility (moving and getting around), self-care (hygiene, dressing, eating, staying alone), getting along (interacting with others), life activities (domestic responsibilities, leisure, work and school), and participation (community and society). Each item is rated from 0 (none) to 4 (extreme or cannot do). The total score of the 36 items will be converted into a metric ranging from 0 (no-disability) to 100 (full- disability) by dividing the total score by the maximum possible value of the total score, which is 144 for 36 items and 128 for 32 items if items D5.5-D5.8 were skipped, and multiplying by 100. In situation where one or two items are missing, it will be handled as described in the manual of WHODAS 2.0[14], the mean score across all items within the domain should be assigned to the missing items. This method should not be used if more than two items are missing. In addition, if domain-wise scores are being computed for domains, the two missing items should not come from the same domain. WHODAS 2.0 is assessed at the baseline and Week 6.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

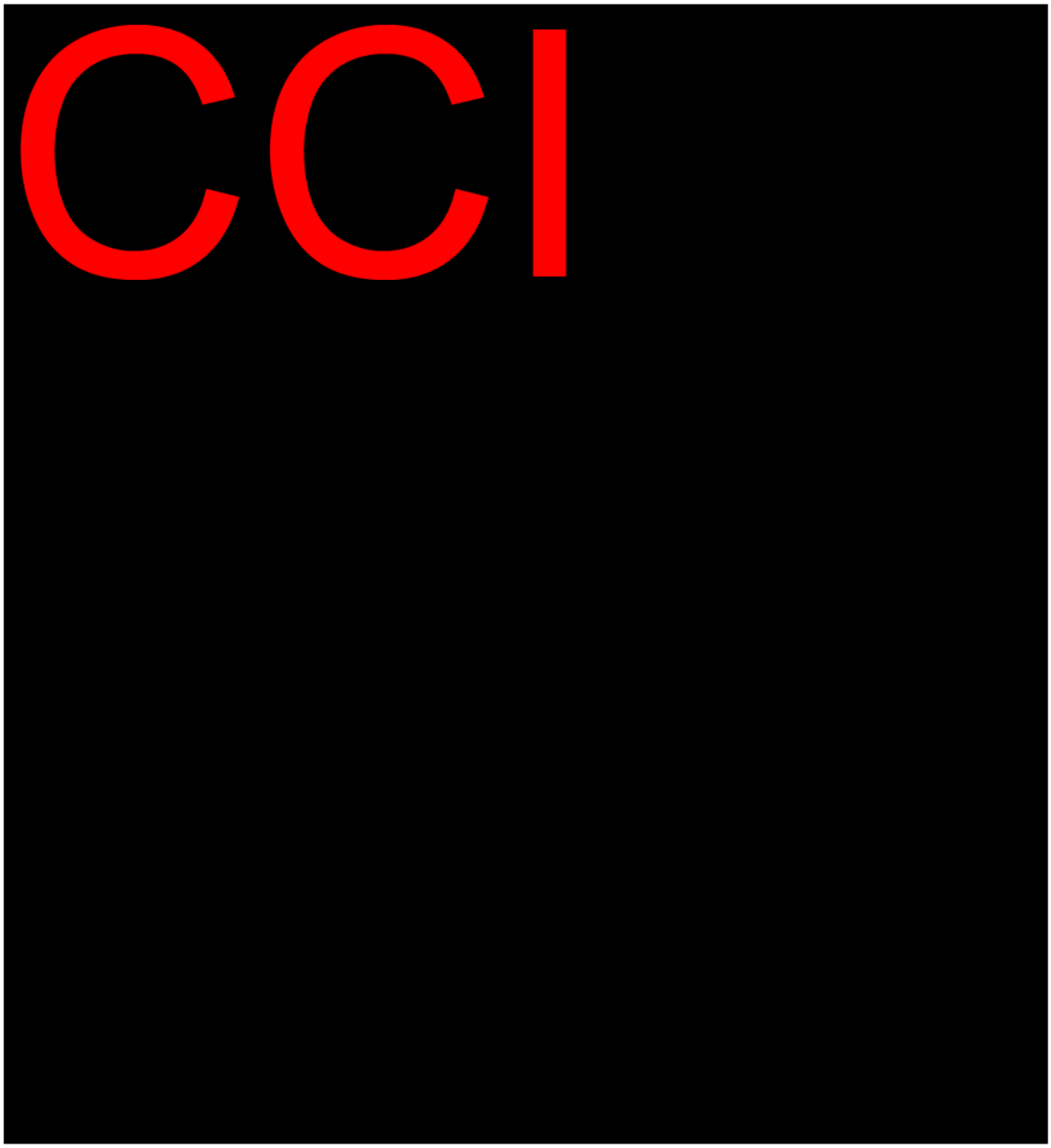
SAP 345-201-00002

10.3.9 Columbia-Suicide Severity Rating Scale (C-SSRS)

Suicidality will be monitored during the trial using the C-SSRS. This trial will use the "baseline/screening" and "Since Last Visit" versions of the scale. The "Baseline/Screening" Version will be used for the first administration (at the screening visit). This version will evaluate a lifetime assessment of suicidal ideations and behaviors, as well a past 12-month evaluation of suicidal ideations and a past 24-month evaluation of suicidal behaviors. Subsequent visits will utilize the "Since Last Visit" Version.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002



Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

CCI

CCI

CCI

CCI

CCI

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: 345-201-00002_Statistical_Analysis_Plan

Document Number: CCI [redacted]

Document Version: 3.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
PPD	Clinical Approval	12-Feb-2022 03:12:58
PPD	Biostatistics Approval	13-Feb-2022 22:20:28

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

Otsuka Pharmaceutical Development & Commercialization, Inc.

Digital Therapeutic

CT-152

Addendum to Statistical Analysis Plan Version 1.0

A Multi-center, Randomized, Controlled Trial to Evaluate the Effectiveness of a Digital Therapeutic (CT-152) as Adjunctive Therapy in Adult Subjects Diagnosed with Major Depressive Disorder

Protocol No. 345-201-00002

Version: 1.0

Date: 16 November 2022

CONFIDENTIAL - PROPRIETARY INFORMATION

Confidential
May not be used, divulged, published or otherwise disclosed
without the consent of Otsuka Pharmaceutical Development & Commercialization, Inc.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

Table of Contents

CCI [redacted]

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

1 Introduction

This addendum to the statistical analysis plan (SAP) version 1.0 is to:

- Remove the key secondary efficacy endpoint and corresponding multiplicity adjustment rule.
- Specify the efficacy endpoints for supportive analyses for the primary efficacy endpoint and for the durability of the effect.
- Modify the criteria for defining Per-protocol (PP) Analysis Sample.

CCI

After the approval of the original SAP (version 1.0, issued on February 13, 2022). The addition and changes will have no impact on the overall type-I error rate. The statistical methodology, data analysis algorithms, and conventions remain the same as those in the original SAP, unless being specified below.

2 Removing Key Secondary Efficacy Endpoint

The change from baseline to Week 6 in GAD-7 total score, mentioned in Section 8.2 in the original SAP, will be considered as a secondary efficacy endpoint rather than the key secondary efficacy endpoint in this study. The hierarchical testing approach used to control the overall type-I error rate on the primary efficacy endpoint and the key secondary efficacy endpoint will no longer be applicable.

To control the overall type I error for the primary efficacy endpoint, it will be tested at a significance level of 0.049 (2-sided) as planned. All rest efficacy endpoints, including secondary efficacy, other efficacy, and CCI , will be tested at a nominal 0.05 level (2-sided) individually without adjusting for multiplicity.

3 Supportive Analyses for the Primary Efficacy Endpoint

The following endpoints from Section 8.3 (Other Efficacy Endpoints) CCI (CCI) in the original SAP will be served as supporting evidence to the primary efficacy endpoint:

1) Change from baseline to Weeks 2 and 4 in the MADRS total score
2) MADRS full response at Weeks 2, 4, and 6
3) MADRS partial response at Weeks 2, 4, and 6
4) MADRS remission at Weeks 2, 4, and 6

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

SAP 345-201-00002

Note that in order to differentiate with MADRS partial response, MADRS response in the original SAP is renamed as MADRS full response.

4 Supportive Analyses for Assessment for Durability Based on MADRS Total Score

The following endpoints from Section 8.3 (Other Efficacy Endpoints)CCI [redacted] (CCI [redacted]) in the original SAP will be served as supporting evidence to the durability of effect assessed based on MADRS total score. They will be analyzed by the same statistical model as used for Endpoints (2) - (4) in Section 3 of this addendum.

1) MADRS full response at Weeks 8 and 10
2) MADRS partial response at Weeks 8 and 10
3) MADRS remission at Weeks 8 and 10

5 Modification on PP Analysis Sample

The criteria 3 and 4 in Section 5.1 in the original SAP will not be used for defining PP Analysis Sample.

The PP Analysis Sample will only be used for sensitivity analysis for the primary efficacy variable and the secondary efficacy variable.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

cci

cci

cci

cci

cci

cci

cci

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: 345-201-00002_SAP v1.0 Addendum

Document Number: CCI [redacted]

Document Version: 2.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
PPD	Biostatistics Approval	16-Nov-2022 17:48:10
PPD	Clinical Approval	16-Nov-2022 18:24:45

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

Otsuka Pharmaceutical Development & Commercialization, Inc.

Digital Therapeutic

CT-152

Protocol No. 345-201-00002

A Multi-center, Randomized, Controlled Trial to Evaluate the Effectiveness of a Digital Therapeutic (CT-152) as Adjunctive Therapy in Adult Subjects Diagnosed with Major Depressive Disorder

Interim Analysis Plan

Version: Final
Date: 13 Oct 2021

Confidential
May not be used, divulged, published or otherwise disclosed
without the consent of Otsuka Pharmaceutical Development & Commercialization, Inc.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

Protocol 345-201-00002 Interim Analysis Plan

Table of Contents

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

List of Appendices

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

List of Abbreviations and Definitions of Terms

Abbreviation	Definition
ADT	Antidepressant therapy
AE	Adverse event
ANCOVA	Analysis of Covariance
DMC	Data Monitoring Committee
DSM-5	*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
EFMT	Emotional Faces Memory Task
FAS	Full analysis set
GAD-7	Generalized Anxiety Disorder-7
IA	Interim analysis
IAP	Interim analysis plan
IDAG	Independent Data Analysis Group
IVRS	Interactive voice response system
IWRS	Interactive web response system
LOE	Lack of efficacy
LS	Least square
MADRS	Montgomery-Asberg Depression Rating Scale
MAR	Missing at random
MCAR	Missing completely at random
MDD	Major depression disorder
MI	Multiple imputation
MMRM	Mixed-effect model repeated measure
MNAR	Missing not at random
OC	Observed-case
OPDC	Otsuka Pharmaceutical Development & Commercialization, Inc.
PMM	Pattern mixture models
SAP	Statistical analysis plan
SMT	Shapes Memory Task
SMS	Short message service
UN	Unstructured

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

1 Introduction

This interim analysis plan (IAP) documents the statistical methodology, data analysis algorithms, and conventions to be applied in the interim analysis (IA) and reporting of efficacy data of Trial 345-201-00002.

The primary objective is to compare the effectiveness of CT-152 with sham, in adult subjects diagnosed with major depression disorder (MDD) who are on antidepressant therapy (ADT) monotherapy.

The safety objective is to evaluate the safety of CT-152 in adult subjects diagnosed with MDD who are on ADT monotherapy.

This is a multi-center, randomized, controlled trial to evaluate the effectiveness of CT-152 in adult subjects diagnosed with MDD who are on ADT monotherapy for the treatment of depression. The trial population will include male or female subjects aged 22 to 64 years old at the time of informed consent, with a current primary diagnosis of MDD based on the criteria in the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (*DSM-5*), single or recurrent episode, without psychotic features and do not meet criteria for MDD with mixed features subtype, and who are on ADT monotherapy.

Subjects will participate in the trial for up to 13 weeks. The trial will include a screening period of up to 3 weeks, a treatment period for 6 weeks, and an extension period for 4 weeks. Eligible subjects will be randomized to 1 of 2 digital mobile applications (CT-152 or sham) on Day 1. During the treatment period (Day 1 [baseline] to Week 6), subjects will have a remote visit at Weeks 2, 4, and 6 and will be contacted by telephone at Weeks 1, 3, and 5. Subjects will be expected to be adherent with the recommended schedule of their digital mobile application exercises during the treatment period. After Week 6, subjects will continue participation in the trial during the extension period (Weeks 7 to 10). In the extension period, the digital mobile applications will remain installed for each group, with the Emotional Faces Memory Task (EFMT) and Shapes Memory Task (SMT) no longer available. Psychotherapy content provided previously will remain available for optional reference in the CT-152 group but no new therapeutic content will be introduced and no required treatment schedule is in place. The 2 groups will each receive brief short message service (SMS) messages in the extension period reminding subjects of the previously completed CT-152 or sham treatment courses, and will continue their ADT. Subjects will have a remote visit at Weeks 8 and 10 and will be contacted by telephone at Weeks 7 and 9. The end of the trial will be Week 10.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

2 Statistical Methods

2.1 Sample Size and Power Justification

The primary efficacy endpoint is the change from baseline at Week 6 in Montgomery-Asberg Depression Rating Scale (MADRS) total Score. The trial will compare CT-152 to Sham, randomized at a ratio of 1:1, with an overall alpha of 0.05 for the primary efficacy endpoint.

CCI [redacted] In this trial, we plan to detect a 3-point difference on the primary efficacy endpoint between the treatment groups. The 3-point treatment difference is above the CCI range. Therefore, the initial sample size is calculated to detect a 3-point difference between CT-152+ADT and Sham+ADT in the change from baseline in MADRS total score with 85% power at a 2-sided $\alpha = 0.05$ level, assuming a common standard deviation of 9. The resulting sample size is 324 evaluable subjects in total (162 subjects in each arm). To compensate for subjects that fail to have evaluable assessments of MADRS total score in the full analysis set (FAS) Sample (estimated at up to 10% of all subjects), a total of 360 subjects (180 subjects in each arm) will be randomized in this trial.

Due to the limitations of applying assumptions on the treatment effect size, and in order to ensure adequate power of the trial, an unblinded IA will be conducted by a Data Monitoring Committee (DMC) on approximately the first 180 randomized subjects who have either completed the Week 6 visit or discontinued prior to Week 6. The final sample size could be increased to 540 subjects (270 subjects in each arm) as per recommendation of the DMC. Using the O'Brien-Fleming boundaries[3], an alpha level of 0.003 (2-sided) is allocated to this interim analysis. The alpha level left for the final analysis is 0.049 (2-sided). The power and sample size were obtained using the PASS 14 (2015) statistical computing software.

2.2 Interim Efficacy Analysis

The interim analysis of efficacy data will be performed by an Independent Data Analysis Group (IDAG), the Labcorp DMC Support Team, in coordination with DMC. Reference is made to the DMC charter for details of its roles and responsibilities. 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 additional subjects are needed to have sufficient power to achieve the primary objective. To this end, the primary efficacy endpoint of change from the baseline to Week 6 in MARDS total score and the durability of effect of

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

CT-152 based on MADRS total score will be analyzed. The analysis for the key secondary efficacy endpoint of the change from the baseline to Week 6 in Generalized Anxiety Disorder-7 (GAD-7) total score and the durability of effect of CT-152 based on GAD-7 total score will also be provided.

The primary analysis will be performed on the FAS Sample which includes all randomized subjects who receive at least 1 occurrence of either CT-152 or sham use, and have both a baseline and at least one postbaseline assessment of MADRS total score. The primary efficacy analysis will be performed by fitting a mixed model repeated measurements (MMRM) analysis with an unstructured (UN) variance covariance structure, in which the change from baseline in MADRS total score (at Weeks 2, 4, and 6) during treatment period will be the dependent variable based on the OC data set. The model will include fixed class effect terms for treatment, visit, treatment by visit interaction, and site to assess heterogeneity of treatment effects. The model will also include an interaction term of visit by baseline MADRS total score as a covariate. The primary comparison between the CT-152+ADT and Sham +ADT at Week 6 will be estimated as the difference between Least Squares (LS) means utilizing the computing software SAS procedure PROC MIXED. In case there is a convergence problem with MMRM model with the unstructured (UN) variance covariance matrix, the following structures other than unstructured will be used in order of 1) heterogeneous toeplitz (TOEPH), 2) heterogeneous autoregressive of order 1 (ARH1), and 3) heterogeneous compound symmetry (CSH) and the first (co)variance structure converging to the best fit will be used as the primary analysis. If a structured covariance has to be used, the empirical "sandwich" estimator of the standard error of the fixed effects parameters will be used in order to deal with possible model misspecification of the covariance matrix.

To assess the durability of effect of CT-152, three MADRS assessments at Weeks 6, 8, and 10 will be included. CCI [redacted] at Week 6 with statistical significance, durability will be demonstrated by a point estimate of the difference in change from baseline at Weeks 8 and 10 above 1.6, when comparing CT-152 and sham. Such point estimates will be provided using the same MMRM model in the primary efficacy analysis with change from baseline in MADRS total score at Week 8 and 10 as the dependent variable.

This key secondary efficacy endpoint will be analyzed by fitting the same MMRM model as described in the primary analysis with change from baseline in GAD-7 total score (at Weeks 2, 4, and 6) as the dependent variable.

Durability based on GAD-7 will also include 3 assessments, at Weeks 6, 8, and 10, demonstrating a numerically larger improvement on point estimate of the difference in

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

change from baseline in GAD-7 total score at Weeks 8 and 10 in CT-152 compared to sham. Such point estimates will be provided using the same MMRM model in the key secondary efficacy analysis with change from baseline in GAD-7 total score at Week 8 and 10 as the dependent variable.

One unblinded IA for assessment of efficacy will be conducted on approximately the first 180 (50%) randomized subjects. This unblinded IA will be carried out when these subjects have either completed the Week 6 visit or discontinued prior to Week 6. The O'Brien-Fleming boundaries[3] will be utilized in primary analysis for the rejection of the null hypothesis to maintain an overall nominal significance level of 0.05 (2-sided). The boundary values for the rejection of the null hypothesis are given in Table 2.2-1. Also, to control the overall type I error, the statistical tests based on the primary efficacy endpoint and the key secondary efficacy endpoint will be done using a hierarchical approach. The primary hypothesis will be tested at the 0.003 level at IA and if successful in rejecting the null hypothesis, analysis of the key secondary endpoint will be tested at the type I error level of 0.003. Similar rules will be carried out at the final analysis with the 0.049 level if the trial doesn't stop at IA. Should the primary outcome not achieve statistical significance, the key secondary endpoint will be categorized as other efficacy endpoint.

Table 2.2-1 Boundaries for Rejection of the Null Hypothesis for the Interim and Final Analyses

Analysis	Two-sided Alpha Level[1]
Interim analysis #1 (~ 50%)	0.003
Final analysis (100%)	0.049

[1] Obtained using PASS 14(2015) sample size software.

The fraction of the sample size might not be consistent with the fraction of the information when the primary analysis is MMRM with high dropout rate. If there is a high dropout rate, the timing for the IA will use the rate of those subjects who complete the Week 6 assessment which would control the Type I error with the alpha spending function.

The IA results will be reviewed by the DMC. Should the interim analysis be performed at a different time, the Lan-DeMets Spending Function[4] with O'Brien-Fleming boundary will be applied to adjust the critical values in order to guarantee an overall 0.05 alpha level.

With the interim data, the DMC will be allowed to make recommendations to stop the trial for efficacy, stop the trial for futility, or continue the trial with/without sample size adjustment, based on the p-value and the conditional power calculated on the primary efficacy endpoint. The details are listed Table 2.2-2. The sample size will be re-estimated

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

only based on the conditional power determined at the interim analysis. The adaptive designs methodology published by Chen, DeMets, and Lan (2004)[5] will be used to increase the sample size based on an interim estimate of the treatment effect size of the primary efficacy endpoint, possibly combined with other external information, without inflating the type I error.

Table 2.2-2 Interim Results and DMC Recommendation

Interim Results	Interim Analysis Recommendation	Sample Size
p-value < 0.003 and all point estimates of difference in change from baseline to Week 6, 8, and 10 in MADRS total score are above 1.6.	Stop for efficacy	N=180 (90/arm)
Conditional power[1] ≥ 85%	No adaption to the planned sample size	N=360 (180/arm)
50% < Conditional power[1] < 85%	Increase sample size up to 540	N=540 (270/arm)
15% < Conditional power[1] ≤50%	No adaption to the planned sample size	N=360 (180/arm)
Conditional power[1] ≤ 15%	Stop for futility	N=180 (90/arm)

[1] The conditional power will be only based on the statistical test for the primary efficacy endpoint (ie, the change from baseline to Week 6 in MADRS total score).

If the study is terminated after claiming efficacy in the interim analysis, a "final" analysis will be performed. This "final" analysis will include the efficacy data occurring between the data cutoff for the interim analysis and the final database lock (i.e. all efficacy data up to database lock), and the alpha level for the "final" analysis will be derived based on the information fraction using the Lan-DeMets Spending Function[4] with O'Brien-Fleming boundary. In order to make sure that the "final" analysis will be positive without doubt, criteria more conservative than the one provided in Table 2.2-1 are needed for making an efficacy claim at an interim analysis.

The sponsor will remain blinded to any modification of the primary model due to non-convergence and should not be informed of any additional analysis or modification requested by the DMC and performed by unblinded statistician.

2.2.1 Sensitivity Analyses for Missing at Random (MAR) Assumption

Traditionally the dropout mechanisms are divided into three types (Little, 1995) [6]: (1) Missing Completely at Random (MCAR), in which the probability of dropout doesn't depend on the observed data and the missing data; (2) Missing at Random (MAR), in which the probability of dropout depends on the observed data, and (3) Missing Not at

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

Random (MNAR), where the probability of dropout depends on the missing data and possibly the observed data.

Most of MNAR methods (Diggle P, Kenward MG, 1994) [7] have treated all observations with dropout as if they fall within the same dropout type. In practice, we would find that different dropout reasons may be related to the outcomes in different ways, for example, detailed dropout reasons for this study are: adverse events (AE), lack of efficacy (LOE), lost to follow-up, protocol deviation, sponsor discontinued study, subject met (protocol specified) withdrawal criteria, subject was withdrawn from participation by the investigator, and subject withdrew consent to participate. Dropout due to an AE and LOE may lead to MNAR dropout. Subject withdrew consent may also lead to MNAR dropout. However, it is debatable whether a dropout caused by subjects withdrew consent is MAR or MNAR. Except AE, LOE, and subject withdrew consent, all the other dropout reasons may be assumed as either MCAR or MAR dropout.

As sensitivity analyses for missing at random (MAR) assumption, analyses for missing not at random (MNAR) will be carried out. Pattern Mixture Models (PMM) based on Multiple Imputation (MI) with mixed missing data mechanisms will be used to investigate the response profile of dropout patients by last dropout reason under MNAR mechanism for the following three scenarios:

1) Dropout reasons due to either AE or LOE as MNAR
2) Dropout reasons due to either AE or LOE or subject withdrew consent as MNAR
3) All dropouts as MNAR

Delta Adjustment Imputation Methods

This MNAR sensitivity analysis is to departure from MAR assumption by progressively increasing the delta until conclusion from the primary analysis is overturned. The delta is 0%, 10%, 20%, 30%, ..., 100% of the expected treatment difference of 3 points and/or the observed treatment difference between CT-152 and Sham from the primary analysis of MMRM model until conclusion of the primary analysis is overturned. When delta=0 it is MAR. When delta > 0 it is MNAR.

1) Using Monte Carlo Markov Chain (MCMC) methodology from PROC MI to impute the intermittent missing data to a monotone missing pattern
2) Using a standard MAR-based multiple imputation approach from PROC MI to impute the monotone missingness data

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

3) For patients in the treated group and with a dropout reason of AE or LOE or subject withdrew consent, a delta will be added for all the values after the dropout time

4) Using analysis of covariance (ANCOVA) model to analyze the completed data through multiple imputation

5) Obtaining the overall results using PROC MIANALYZE

The DMC may make recommendation by taking into account the sensitivity analysis result. A decision of stopping trial for efficacy might be made if the p-value of the sensitivity analysis for the primary efficacy endpoint using Delta Adjustment Imputation Method under 100% penalty for all dropouts in CT-152 is < 0.05.

3 Data Analysis

All interim efficacy monitoring will be performed by the DMC. Reference is made to the DMC Charter for details of constituents and roles and responsibilities of DMC. In addition, an IDAG has been formed. The IDAG will work closely with the DMC and will assist the DMC in the retrieval and analysis of efficacy data. The DMC will follow rules delineated in the IAP and the DMC charter in disposing its responsibilities.

3.1 Planned Interim Analyses of Unblinded Efficacy data

A set of pre-specified tables (see Appendix 1) will be produced at the planned interim analysis following statistical methods described in the Trial 345-201-00002 SAP. The Otsuka Pharmaceutical Development & Commercialization, Inc. (OPDC) Biometrics will provide blinded SAS datasets and SAS programming codes to the IDAG statistician. The OPDC Biometrics will also provide authorization for the IDAG statistician to get the randomization directly from IVRS or IWRS vendor. The IDAG will produce all pre-specified summary tables using the SAS codes and analysis datasets provided by OPDC Biometrics.

4 Confidentiality of Results

Except for the situation when the DMC makes a determination of stopping the trial because of compelling efficacy reasons, all deliberations, results and findings of the DMC and the IDAG will remain confidential within their jurisdictions, and the Sponsor and Investigators will remain blinded of any or all findings of unblinded analysis of efficacy data.

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

5 References

1 Duru G, Fantino B. The clinical relevance of changes in the Montgomery- Asberg Depression Rating Scale using the minimum clinically important difference approach. Curr Med Res Opin. 2008;24(5):1329-1335.

2 Masson SC, Tejani AM. Minimum clinically important differences identified for commonly used depression rating scales. J Clin Epidemiol. 2013;66(7):802-807.

3 O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35(3):549-56.

4 Lan KK, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika*. 1983;70(3):659-63.

5 Chen YH, DeMets DL, Lan KK. Increasing the sample size when the unblinded interim result is promising. *Stat Med*. 2004, 23(7) 1023-1038.

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

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

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

CCI

CCI

CCI

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: 345-201-00002_InterimAnalysisPlan

Document Number: CCI

Document Version: 3.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
PPD	Biostatistics Approval	14-Oct-2021 21:18:09
PPD	Clinical Approval	15-Oct-2021 18:29:44

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

CCI

[redacted]

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.

CCI

[redacted]

Until the information herein is released by Otsuka to the public domain, the contents of this document are Otsuka confidential information and should not be duplicated or re-distributed without prior written consent of Otsuka.