Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational New Drug
Centanafadine Sustained Release Tablets (EB-1020)

Protocol No. 405-201-00056
IND No. 155,243

A Phase 2, Multicenter, Randomized, Double-blind, Placebo-controlled Trial to Assess the Efficacy, Safety, and Tolerability of Centanafadine Sustained-release Tablets After Oral Administration in Adult Subjects with Binge Eating Disorder

A Trial of Centanafadine Efficacy, Safety, and Tolerability in Adult Subjects with Binge Eating Disorder

Statistical Analysis Plan

Version: Final
Date: 6 September 2022

Protocol Date: 24 Jan 2022

Table of Contents

List of In-text Figures

List of In-text Tables

List of Appendices

1 Introduction

This statistical analysis plan (SAP) describes the statistical methodology and data analysis algorithms and conventions to be applied for statistical analysis and reporting of data collected under the clinical protocol 405-201-00056. All amendments to the protocol are taken into consideration in developing the SAP, which is consistent with latest version of protocol amendment 2 dated on 24 Jan 2022.

2 Trial Objectives

Primary objective is to assess the efficacy of 2 doses of centanafadine sustained release (SR) (ie, 200 and 400 mg total daily dose [TDD]) compared with placebo in adults with moderate to severe Binge Eating Disorder (BED).

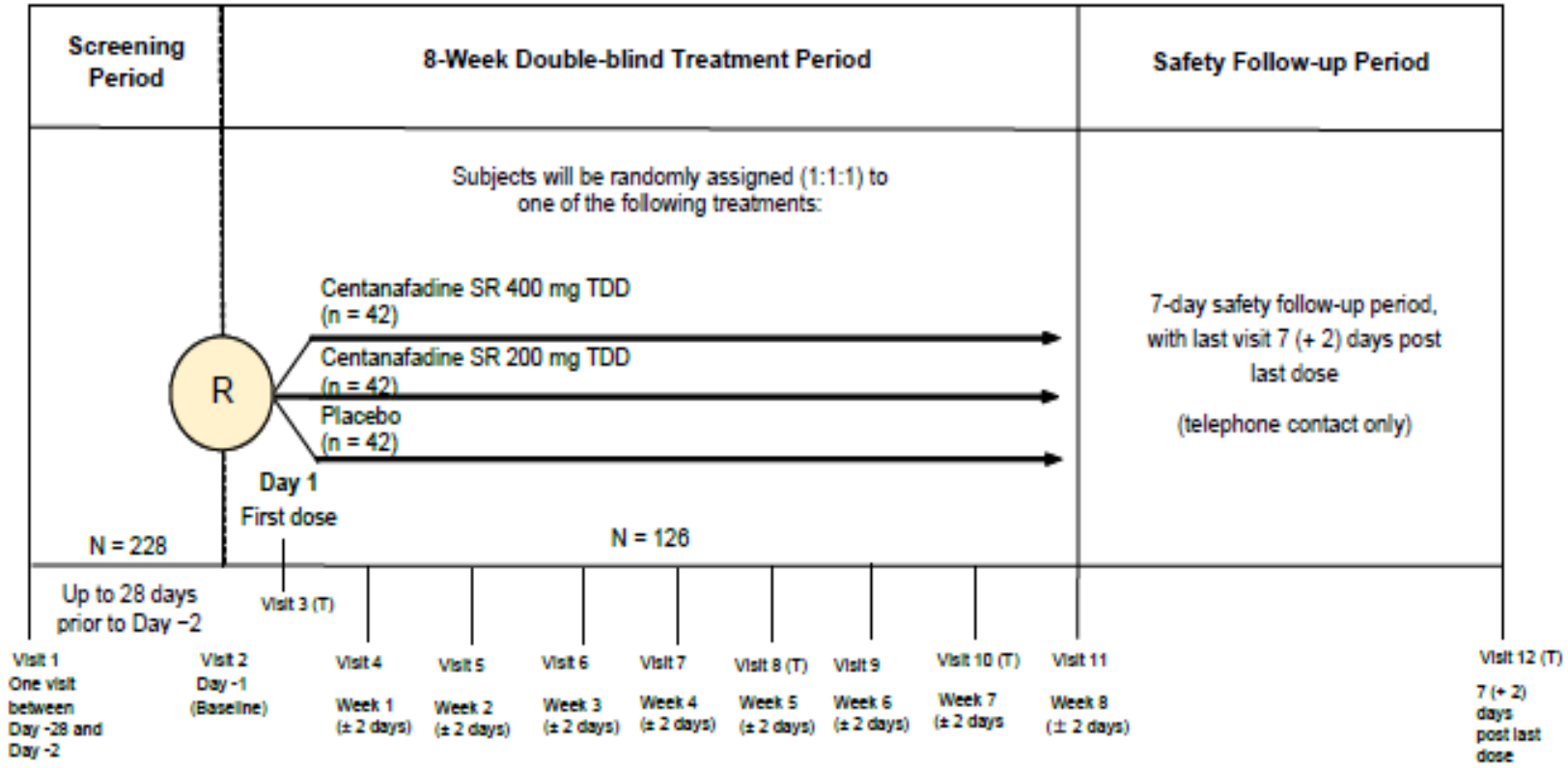
Secondary objective is to assess the safety and tolerability of centanafadine SR tablets administered daily (200 and 400 mg TDD) in the treatment of subjects with BED.

3 Trial Design

3.1 Type/Design of Trial

This is a phase 2, multicenter, randomized, double-blind, placebo-controlled trial to assess the efficacy, safety, and tolerability of centanafadine SR tablets for the treatment of adult subjects with BED. Subjects with a primary clinical diagnosis of moderate to severe BED based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) diagnosis criteria and confirmed by the Structured Clinical Interview for DSM-5 Disorders, Clinical Trials Version (SCID-5-CT) is eligible for enrollment. The trial consists of a screening period of up to 28 days that includes completion of a binge eating diary for at least 14 days (that shows at least 3 binge eating days in each week during the previous 14 days of a subject's screening period as reviewed at baseline [minimum screening period is 2 weeks]), a baseline visit, an 8-week (56-day) double-blind treatment period, and a 7-day safety follow-up period. All baseline assessments are conducted prior to first dose of investigational medicinal product (IMP). The design scheme is summarized in Figure 3.1-1.

Eligible subjects are randomized 1:1:1 at baseline to receive 1 of 2 doses of centanafadine SR or placebo daily for 8 weeks. All subjects who terminate early are required to participate in a 7-day safety follow-up period, which consists of telephone contact 7 (+ 2) days after the last dose of IMP. For subjects who early terminate, they are instructed to refrain from using prohibited concomitant medications, including BED treatments, until after the safety follow-up visit.



N = number of subjects; R = randomization; T = telephone visit.

Figure 3.1-1 Trial Design Schematic

3.2 Trial Treatments

Eligible subjects are randomized in a 1:1:1 ratio at baseline to receive 1 of 2 doses of centanafadine SR or placebo daily for 8 weeks:

- High dose centanafadine: 400 mg centanafadine SR oral tablets TDD (2 × 100 mg BID, approximately 4 - 6 hours apart)
- Low dose centanafadine: 200 mg centanafadine SR oral tablets TDD (1 × 100 mg) + 1 placebo oral tablet (BID, approximately 4 - 6 hours apart)
- Placebo: 2 placebo oral tablets (BID, approximately 4 - 6 hours apart)

All doses of centanafadine SR and matching placebo should be taken orally BID (with up to 240 mL of water). Doses should be taken at approximately the same time each day, with the first dose taken in the morning and the second dose taken 4 to 6 hours after the morning dose is administered. All doses can be administered without regard to meals.

If a subject forgets to take their morning dose at their normal dosing time, the subject should take the dose as soon as they remember that same day. If a subject forgets to take their second dose and 6 hours have elapsed since the time of the morning dose, they should skip the dose and resume dosing the next day. If a subject does not remember to take a missed dose on the same day it is missed, the subject should skip that dose and resume the normal dosing schedule for the following dose. Subjects should not take more than 2 doses in one day.

3.3 Trial Population

Subjects are 18 to 65 years of age (inclusive), with a primary diagnosis of BED of at least moderate severity, with a history of at least 2 binge eating days per week for 6 months prior to screening and at least 3 binge eating days in each week during the previous 14 days of a subject's screening period according to the subject's binge diary as reviewed at baseline, a rating of 4 or higher on the Clinical Global Impression - Severity (CGI-S) at screening and baseline, and a body mass index (BMI) of 18 to 45 kg/m^2, inclusive. Subjects with a lifetime history or current diagnosis of bulimia nervosa or anorexia nervosa are excluded.

Approximately 228 adult subjects with BED are anticipated to be screened with the expectation that approximately 126 eligible subjects are to be randomized (42 subjects per treatment arm). The randomization is stratified by trial site.

4 Sample Size

Assuming a treatment difference of 1.1 points with a standard deviation of 1.7 in the mean change from baseline in binge eating days per week at Weeks 7-8 in either centanafadine dose group (high dose centanafadine or low dose centanafadine) compared to the placebo group (effect size = 0.65), a sample size of 114 evaluable subjects (38 in each treatment arm) yield at least 80% power to detect the treatment effects at a 2-tailed significance level of 0.05. Approximately 126 subjects are randomized to ensure 114 evaluable subjects in the study. The power and sample size were obtained using the PASS 14 (2015) statistical computing software.

5 Statistical Analysis Datasets

The following datasets are defined for this trial:

- Randomized Analysis Set comprises all subjects who are randomized in the double-blind treatment period. Subjects are considered randomized subjects when they are assigned a treatment group at the baseline visit.
- Safety Analysis Set comprises those randomized subjects in the double-blind treatment period who receive at least 1 dose of double-blind IMP. Subjects will only be excluded from this population if there is documented evidence (ie, numbers of drug dispensed are equal to numbers of drug returned, or no IMP dispensed at all) that the subject did not take IMP. If a subject is dispensed IMP and is lost to follow-up, he/she is considered as exposure to IMP.
- Efficacy Analysis Set comprises all subjects in the Safety Analysis Set who have a baseline value and at least 1 valid post-randomization efficacy evaluation (i.e., either binge eating diary data or CGI-S) in the double-blind treatment period.
- Per Protocol Analysis Set comprises those subjects in the efficacy analysis set who complete at least first 2 weeks of double-blind IMP (i.e., [last day of IMP - first day of IMP + 1] $\geq$ 14 days) and have at least one post baseline primary efficacy endpoint on or after week 2 without major protocol violations deemed to compromise the assessment of efficacy. These major protocol violations will be any of the followings:

    1) Subjects who are less than 80% or more than 120% compliant with double-blind IMP or miss at least 7 consecutive days of IMP prior to last date of binge dairy collection.
    2) Subjects who report concomitant medication use which impact the primary efficacy endpoint.
    3) Subjects who have major protocol deviations which impact assessment of primary efficacy endpoint.
    4) Subjects who receive the actual treatment which is inconsistent with the treatment assigned based on the randomization schedule.

The core dataset for all efficacy analyses is the Efficacy Analysis Set, which is created based on the intent-to-treat (ITT) principle. In order to handle missing data and restrictions imposed by different types of analyses (e.g., change from baseline analysis), datasets derived from Efficacy Analysis Set are used for the efficacy analysis. Safety analysis set is mainly used for safety analysis.

5.1 Baseline and Last Visit Definition

In general, baseline is defined as the last available measurement prior to the first dose of IMP in the double-blind treatment period. Last visit is defined as the last available measurement by completion or termination in the double-blind treatment period.

However, baseline for binge eating days per week is the number of confirmed binge eating days divided by the number of non-missing diary days within 14 days prior to randomization, and multiplied by 7. Similarly, baseline for number of binge episodes per week is the number of confirmed binge episodes (by principal investigator) divided by the number of non-missing diary days within 14 days prior to randomization, multiplied by 7.

5.2 Mapped Trial Week Windows

Trial week windows are used to map the scheduled visits using trial day intervals. Trial day is calculated as: trial day = date of assessment − date of 1st IMP administration + 1. The visit window convention in Table 5.2-1 applies to analysis for the efficacy variables regarding episodes of binge eating, which are originally collected from subjects' diary cards.

Table 5.2-1 Mapped Trial Week Windows for Counting Days or Episodes of Binge Eating

Scheduled Visit	Mapped Week	Trial Day Interval [a]
Visit 2	Baseline	−14 ~ −1
Visit 4	Week 1	1 ~ 7
Visit 5	Week 2	8 ~ 14
Visit 6	Week 3	15 ~ 21
Visit 7	Week 4	22 ~ 28
Visit 9	Week 5	29 ~ 35
Visit 9	Week 6	36 ~ 42
Visit 11	Week 7	43 ~ 49
Visit 11	Week 8	50 ~ 59 [b]

Note: Day 0 is scheduled as randomization date per protocol.
a Relative to the first day of double-blind IMP in the double-blind treatment period.
b Evaluations occurring more than three days after the last dosing date and evaluations occurring during the follow-up period will be excluded from analyses.

The visit window convention in Table 5.2-2 applies to analysis for other efficacy variables, body weight and BMI. If more than one observation falls within a particular trial day interval, then the last observation within that interval is used.

Table 5.2-2 Mapped Trial Week Windows for Other Efficacy Analysis

Scheduled Visit	Mapped Week	Trial Day Interval [a]
Visit 4	Week 1	1 ~ 10
Visit 5	Week 2	11 ~ 17
Visit 6	Week 3	18 ~ 24
Visit 7	Week 4	25 ~ 35
Visit 9	Week 6	36 ~ 49
Visit 11	Week 8	50 ~ 63 [b]

Note: Day 0 is scheduled as randomization date per protocol.
a Relative to the first day of double-blind IMP in the double-blind treatment period.
b Evaluations occurring more than three days after the last double-blind dosing date and evaluations occurring during the follow-up period will be excluded from analyses.

5.3 Handling of Missing Data

5.3.1 Missing Data due to Discontinuation

In general, missing data will be handled by analysis of mixed-effect model repeated measures (MMRM) methodology based on observed-case data from scheduled visits under the assumption of missing at random. MMRM assume data are missing at random (MAR), which is a reasonable assumption in longitudinal clinical trials[1] in the binge eating disorder population. However, the possibility of "missing not at random" (MNAR) data can never be ruled out. CCI [redacted].

CCI [redacted].

The observed-cases (OC) dataset will consist of actual observations recorded at each visit during the double-blind treatment period and no missing data will be imputed. MMRM,

model-based imputation model, 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 or, if no observation is recorded at that visit, data carried forward from the previous visit. Baseline data will not be carried forward to impute missing values for the LOCF data set. For categorical efficacy variables, OC analyses will be performed in addition to LOCF analyses.

5.3.2 Intermittent Missing Episode Data due to Incomplete Diary

All binge episodes are captured daily by the subject in a binge eating diary. Details about binge eating episodes per day are recorded. At each visit, the investigator or designee review the completed diary with the subject and assess the number of binges for each day. The investigator assessment of number of binges for each day are recorded in the electronic case report form (eCRF).

Key opinions leaders indicated that intermittent missing episode data due to incomplete diary is rare in a well-designed clinical trial setting. No evidence is shown missing diary data is likely linked to worsening BED. Missing diary for recording binge eating episodes for BED population often results from subjects without binge episode forget to log in the diary cards. It is reasonable to assume MAR for missing diary data. Number of episodes on a missing-diary day will be generally considered as average of number of daily episodes based on non-missing diary within the same week. Details are provided in Section 8.2.3.1. In addition, the missing diary data pattern will be investigated as needed.

6 Primary and Secondary Outcome Variables:

6.1 Primary Outcome Variables

Primary efficacy endpoint is change from baseline in binge eating days per week at Week 7-8, having binge eating day defined as a day with at least one confirmed binge eating episode.

6.2 Secondary Outcome Variables

Key secondary efficacy endpoint is change from baseline in Clinical Global Impression Severity (CGI-S) score at Week 8.

Secondary safety endpoints include assessments in adverse events (AEs), clinical laboratory tests (hematology, serum chemistry and urinalysis), vital sign measurements, 12-lead electrocardiograms (ECGs), Study Medication Withdrawal Questionnaire

(SMWQ), Hamilton Anxiety Rating Scale (HAM-A), Montgomery-Asberg Depression Rating Scale (MADRS), and the Columbia Suicide Severity Rating Scale (C-SSRS).

6.3 Other Outcome Variables

Other efficacy endpoints include

- change from baseline in CGI-S score at Weeks 1, 2, 3, 4, and 6,
- Clinical Global Impression - Change (CGI-C) score,
- change from baseline in Yale-Brown Obsessive-Compulsive Scale Modified for Binge Eating (Y-BOCS-BE) score,
- proportion of subjects with four-week cessation from binge at Week 8/ET,
- change from baseline in number of binge episodes per week,
- change from baseline in Patient Global Impression - Severity (PGI-S) score,
- Patient Global Impression - Change (PGI-C) score,
- change from baseline in 36-item Short-form Health Survey Version 2 (SF-36v2) physical component summary (PCS) and mental component summary (MCS) scores
- change from baseline in Eating Disorder Examination Questionnaire - 7-Item Version (EDE-Q7) global score.

7 Disposition and Demographic Analysis

7.1 Subject Disposition

Subject disposition will be summarized for the randomized analysis set by treatment group in addition to by center. Subjects who are evaluated at the last scheduled visit during the double-blind treatment period (Week 8) will be defined as trial completers. Subject completion rate and reasons for discontinuation will be summarized for the randomized analysis set by treatment group.

7.2 Demographic and Baseline Characteristics

For the randomized analysis set, demographic characteristics will be summarized for the randomized analysis set by treatment group and overall, including age, gender, race, ethnicity, height, weight, and body mass index (BMI) at baseline. BMI will also be classified as Normal/Underweight (<25 kg/m^2), Overweight (25 kg/m^2 to <30 kg/m^2), Obesity Class I (30 kg/m^2 to <35 kg/m^2), Obesity Class II (35 kg/m^2 to <40 kg/m^2), Obesity Class III (≥ 40 kg/m^2). 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 Baseline Disease Evaluation

Summary of medical and psychiatric history will be presented for the randomized analysis set by treatment group and overall.

Binge eating days and number of binge eating episodes at baseline will be summarized for the randomized analysis set by treatment group and overall.

Descriptive statistics for efficacy scales at baseline will be provided for the randomized analysis set by treatment group and overall, including binge eating days per week, binge eating episodes per week, CGI-S score, Y-BOCS-BE total score and subscale scores for obsession and compulsion, PGI-S score, EDE-Q7 global score and subscale scores for dietary restraint, shape/weight overvaluation, and body dissatisfaction, SF-36v2 PCS and MCS scores, HAM-A total score, and MADRS total score.

7.4 Treatment Compliance

Based on the IMP panel of the eCRF, compliance of taking IMP is calculated by dividing the number of tablets taken by the total number of tablets the subjects are scheduled to take during the trial period. For lost-to-follow up subjects, last IMP end date record are used as the treatment end date. Compliance rate will be summarized for the randomized analysis set by treatment group in categories of < 80%, 80% to <90%, 90% to <100%, ≥100% etc.

7.5 Prior and Concomitant Medications

Number and proportion of subjects taking concomitant medications will be presented for randomized analysis set by treatment group and drug classification using the World Health Organization (WHO) drug dictionary for the following three intervals: prior to start of double-blind treatment period, during double-blind treatment period, and post study therapy.

7.6 Protocol Deviations

Major protocol deviations will be summarized for randomized analysis set by center and type of deviation for randomized subjects in addition to by treatment group and type of deviation. Listing of protocol deviation will be provided. In addition, protocol deviations due to COVID-19 will be summarized by center and type of deviation for randomized subjects. Listing of subjects with protocol deviations affected by the COVID-19 will also be provided.

8 Efficacy Analysis

8.1 Estimands

8.1.1 Primary Estimand

The primary clinical question of interest is: what is the treatment difference in binge eating days per week after 8 weeks of centanafadine treatment compared to placebo treatment in adult subjects with moderate to severe BED, where no subject would discontinue treatment due to any reason?

The primary estimand has the following attributes:

- Population: adult subjects with moderate to severe BED who meet the inclusion/exclusion criteria.
- Endpoint: change from baseline in binge eating days per week at Weeks 7-8.
- Intercurrent events: premature treatment discontinuation.
- Treatment condition: one of the randomized treatment groups of the following per protocol: high dose centanafadine 400 mg TDD, low dose centanafadine 200 mg TDD, or placebo.
- Population-level summary: difference in mean of endpoint between treatment conditions.
- Rationale for estimand: the hypothetical estimand is to evaluate the pharmacological effect if no withdrawals had occurred. It is justifiable since the focus is on the pharmacological effect of the drug additional to non-specific effects. Subjects who withdraw from a symptomatic IMP treatment either could have lost the treatment effect if the subjects had not taken any other symptomatic medication after withdrawal or could have the treatment effect masked if the subjects had taken other symptomatic medication after withdrawal. In those cases, any observations taken after subjects stop IMP will most likely not contribute to relevant information about the pharmacological effect of the drug. Therefore, the last collected efficacy assessment after premature trial discontinuation will be performed only once at the ET Visit.

8.1.2 Secondary and Other Estimands

The clinical question of interest for the key secondary estimand is: what is the treatment difference in CGI-S Score after 8 weeks of centanafadine treatment compared to placebo treatment in adult subjects with moderate to severe BED, where no subject would discontinue treatment due to any reason?

The key secondary estimand has similar attributes to the primary estimand except that key secondary endpoint is change from baseline in CGI-S score at Week 8. Other estimands have similar attributes to the primary and key secondary estimands.

8.2 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline in binge eating days per week at Week 7-8, having binge eating day defined as a day with at least one confirmed binge eating episode by principal investigator.

8.2.1 Primary Efficacy Analysis

The primary efficacy endpoint is the change from baseline in binge eating days per week at Weeks 7-8, having binge eating day defined as a day with at least one binge eating episode. The primary estimand defining the treatment effect of interest in the trial uses the hypothetical strategy specified in the draft ICH E9 (R1) Addendum[8]. In the hypothetical strategy, the event of withdrawing IMP is considered MAR, and the primary endpoint of the trial could be considered as a combination of the responses of on-treatment completers at Week 8 and the imputation of the endpoint to Week 8 following the trend in each treatment group using the MMRM method for subjects who withdraw IMP during the trial. All data collected during the trial treatment period will be used for statistical analysis.

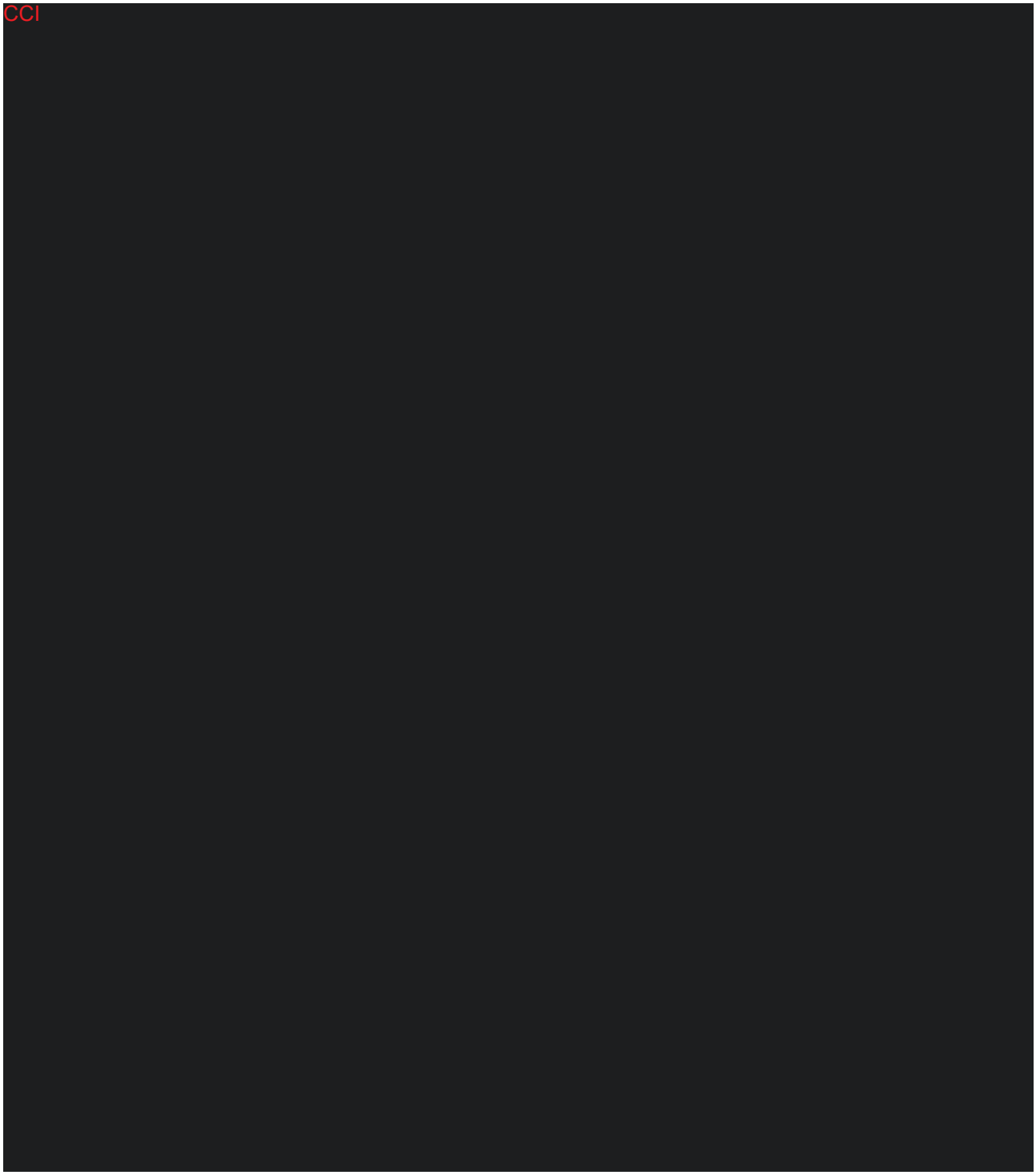
The comparison between the centanafadine groups and the placebo group will be tested at a significance level of 0.05 (2-sided) in a fixed sequence testing of 1) high dose centanafadine versus placebo, and 2) low dose centanafadine versus placebo, in which Test 2) won't be performed unless Test 1) is statistically significant.

The primary efficacy analysis will be performed by fitting an MMRM analysis with an unstructured variance covariance structure, in which change from baseline in binge eating days per week at the scheduled visits will be the dependent variable. The model will include fixed effect terms for treatment, trial center, visit week, and an interaction term of treatment by visit week. The model will also include the interaction term of baseline values of binge eating days per week by visit week as covariates. The level of visit in the MMRM will include Week 1, Week 2, Week 3, Week 4, Week 6, and Week 8. The primary comparison between centanafadine groups and the placebo group at Week 7-8 will be estimated as the difference between least squares means utilizing the computing software SAS procedure PROC MIXED.

If there is a convergence problem with the unstructured variance covariance matrix of the MMRM model, the following structures other than unstructured will be used in order of 1) heterogeneous toeplitz, 2) heterogeneous autoregressive of order 1, and 3)

heterogeneous compound symmetry 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 "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.

8.2.2 Sensitivity Analyses



CCI

[redacted]

8.2.2.3 Sensitivity Analyses for Violation of Normality Assumption

MMRM model for the primary efficacy endpoint is a maximum likelihood method that relies on normality assumption. Residual analyses will be carried out to examine model assumption by using Shapiro-Wilk test.

In the case of gross violations of the normality assumptions, nonparametric van Elteren test[9] will be performed to compare treatment effect in primary efficacy endpoint. The stratification factor is trial center. 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.

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[10]. 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.

8.2.3 Technical Computational Details for Primary Efficacy Analysis

8.2.3.1 Calculation of Binge Eating Days per Week

Binge eating day is defined as a day with at least one confirmed binge eating episode. Binge eating days weekly[11] is calculated as the number of confirmed binge eating days divided by the number of non-missing diary days and multiplied by 7 within specified trial day intervals for the mapped week in Table 5.2.-1. In addition, it is required to have non-missing diary data for at least 5 days in calculation of weekly binge eating days. If subjects have missing diary data for more than 2 days at a specific week prior to termination/completion, weekly binge eating days will be set as missing.

Binge eating days per week at Week 1, 2, 3 and 4 (scheduled visits) is equivalent to the calculated weekly binge eating days as specified above. Binge eating days per week for week 6 (scheduled visit) is average of binge eating days on the mapped Week 5 and binge eating days on the mapped Week 6. Similarly, binge eating days per week for week 8(scheduled visit) is average of binge eating days on the mapped Week 7 and/or binge eating days on the mapped Week 8. The derived binge eating days per week at Week 1, 2, 3, 4, 6 and 8 (scheduled visits) will be included in MMRM model for primary efficacy analyses.

8.2.3.2 Pooling Centers

Small centers will be defined as centers that do not have at least one subject with the evaluable primary efficacy endpoint) in each treatment arm. All small centers will be pooled to form “pseudo centers” for the purpose of analysis according to the following algorithm. Small centers will be ordered from the largest to the smallest based on the number of evaluable subjects (ie, subjects who have baseline and at least one post-baseline value for the primary endpoint). The process will start by pooling the largest of the small centers with the smallest of the small centers until a non-small center is formed. This process will be repeated using the centers left out of the previous pass. In case of

ties in center size, the center with the smallest center code will be selected. If any centers are left out at the end of this process, they will be pooled with the smallest pseudo centers, or if no pseudo centers exist, they will be pooled with the smallest non-small center.

8.2.3.3 Data Processing Convention

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 center 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;
```

8.3 Key Secondary Efficacy Endpoint

The key secondary endpoint is change from baseline in CGI-S score at Week 8. CGI-S assessment collected at scheduled visits will be mapped to the weeks per specification in Table 5.2-2. This key secondary efficacy endpoint will be analyzed by fitting the MMRM model with treatment group, trial center, visit week, and an interaction term of treatment by visit week as fixed effect terms, and interaction term of baseline value by week as covariates. The similar approach of selection of covariance matrix for MMRM model will be applied as for primary efficacy endpoint.

Sensitivity analyses for the key secondary endpoint will be performed as specification in Section 8.2.2.1 as needed.

8.4 Control of Experiment-wise Type 1 Error

To control the overall experiment-wise type I error at the 0.05 level, a fixed sequence testing approach will be applied. The statistical test will be performed in the following order:

1) Change from baseline to Week 7-8 in binge eating days per week between high dose centanafadine and placebo;
2) Change from baseline to Week 7-8 in binge eating days per week between low dose centanafadine and placebo;
3) Change from baseline to Week 8 in CGI-S score between high dose centanafadine and placebo;

4) Change from baseline to Week 8 in CGI-S score between low dose centanafadine and placebo.

The test for the subsequent hypothesis will be performed only if the test for the previous hypothesis is statistically significant.

8.5 Other Efficacy Endpoints

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

Change from baseline in CGI-S score at Weeks 1, 2, 3, 4, and 6 will be analyzed within same methods as specified in Section 8.3.

Cochran-Mantel-Haenszel (CMH) row mean scores differ test controlling trial center will be applied to CGI-C score for LOCF dataset and for OC datasets. Patient Global Impression - Change (PGI-C) score will be analyzed similarly as CGI-C.

Changes from baseline in Y-BOCS-BE total score will be analyzed similarly using MMRM as key secondary efficacy endpoint in Section 8.3. Changes from baseline in PGI-S score and SF 36v2 PCS and MCS scores will be analyzed similarly as changes from baseline in Y-BOCS-BE total score.

Four-week cessation from binge is defined as no binge eating episodes for 28 consecutive days before the last visit. Proportion of subjects with four-week cessation from binge at Week 8/ET will be evaluated by the CMH statistics test controlling for trial center. Risk difference of cessation rate between centanafadine groups and placebo group may be provided as needed.

Change from baseline in number of binge episodes per week will be analyzed similarly using MMRM model as primary efficacy endpoint in Section 8.2.1. Number of binge episodes weekly is number of confirmed binge eating episodes multiplied by 7 and divided by the number of non-missing diary days within trial days interval specified in Table 5.2-1. Number of binge episodes per week at Week1, 2, 3, 4, 6 and 8 (scheduled visits) will be calculated similarly as number of binge eating days as specified in Section 8.2.3.1.

Change from baseline in EDE-Q7 global score at Week 8 and last visit will be analyzed using analysis of covariance (ANCOVA) including treatment and trial center as fixed factors and baseline as covariate.

8.6 Exploratory Analysis

Exploratory efficacy analyses will be presented in Statistical Documentation as needed.

Line plots for mean of primary efficacy endpoint will be presented by week to explore missing patterns across the treatment groups.

Treatment-by-center interaction will be assessed at Week 8 by including the treatment-by-center-by-visit interaction in the models.

The primary, key secondary and other efficacy endpoints will be analyzed using ANCOVA model, including treatment and trial center as fixed factors and baseline as covariate, will be also performed by week, except the endpoints of frequency of subjects with four-week cessation from binge at Week 8/ET, CGI-C and PGI-C.

Y-BOCS-BE subscale scores for obsession and compulsion maybe explored using similar MMRM model and ANCOVA for Y-BOCS-BE total score. In addition, percentage of subjects in different severity (i.e., sub-clinical, mild, moderate, severe, and extreme) will be provided by week and treatment group.

EDE-Q7 subscale scores for dietary restraint, shape/weight overvaluation, and body dissatisfaction will be analyzed using same model for EDE-Q7 total score.

Cohen's D effect size will be provided for the efficacy endpoints as continuous variables, if necessary. Cohen's D effect size is the difference between the two means divided by their standard deviation[12]. Least square mean (LSMean) difference of treatment effects between centanafadine and placebo groups, and the standard error of the difference will be obtained from MMRM, or ANCOVA. Let n_1 and n_2 denote the respective sample sizes of the two groups for pair-wise comparison,

Cohen's D Effect Size = (LSMean $_{\text{centanafadine}}$ – LSMean $_{\text{placebo}}$) / σ, where $\sigma = \frac{\text{Stderr}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$.

Subgroup analyses for primary efficacy endpoints may be explored by age (<40 years old vs >=40 years old), gender (Male vs Female), race(Caucasian vs non-Caucasian) or other baseline characteristics.

9 Safety Analyses

In general, analysis of safety data will be performed for safety analysis set on observed case (including last visit) unless otherwise specified.

9.1 Extent of Exposure

The number and percentage of subjects who receive IMP, will be presented by week and by treatment group. Each dosing week will be based on the actual week, e.g., Day 1 to 7 in Week 1, Day 8 to 14 in week 2, etc. This summary will be performed on randomized analysis set.

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

9.2 Adverse Events

All 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 will be summarized by treatment group:

a) Treatment-emergent AEs (TEAEs)
b) TEAEs by severity
c) TEAEs potentially causally related to the IMP
d) TEAEs with an outcome of death
e) Serious TEAEs
f) TEAEs leading to discontinuation of the IMP
g) Treatment-emergent Adverse Events of Special Interest (AESI) related to Rash
h) TEAEs related to Abuse
i) TEAEs involving Medication Handling Irregularities (MHI)

AEs will be classified by primary SOC and PT according to MedDRA. AEs that are gender-specific, e.g., ovarian cancer, will have their incidence rates evaluated for the specific gender. TEAEs are all adverse events which started after start of IMP; or if the event was continuous from end of the single-blind placebo run-in period and was worsening, serious, study drug related, or resulted in death, discontinuation, interruption, or reduction of study therapy. Adverse events occurring up to 30 days after the last day of double-blind dosing will be included in the summary tables.

Incidence of TEAEs will be summarized by treatment group. Incidence of TEAE of at least 5% in either centanafadine group and greater than placebo will be provided by treatment group, SOC and MedDRA PT.

AESI related to rash will be selected from AESI worksheet and AE CRF. TEAEs related to abuse will be searched based on specified MedDRA PTs. TEAEs involving MHI will be selected from Events Subject to Additional Monitoring (ESAM) form and AE CRF.

9.3 Vital Sign Data

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

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

9.4 Clinical Laboratory Data

Clinical laboratory tests include serum chemistry, hematology, and urinalyses analyses. Potentially clinically relevant laboratory test abnormalities will be listed by subject and by test. Criteria for potentially clinically relevant laboratory test abnormalities are provided in Appendix 2. Subjects with aspartate aminotransferase (AST) or alanine aminotransferase (ALT) $\geq 3 \times$ upper limit of normal value and total bilirubin $\geq 2 \times$ upper limit of normal value or baseline will be listed.

The incidences of potentially clinically relevant laboratory tests abnormalities based on observations during scheduled and unscheduled post-baseline visits will be tabulated by treatment group. Summary statistics for clinical laboratory measurements and changes from baseline will be presented by treatment group and scheduled visit. Change from baseline in fasting triglycerides and total cholesterol at Week 8/ET will be analyzed using ANCOVA including treatment group as fixed factor and baseline as covariate.

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

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

9.5 Physical Examination Data

By-subject listings will be provided for physical examination findings.

Body weight and BMI collected at scheduled visits will be mapped to the weeks per specification in Table 5.2-2. Potentially clinically relevant changes in body weight will be summarized by treatment group and/or by week.

In addition, change from baseline in body weight and percentage of change from baseline in body weight will be analyzed similarly using MMRM model including treatment group, trial center, visit week, and an interaction term of treatment by visit week as fixed effect terms, and interaction term of baseline weight by week as covariates. Change from baseline in body weight percentage of change from baseline in body weight at all mapped visits and last visit will also be explored using ANCOVA model, including treatment and trial center as fixed factors and baseline weight as covariate.

Change from baseline in BMI will be summarized by treatment group and week. In addition, frequencies and percentage of subjects will be provided for classification of BMI by treatment group and week. Shift table will be provided as needed.

9.6 Electrocardiogram Data

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

- QTcF is the length of the QT interval corrected for heart rate by the Fridericia formula: $QTcF=QT/(RR)^{0.33}$
- QTcN is the length of the QT interval corrected for heart rate by the FDA Neuropharm Division formula: $QTcN=QT/(RR)^{0.37}$

Potentially clinically relevant changes in the 12-lead ECG identified for the safety analysis set will be listed and summarized by treatment group. Criteria for identifying ECG measurements of potential clinical relevance are provided in Appendix 3.

Categorical changes in ECG parameters will be summarized by treatment group based on the following criteria in Table 9.6-1.

Table 9.6-1 Categorical Change Criteria in QT/QTc Parameters		
Classification	**Category**	**Criteria**
QT	New Onset (> 450 msec)	New onset (>450 msec) in QT means a subject who attains a value > 450 msec during treatment period but not at baseline.
QTc *	New Onset (> 450 msec)	New onset (> 450 msec) in QTc means a subject who attains a value > 450 msec during treatment period but not at baseline.
	New Onset (> 450 msec) And > 10% Increase	New onset (> 450 msec) and > 10% increase in QTc means a subject who attains a value > 450 msec and > 10% increase during treatment period but not at baseline
	New Onset (> 500 msec)	New onset (> 500 msec) in QTc means a subject who attains a value > 500 msec during treatment period but not at baseline.
	Increase 30 - 60 msec	Increase from baseline value > 30 and ≤ 60 msec in QTc
	Increase > 60 msec	Increase from baseline value > 60 msec in QTc

* QTc categorical change criteria apply to QTcF and QTcN.

Summary statistics of changes in ECG parameters will be provided by treatment group and scheduled visit. If ECG assessments are repeated for the same trial visit, the last repeated values will be used for generation of mean change from baseline. This is accomplished by sorting subject data by visit date and visit time (if applicable) within the same trial visit.

9.7 Suicidality

Suicidality is assessed based on C-SSRS data. The “Baseline” version of the C-SSRS is completed at screening and the “Since Last Visit” version is completed at all other specified visits (including ET, if applicable). Baseline of C-SSRS data is defined as the last evaluable assessment prior to first IMP. Descriptive statistics of the incidence of suicidality and incidence of suicidality by type (suicidal behavior and suicidal ideation) is reported by treatment group.

Baseline and Since Last Visit C-SSRS data is summarized to report the incidence of suicidality; suicidal behavior only, emergence of suicidal behavior; suicidal ideation only, emergence of suicidal ideation, emergence of serious suicidal ideation, and worsening of suicidal ideation at trial visits and last visit in addition to overall during the double-blinded period.

Suicidality is defined as reporting any suicidal ideation or behavior. Suicidal behavior only is defined as reporting any type of suicidal behaviors (actual attempt, interrupted attempt, aborted attempt and preparatory acts or behavior) throughout assessment period. Emergence of suicidal behavior is defined as having no suicidal behavior at baseline and reporting any type of behavior at post-baseline.

Suicidal ideation only is defined as reporting any type of suicidal ideation (wish to be dead, non-specific active suicidal thoughts, active suicidal ideation with any methods, active suicidal ideation with some intent and active suicidal ideation with specific plan,). Emergence of suicidal ideation is defined as having no suicidal ideation at baseline and reporting any type of ideation during treatment. Emergence of serious suicidal ideation is defined as having no suicidal ideation at baseline and reporting serious suicidal ideation with score of 4 or 5 on suicidal ideation severity rating during the treatment. Worsening of suicidal ideation is defined as having more severe in most severe suicidal ideation rating at post baseline than at baseline. Numbers of subjects reporting suicidal behavior are also reported by type of suicidal behavior.

Listings for treatment-emergent suicidal behavior, suicidal ideation, serious suicidal ideation and worsening of suicidal ideation will also be provided.

9.8 Other Safety Data

Change from baseline in HAM-A total score, MADRS total score will be summarized by treatment group in descriptive statistics. SMWQ total score will be summarized by treatment group.

Classification of MHIs and Classification of Events Reported as findings in ESAM form will be summarized in frequencies and proportion by treatment group.

10 Rules for Scoring

10.1 Clinical Global Impression - Severity

The CGI-S is a standardized, clinician-administered global rating scale that measures disease severity on a 7-point scale. A higher score on the CGI-S represents a higher severity of disease. It is rated as: 1 = normal, not at all ill; 2 = borderline ill; 3 = mildly ill; 4 = moderately ill;5 = markedly ill; 6 = severely ill; 7 = among the most extremely ill subjects.

10.2 Clinical Global Impression - Change

The CGI-C is a single-item, 7-point scale that requires the clinician to assess how much the subject’s illness has improved or worsened relative to their baseline state at the beginning of treatment, rated as: 1 = very much improved; 2 = much improved; 3 = minimally improved; 4 = no change; 5 = minimally worse; 6 = much worse; or 7 = very much worse.

10.3 Yale-Brown Obsessive-Compulsive Scale Modified for Binge Eating

The Y-BOCS-BE is a clinician-rated scale that assesses obsession with binge eating thoughts and compulsiveness of binge eating behaviors. It can be divided into 2 subscales: obsessions (i.e., items 1-5) and compulsions (i.e., items 6-10) . The Y-BOCS-BE is a 10-item scale, with each item rated from 0 (no symptoms) to 4 (extreme symptoms) and total scores ranging from 0 to 40, sub-scales scores ranging from 0 to 20. A score of 0 to 7 is sub-clinical; 8 to 15 is mild; 16 to 23 is moderate; 24 to 31 is severe; and 32 to 40 is extreme. Any missing item will lead total/sub-scale scores missing.

10.4 Eating Disorder Examination Questionnaire 7-Item Version

The EDE-Q7 is a brief version of the EDE-Q which comprises 7 items (to generate a global score and 3 subscales scores for dietary restraint, shape/weight overvaluation, and body dissatisfaction. Severity rating ranges from 0 to 6, where 0 represents absence of the feature and 6 represents an extreme degree. Frequency ratings range from 0 to 6 where 0= 'No days' or absence of the feature and 6= 'every day'. Global scores range from 0 to 42; dietary restraint subscales scores range from 0 to 18 (i.e., sum of scores from question 1 to 3), shape/weight overvaluation subscales scores range from 0 to 12 (i.e., sum of scores from question 4 to 5), and body dissatisfaction subscales scores range from 0 to 12(i.e., sum of scores from question 6 to 7). Any missing item will lead total/sub-scale scores missing.

10.5 Patient Global Impression - Severity

The PGI-S is a single-item self-report of the patient's severity of symptoms. Severity is rated on a 7-point scale as: 1 = no symptoms; 2 = minimal; 3 = mild; 4 = moderate; 5 = marked; 6 = severe; 7 = very severe.

10.6 Patient Global Impression - Change

The PGI-C is a single-item, 7-point patient self-report that requires the patient to assess how much his/her illness has improved or worsened relative to baseline, rated as: 1 = very much improved; 2 = much improved; 3 = minimally improved; 4 = no change; 5 = minimally worse; 6 = much worse; or 7 = very much worse.

10.7 36-Item Short-Form Health Survey

The 36-item Short-form Health Survey Version 2 (SF-36v2) is a patient-reported questionnaire with a standard recall period of 4 weeks which measures generic health-related quality of life on 2 broad domains, physical and mental composites, across eight health domain scales: physical functioning, pain, role physical, general health,

vitality/fatigue, social functioning role emotional, and mental health. The SF-36v2 uses norm-based scoring to generate scores on a scale of 0 to 100 where lower scores on PCS and MCS represent a lower health-related quality of life and a score of 50 references the normal US population. Those scores will be derived through QualityMetric's software[13].

10.8 Hamilton Anxiety Rating Scale

The HAM-A consists of 14 items, each defined by a series of symptoms, and measures both psychic anxiety (mental agitation and psychological distress) and somatic anxiety (physical complaints related to anxiety). Each item is scored on a 5-point scale, ranging from 0 = not present to 4 = severe. HAM-A total scores range from 0 to 56. Any missing item will lead total scores missing.

10.9 Montgomery-Asberg Depression Rating Scale

The MADRS is a clinician-rated assessment of the subject's severity of depression. The measure contains 10 items that measure apparent and reported sadness, inner tension, reduced sleep and appetite, difficulty concentrating, lassitude, inability to feel, and pessimistic and suicidal thoughts. Each item is scored in a range of 0 to 6 points, with higher scores indicating increased depressive symptoms. MADRS total scores range from 0 to 60. Any missing item will lead total scores missing.

10.10 Study Medication Withdrawal Questionnaire

The SMWQ is a questionnaire to assess withdrawal symptoms that will be assessed at week 8/ET and 7 days after the last dose of IMP in the follow-up period. The SMWQ is a modification of the Amphetamine Withdrawal Questionnaire in which the terms "amphetamines and methamphetamine" are replaced with the term "the study medication." This questionnaire consists of 10 questions. Each is scored in 0 = Not at all, 1 = Very little, 2 = A little, 3 = Quite a lot, 4= Very much. The SMWQ total score ranges from 0 to 40. Any missing item will lead total scores missing.

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

Suicidality will be monitored during the trial using the C-SSRS. The C-SSRS is a semi-structured interview that captures the occurrence, severity, and frequency of suicide-related thoughts and behaviors during the assessment period. The interview includes definitions and suggested questions to solicit the type of information needed to determine if a suicide-related thought or behavior has occurred.

The C-SSRS has a "Screening/Baseline" version, which will be completed at screening and a "Since Last Visit" version that will be completed at all scheduled visits (including the ET visit, if applicable). There are a maximum of 19 items to be completed: 7

required, 10 potential additional items if there is a positive response to a required item, and 2 items for suicide/suicide behavior present during the interview. The C-SSRS uses dichotomous scales (i.e., yes or no), Likert scales, and text or narrative to further describe the thoughts or behaviors.

11 Pharmacokinetic Analyses

Sparse PK samples will be evaluated to confirm compliance with dosing. Concentrations will be listed in the bioanalytical report.

12 Pharmacodynamic Analyses

No pharmacogenomic analyses is planned.

13 Pharmacogenomic Analyses

No pharmacogenomic analyses is planned.

14 Interim Analysis

No interim analyses is planned.

15 Changes in the Planned Analyses

Trial center will not be included in ANCOVA model for analyses of change from baseline in fasting triglycerides and total cholesterol at Week 8/ET.

16 References

1. Siddiqui O, Hung JHM, O'Neill R. MMRM vs. LOCF: A comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharmaceutical Stats. 2009; 19(2):227-46.
2. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. Applied Statistics. 1994; 43:49-93.
3. Little RJA. Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc. 1993; 88:125-34.
4. Little RJA. Modeling the drop-out mechanism in repeated measures studies. J Am Stat Assoc. 1995; 90:1112-21.
5. Hedeker D, Gibbons RD. Application of random effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods. 1997; 2:64-78.
6. Ali MW, Siddiqui O. Multiple imputation compared with some information dropout procedures in the estimation and comparison of rates of change in longitudinal clinical trials with dropouts. J Biopharmaceutical Stats. 2000;10(2):165-81.

7. Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. Biometrics. 1989; 45:939-55.
8. ICH Harmonised Guideline Estimands and Sensitivity Analysis in Clinical Trials E9(R1), 16 June 2017.
9. van Elteren, PH. On the combination of independent two sample tests of Wilcoxon. Bull Int Stat Inst. 1960; 37:351-61.
10. Mehrotra D, Li X, Liu J, Lu K. Analysis of Longitudinal Clinical Trials with Missing Data Using Multiple Imputation in Conjunction with Robust Regression. Biometrics 2012; 68:1250-1259.
11. McElroy LS, Hudson J, Ferreira-Cornwell MC., Radewonuk J., Whitake T., & Gasior M. Lisdexamfetamine Dimesylate for Adults with Moderate to Severe Binge Eating Disorder: Results of Two Pivotal Phase 3 Randomized Controlled Trials. Neuropsychopharmacology (2016) 41, 1251–1260
12. Cortina, J. M., & Nouri, H. (2000). Effect size for ANOVA designs. Thousand Oaks, Calif.: Sage Publications.
13. https://www.qualitymetric.com/scientific-consulting/scoring-interpretation/.

Appendix 1 Criteria for Identifying Vital Signs and Weight of Potential Clinical Relevance

Variable	Criterion Value[a]	Change Relative to Baseline[a]
Heart Rate [b]	> 100 bpm < 50 bpm	≥ 10 bpm increase ≥ 10 bpm decrease
Systolic Blood Pressure [b]	≥ 140 mmHg < 90 mmHg	≥ 20 mmHg increase ≥ 20 mmHg decrease
Diastolic Blood Pressure [b]	≥ 90 mmHg < 60 mmHg	≥ 10 mmHg increase ≥ 10 mmHg decrease
Orthostatic Hypotension	≥ 30 mmHg decrease in systolic blood pressure or a ≥ 20 mmHg in diastolic blood pressure after at least 3 minutes of standing compared to the previous supine blood pressure.	Not Applicable (baseline status not considered)
Orthostatic Tachycardia	≥ 25 bpm increase in heart rate from supine to 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 Criteria for Identifying Laboratory Values of Potential Clinical Relevance

Laboratory Tests	Criteria
Chemistry	
AST (SGOT)	≥ 3 x upper limit of normal (ULN)
ALT (SGPT)	≥ 3 x ULN
Alkaline phosphatase	≥ 3 x ULN
BUN	≥ 30 mg/dL
Creatinine	≥ 2.0 mg/dL
Uric Acid	
Men	≥ 10.5 mg/dL
Women	≥ 8.5 mg/dL
Bilirubin (total)	≥ 2.0 mg/dL
Creatine Phosphokinase (CPK)	> 3 x ULN
Hematology	
Hematocrit	
Male	≤ 37 % and decrease of ≥ 3 percentage points from Baseline
Female	≤ 32 % and decrease of ≥ 3 percentage points from Baseline
Hemoglobin	
Men	≤ 11.5 g/dL
Women	≤ 9.5 g/dL
White blood count	≤ 2,800/ mm^3 or ≥ 16,000/ mm^3
Eosinophils	≥ 10%
Neutrophils	≤ 15%
Absolute neutrophil count	≤ 1,500/ mm^3
Platelet count	≤ 75,000/ mm^3 or ≥ 700,000/ mm^3
Urinalysis	
Protein	Increase of ≥ 2 units
Glucose	Increase of ≥ 2 units
Additional Criteria	
Chloride	≤ 90 mEq/L or ≥ 118 mEq/L
Potassium	≤ 2.5 mEq/L or ≥ 6.5 mEq/L
Sodium	≤ 126 mEq/L or ≥ 156 mEq/L
Calcium	≤ 8.2 mg/dL or ≥ 12 mg/dL
Glucose	
Fasting	≥ 100 mg/dL
Non-Fasting	≥ 200 mg/dL
Total Cholesterol, Fasting	≥ 240 mg/dL
LDL Cholesterol, Fasting	≥ 160 mg/dL
HDL Cholesterol, Fasting	
Male	< 40 mg/dL
Female	< 50 mg/dL
Triglycerides, Fasting	≥ 150 mg/dL

Appendix 3 Criteria for Identifying ECG Measurements of Potential Clinical Relevance

Variable	Criterion Value[a]	Change Relative to Baseline[a]
Rate		
Tachycardia	≥ 120 bpm	increase of ≥ 15 bpm
Bradycardia	≤ 50 bpm	decrease of ≥ 15 bpm
Rhythm		
Sinus tachycardia[b]	≥ 120 bpm	increase of ≥ 15 bpm
Sinus bradycardia[c]	≤ 50 bpm	decrease of ≥ 15 bpm
Supraventricular premature beat	all	not present → present
Ventricular premature beat	all	not present → present
Supraventricular tachycardia	all	not present → present
Ventricular tachycardia	all	not present → present
Atrial fibrillation	all	not present → present
Atrial flutter	all	not present → present
Conduction		
1° atrioventricular block	PR ≥ 200 msec	increase of ≥ 50 msec
2° atrioventricular block	all	not present → present
3° atrioventricular block	all	not present → present
Left bundle-branch block	all	not present → present
Right bundle-branch block	all	not present → present
Pre-excitation syndrome	all	not present → present
Other intraventricular conduction block[d]	QRS ≥ 120 msec	increase of ≥ 20 msec
Infarction		
Acute or subacute	all	not present → present
Old	all	not present → present ≥ 12 weeks post study entry
ST/T Morphological		
Myocardial Ischemia	all	not present → present
Symmetrical T-wave inversion	all	not present → present
Increase in QTc	QTcF > 450 msec (male) QTcF > 470 msec (female)	

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

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

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

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

Appendix 4 List of Proposed Summary Tables

CT-1.1 Subject Disposition
CT-1.2 Subject Disposition by Center
CT-1.3 Subject Completion Rates by Week (Randomized Analysis Set)
CT-2 Reasons for Discontinuation (Randomized Analysis Set)
CT-3.1 Demographic Characteristics (Randomized Analysis Set)
CT-3.2.1 Medical History (Randomized Analysis Set)
CT-3.2.2 Psychiatric History (Randomized Analysis Set)
CT-3.3 Baseline Disease Evaluations (Randomized Analysis Set)
CT-4.1 Concomitant Medications: Medications Taken Prior to Start of Study Period (Randomized Analysis Set)
CT-4.2 Concomitant Medications: Medications Taken During the Study Period (Randomized Analysis Set)
CT-4.3 Concomitant Medications: Medications Taken Post Study Period (Randomized Analysis Set)
CCI [redacted]
CT-5.2.1 Summary of Change from Baseline in Binge Eating Days per Week - MMRM (Efficacy Analysis Set)
CCI [redacted]
CT-5.3.1 Summary of Change from Baseline in Clinical Global Impression - Severity Score - MMRM (Efficacy Analysis Set)
CCI [redacted]

CCI [redacted]

CT-5.3.5 Summary of Change from Baseline in Clinical Global Impression - Severity Score - ANCOVA (Efficacy Analysis Set)
CT-5.4.1 Summary of Clinical Global Impression - Change Score - LOCF (Efficacy Analysis Set)
CT-5.4.2 Summary of Clinical Global Impression - Change Score - OC (Efficacy Analysis Set)
CT-5.5.1.1 Summary of Change from Baseline in Yale-Brown Obsessive-Compulsive Scale Modified for Binge Eating Total Score - MMRM (Efficacy Analysis Set)
CT-5.5.1.2 Summary of Change from Baseline in Yale-Brown Obsessive-Compulsive Scale Modified for Binge Eating Total Score - ANCOVA (Efficacy Analysis Set)
CT-5.5.2.1 Summary of Change from Baseline in Yale-Brown Obsessive-Compulsive Scale Modified for Binge Eating Obsession Score - MMRM (Efficacy Analysis Set)
CT-5.5.2.2 Summary of Change from Baseline in Yale-Brown Obsessive-Compulsive Scale Modified for Binge Eating Obsession Score - ANCOVA (Efficacy Analysis Set)
CT-5.5.3.1 Summary of Change from Baseline in Yale-Brown Obsessive-Compulsive Scale Modified for Binge Eating Compulsion Score - MMRM (Efficacy Analysis Set)
CT-5.5.3.2 Summary of Change from Baseline in Yale-Brown Obsessive-Compulsive Scale Modified for Binge Eating Compulsion Score - ANCOVA (Efficacy Analysis Set)
CT-5.5.4 Categorical Analysis of Yale-Brown Obsessive-Compulsive Scale Modified for Binge Eating (Efficacy Analysis Set)
CT-5.6.1 Summary of Change from Baseline in Number of Binge Episodes per Week - MMRM (Efficacy Analysis Set)
CT-5.6.2 Summary of Change from Baseline in Number of Binge Episodes per Week - MMRM (Efficacy Analysis Set)
CT-5.7 Proportion of Subjects with 4-week Cessation from Binge at Week 8/ET (Efficacy Analysis Set)
CT-5.8.1 Summary of Change from Baseline in Patient Global Impression - Severity Score - MMRM (Efficacy Analysis Set)
CT-5.8.2 Summary of Change from Baseline in Patient Global Impression - Severity Score - ANCOVA (Efficacy Analysis Set)
CT-5.9.1 Summary of Patient Global Impression - Change Score - LOCF (Efficacy Analysis Set)
CT-5.9.2 Summary of Patient Global Impression Improvement - Change Score - OC (Efficacy Analysis Set)
CT-5.10.1.1 Summary of Change from Baseline in 36-item Short-form Health Survey Version 2 (SF-36v2) Physical Component Summary Score - MMRM (Efficacy Analysis Set)
CT-5.10.1.2Summary of Change from Baseline in 36-item Short-form Health Survey Version 2 (SF-36v2) Physical Component Summary Score - ANCOVA (Efficacy Analysis Set)

CT-5.10.2.1 Summary of Change from Baseline in 36-item Short-form Health Survey Version 2 (SF-36v2) Mental Component Summary Score - MMRM (Efficacy Analysis Set)
CT-5.10.2.2 Summary of Change from Baseline in 36-item Short-form Health Survey Version 2 (SF-36v2) Mental Component Summary Score - ANCOVA (Efficacy Analysis Set)
CT-5.11.1 Eating Disorder Examination Questionnaire - 7-Item Version (EDE-Q7) Global Score - ANCOVA (Efficacy Analysis Set)
CT-5.11.2 Eating Disorder Examination Questionnaire - 7-Item Version (EDE-Q7) Dietary Restraint Subscale Score - ANCOVA (Efficacy Analysis Set)
CT-5.11.3 Eating Disorder Examination Questionnaire - 7-Item Version (EDE-Q7) Shape/Weight Overvaluation Subscale Score - ANCOVA (Efficacy Analysis Set)
CT-5.11.4 Eating Disorder Examination Questionnaire - 7-Item Version (EDE-Q7) Body Dissatisfaction Subscale Score - ANCOVA (Efficacy Analysis Set)
CT-6.1.1 Columbia-Suicide Severity Rating Scale - Suicidality (Safety Analysis Set)
CT-6.1.2 Columbia-Suicide Severity Rating Scale - Suicidal Behavior by Type (Safety Analysis Set)
CT-6.1.3 Columbia-Suicide Severity Rating Scale - Suicidal Ideation by Type (Safety Analysis Set)
CT-6.1.4 Columbia-Suicide Severity Rating Scale - Treatment Emergent Suicidal Behavior and Ideation (Safety Analysis Set)
CT-6.1.5 Columbia-Suicide Severity Rating Scale - Listing of Treatment Emergent Suicidal Behavior (Safety Analysis Set)
CT-6.1.6 Columbia-Suicide Severity Rating Scale - Listing of Treatment Emergent Suicidal Ideation (Safety Analysis Set)
CT-6.1.7 Columbia-Suicide Severity Rating Scale - Listing of Treatment Emergent Serious Suicidal Ideation (Safety Analysis Set)
CT-6.1.8 Columbia-Suicide Severity Rating Scale - Listing of Worsening Suicidal Ideation (Safety Analysis Set)
CT-6.2 Summary of Change from Baseline in Hamilton Anxiety Rating Scale Total Score (Safety Analysis Set)
CT-6.3 Summary of Change from Baseline in Montgomery-Asberg Depression Rating Scale Total Score (Safety Analysis Set)
CT-6.4 Summary of Study Medication Withdrawal Questionnaire (SMWQ) Total Score (Safety Analysis Set)

CT-7.1 Extent of Exposure (Safety Analysis Set)
CT-7.2 Number and Percentage of Subjects Receiving Study Medication and Mean and Range of Average Daily Dose (Safety Analysis Set)
CT-8.1.1 Adverse Events (All Causalities) (Safety Analysis Set)
CT-8.1.2 Adverse Events During the Single-Blind Placebo Run-in Period (All Causalities) (Safety Analysis Set)
CT-8.2.1 Incidence of Treatment-Emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.2.2 Incidence of Treatment-Emergent Adverse Events by System Organ Class, MedDRA Preferred Term and Severity (Safety Analysis Set)

CT-8.2.3 Incidence of Treatment-Emergent Adverse Events of at Least 5% in Any Centanafadine Group and Greater Than Placebo by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.2.4 Incidence of Non-Serious Treatment-Emergent Adverse Events of at Least 5% in Any Centanafadine Group and Greater Than Placebo by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.2.5 Incidence of TREATMENT-EMERGENT ADVERSE Events of at Least 2% in Any Centanafadine Group and Greater Than Placebo by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.2.6 Incidence and Occurrence (Number of Events) of Serious Treatment-Emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.2.7 Incidence and Occurrence (Number of Events) of Potentially Drug-Related Serious Treatment-Emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.2.8 Incidence and Occurrence (Number of Events) of Non-Serious Treatment-Emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.3.1 Incidence of Potentially Drug-Related Treatment-Emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.3.2 Incidence of Potentially Drug-Related Treatment-Emergent Adverse Events by System Organ Class, MedDRA Preferred Term and Severity (Safety Analysis Set)
CT-8.4.1 Incidence of Deaths Due to Treatment-Emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.4.2 Incidence of Deaths Due to Treatment-Emergent Adverse Events by System Organ Class and MedDRA Preferred Term and Severity (Safety Analysis Set)
CT-8.5.1 Incidence of Serious Treatment-Emergent Adverse Events by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.5. 2 Incidence of Serious Treatment-Emergent Adverse Events by System Organ Class, MedDRA Preferred Term and Severity (Safety Analysis Set)
CT-8.6.1 Incidence of Treatment-Emergent Adverse Events Resulting in Discontinuation of Investigational Medicinal Product by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.6.2 Incidence of Treatment-Emergent Adverse Events Resulting in Discontinuation of Investigational Medicinal Product by System Organ Class, MedDRA Preferred Term and Severity (Safety Analysis Set)
CT-8.7.1 Incidence of Treatment-Emergent Adverse Events of Special Interests related to Rash by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.7.2 Incidence of Treatment-Emergent Adverse Events of Special Interests related to Rash by System Organ Class and MedDRA Preferred Term and Severity (Safety Analysis Set)
CT-8.8.1 Incidence of Treatment-Emergent Adverse Events related to Abuse by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.8.2 Incidence of Treatment-Emergent Adverse Events related to Abuse by System Organ Class and MedDRA Preferred Term and Severity (Safety Analysis Set)

CT-8.9.1 Incidence of Treatment-Emergent Adverse Events involving Medication Handling Irregularities by System Organ Class and MedDRA Preferred Term (Safety Analysis Set)
CT-8.9.2 Incidence of Treatment-Emergent Adverse Events involving Medication Handling Irregularities by System Organ Class and MedDRA Preferred Term and Severity (Safety Analysis Set)
CT-9.1 Listing of Deaths
CT-9.2 Listing of Serious Adverse Events
CT-9.3 Listing of Adverse Event Leading Discontinuations of Investigation Medicinal Products
CT-9.4 Listing of Treatment-Emergent Adverse Events of Special Interests related to Rash
CT-9.5 Listing of Treatment-Emergent Adverse Events related to Abuse
CT-9.6 Listing of Treatment-Emergent Adverse Events involving Medication Handling Irregularities
CT-10.1 Criteria for Laboratory Test Values with Potential Clinical Relevance
CT-10.2.1 Listing of Laboratory Test Values with Potential Clinical Relevance by Subject (Safety Analysis Set)
CT-10.2.2 Listing of Laboratory Test Values with Potential Clinical Relevance by Test (Safety Analysis Set)
CT-10.3 Incidence of Laboratory Test Values with Potential Clinical Relevance (Safety Analysis Set)
CT-10.4.1.1 Summary of Change from Baseline in Clinical Laboratory Test Results - Serum Chemistry (Safety Analysis Set)
CT-10.4.1.2 Summary of Change from Baseline in Clinical Laboratory Test Results - Lipid Parameters - ANCOVA (Safety Analysis Set)
CT-10.4.2 Summary of Change from Baseline in Clinical Laboratory Test Results - Hematology (Safety Analysis Set)
CT-10.4.3 Summary of Change from Baseline in Clinical Laboratory Test Results - Urinalysis (Safety Analysis Set)
CT-10.5.1 Incidence of Potentially Liver Injury Related Laboratory Test Abnormalities (Safety Analysis Set)
CT-10.5.2 Listing of Potentially Liver Injury Related Laboratory Test Abnormalities (Safety Analysis Set)
CT-11.1 Criteria for Potentially Clinically Relevant Abnormalities in Vital Signs
CT-11.2 Listing of Potentially Clinically Relevant Abnormalities in Vital Signs (Safety Analysis Set)
CT-11.3 Incidence of Potentially Clinically Relevant Abnormalities in Vital Signs (Safety Analysis Set)
CT-11.4 Summary of Change from Baseline in Vital Signs (Safety Analysis Set)
CT-12.1 Criteria for Potentially Clinically Relevant Abnormalities in ECG Evaluations
CT-12.2 Listing of Potentially Clinically Relevant Abnormalities in ECG Evaluations (Safety Analysis Set)
CT-12.3 Incidence of Potentially Clinically Relevant Changes in ECG Evaluations (Safety Analysis Set)
CT-12.4 Mean Change from Baseline in Electrocardiogram Results (Safety Analysis Set)

CT-12.5 Incidence of Categorical Changes in QT/QTc (Safety Analysis Set)
CT-13.1.1 Summary of Change from Baseline in Body Weight by Mapped Week - MMRM (Safety Analysis Set)
CT-13.1.2 Summary of Change from Baseline in Body Weight by Mapped Week - ANCOVA (Safety Analysis Set)
CT-13.2.1 Summary of Percentage of Change from Baseline in Body Weight by Mapped Week - MMRM (Safety Analysis Set)
CT-13.2.2 Summary of Percentage of Change from Baseline in Body Weight by Mapped Week - ANCOVA (Safety Analysis Set)
CT-13.3 Summary of Proportion of Subjects with Potentially Clinically Relevant Weight Gain or Weight Loss by Mapped Week (Safety Analysis Set)
CT-13.4.1 Summary of Change from Baseline in BMI by Mapped Week (Safety Analysis Set)
CT-13.4.2 Classification of BMI by Mapped Week (Safety Analysis Set)
CT-14 Summary of Medication Handling Irregularity (Safety Analysis Set)
CT-15 Summary of Events Reported as Findings in Events Subject to Additional Monitoring (Safety Analysis Set)

Appendix 5 Proposed Subjects Data Listing

DREAS-1 Discontinued Subjects and Reasons for Discontinuation
DREAS-2 Adverse Events for Subjects Who Discontinued Trial due to Withdrew Consent

PDEV-1.1.1 Summary of Protocol Deviations by Type of Deviation
PDEV-1.1.2 Summary of Protocol Deviations by Center and Type of Deviation
PDEV-1.2.1 Summary of Protocol Deviations due to COVID-19 by Type of Deviation
PDEV-1.2.2 Summary of Protocol Deviations due to COVID-19 by Center and Type of Deviation
PDEV-2 Protocol Deviations by Subject
PDEV-3 Protocol Deviation Criteria

SUBEX-1 Inclusion/Exclusion from Efficacy Analysis

DEMOG-1 Demographic Characteristics

SMED-1.1 Summary of Investigational Medicinal Product Compliance
SMED-1.2 Investigational Medicinal Product Compliance

EFF-1.1 Change from Baseline in Number of Binge Eating Days per Week
EFF-1.2 Change from Baseline in Number of Binge Eating Episode per Week
EFF-1.3 Status for Subjects with Four-week Cessation from Binge at Week 8/ET
EFF-2.1 Change from Baseline in Clinical Global Impression - Severity Score
EFF-2.2 Clinical Global Impression - Change Score
EFF-3 Change from Baseline in Yale-Brown Obsessive-Compulsive Scale Modified for Binge Eating Total Score and Subscale Scores
EFF-4.1 Change from Baseline in Patient Global Impression - Severity Score
EFF-4.2 Patient Global Impression - Change Score
EFF-5 Change from Baseline in 36-item Short-form Health Survey Version 2 - Physical Component Summary and Mental Component Summary Scores
EFF-6 Change from Baseline in Eating Disorder Examination Questionnaire - 7-Item Version Global Score and Subscale Scores

AE-1 Adverse Events

LAB-1 Laboratory Test Results: Serum Chemistry
LAB-2 Laboratory Test Results: Hematology
LAB-3 Laboratory Test Results: Urinalysis
LAB-4 Laboratory Test Results: Urine Drug Screen
LAB-5 Laboratory Test Results: Urine Pregnancy Test
LAB-6 Laboratory Test Results: Other Laboratory Tests

STAT-1.1.1 SAS Output for Change from Baseline in Binge Eating Days per Week - MMRM (Efficacy Analysis Set)

STAT-1.1.2 SAS Output for Change from Baseline in Binge Eating Days per Week - ANCOVA (Efficacy Analysis Set)
STAT-1.2.1 Residual Analysis of MMRM for Change from Baseline in Binge Eating Days per Week - Shapiro-Wilk Test (Efficacy Analysis Set)
STAT-1.2.2 SAS Output for Residual by Week from MMRM for Change from Baseline in Binge Eating Days per Week (Efficacy Analysis Set)
STAT-2.1 SAS Output for Change from Baseline in Clinical Global Impression - Severity Score - MMRM (Efficacy Analysis Set)
STAT-2.2 SAS Output for Change from Baseline in Clinical Global Impression - Severity Score - ANCOVA (Efficacy Analysis Set)
STAT-3 Estimated Effect Size for Efficacy Endpoints by End of Week 8
STAT-4.1.1 Summary of Change from Baseline in Binge Eating Days per Week by Gender - MMRM (Efficacy Analysis Set)
STAT-4.1.2 Summary of Change from Baseline in Binge Eating Days per Week by Race - MMRM (Efficacy Analysis Set)
STAT-4.1.3 Summary of Change from Baseline in Binge Eating Days per Week by Age - MMRM (Efficacy Analysis Set)

CF-1.1 Plot for Change from Baseline in Binge Eating Days per Week
CF-1.2 Forest plot for Sensitivity Analysis of Change from Baseline in Binge Eating Days per Week
CF-2.1 Plot for Change from Baseline in Clinical Global Impression - Severity Score
CF-2.2 Forest Plot for Sensitivity Analysis of Change from Baseline in Clinical Global Impression - Severity Score
CF-3 Incidence of Treatment-Emergent Adverse Events of at Least 5% in Any Centanafadine Group and Greater Than Placebo
CF-4.1 Mean of Change from Baseline in Binge Eating Days per Week by Dropout Cohort - OC: Centanafadine 400mg
CF-4.2 Mean of Change from Baseline in Binge Eating Days per Week by Dropout Cohort - OC: Centanafadine 200mg
CF-4.3 Mean of Change from Baseline in Binge Eating Days per Week by Dropout Cohort - OC: Placebo

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: P40520100056_Statistical Analysis Plan_Final

Document Number: 1000177558

Document Version: 3.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
PPD	Clinical Approval	08-Sep-2022 13:53:25
PPD	Biostatistics Approval	06-Sep-2022 22:06:22