Statistical Analysis Plan

A 12-week, Randomized, Double-blind, Parallel-group, Placebo-controlled, Fixed-dosed, Multicenter Study to Evaluate the Efficacy, Safety, and Tolerability of Dasotraline in Adults with Moderate to Severe Binge Eating Disorder

DASOTRALINE (SEP-225289)

Protocol SEP360-321

Version Number: V 1.0

Version Date: June 22, 2018

Prepared by INC Research
3201 Beechleaf Court
Raleigh, NC 27604-1547

Authorization Signature Page

A 12-WEEK, RANDOMIZED, DOUBLE-BLIND, PARALLEL-GROUP, PLACEBO-CONTROLLED, FIXED-DOSED, MULTICENTER STUDY TO EVALUATE THE EFFICACY, SAFETY, AND TOLERABILITY OF DASOTRALINE IN ADULTS WITH MODERATE TO SEVERE BINGE EATING DISORDER

Author:

Name: Date

Position: Principal Biostatistician
Company: INC Research

Approved by:

Manager, Biostatistics, Date
INC Research

Sr. Director, Clinical Research, Date
Sunovion Pharmaceuticals Inc.

Sr. Director, Biostatistics, Date
Sunovion Pharmaceuticals Inc.

Executive Director, Biostatistics, Date
Sunovion Pharmaceuticals Inc.

Table of contents

List of Abbreviations and Definition of Terms

Abbreviation	Full Form
AE	Adverse event
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
APMP	Abuse Potential Monitoring Plan
AST	Aspartate aminotransferase
BDRM	Blind Data Review Meeting
BE-CGI-S	Binge-eating Clinical Global Impression-Severity
BED	Binge eating disorder
BMI	Body mass index
BUN	Blood urea nitrogen
CFR	Code of Federal Regulations
CI	Confidence interval
CRF	Case report form
C-SSRS	Columbia-suicide severity rating scale
CSSA	Cocaine Selective Severity Assessment
CST	Clinical Surveillance Team
DAS	Dasotraline
DB	Double-Blind
DBL	Database lock
DEAE	Discontinuation-emergent adverse event
DESS	Discontinuation-Emergent Signs and Symptoms
DHPG	3,4-dihydroxyphenylglycol
DNA	Deoxyribonucleic acid
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
DSMB	Data and Safety Monitoring
ECG	Electrocardiogram
EDC	Electronic data capture
EDE-Q	Eating Disorder Examination Questionnaire
EDE-Q7	Eating Disorder Examination Questionnaire Brief Version
EDE-QM	Eating Disorder Examination Questionnaire Modified
EOT	End of treatment
FDA	U.S. Food and Drug Administration
GEE	Generalized Estimating Equations
GLMM	Generalized linear mixed model

Abbreviation	Full Form
HAM-A	Hamilton Anxiety Rating Scale
HDL	High-density lipoprotein (cholesterol)
HR	Heart rate
ICF	Informed consent form
IPW	Inverse probability weighting
IRT	Interactive Response Technology
ITT	Intention-to-Treat
IXRS	Interactive Response System
LDL	Low-density lipoprotein (cholesterol)
LOCF	Last Observation Carried Forward
MADRS	Montgomery-Asberg Depression Rating Scale
MCS	Mental Component Summary
MINI	Mini International Neuropsychiatric Interview
MMRM	Mixed-effects Models Repeated Measures
MNAR	Missing not at random
NE	Norepinephrine
NET	Norepinephrine transporter
NNT	Number Needed to Treat
PCS	Physical Component Summary
PGx	Pharmacogenomics
PK	Pharmacokinetic(s)
PMM	Pattern mixture model
PP	Per-Protocol
PR	Time between P wave and QRS in electrocardiography
PT	Preferred term
QRS	Electrocardiographic wave (complex or interval)
QT	Electrocardiographic interval from the beginning of the QRS complex to the end of the T wave
QTc	QT interval corrected for heart rate
QTcB	corrected QT interval using Bazett's formula
QTcF	corrected QT interval using Fridericia's formula
RR	RR interval
SAE	Serious adverse event
SAP	Statistical analysis plan
SCID	Structured clinical interview for DSM disorder
SD	Standard deviation
SDS	Sheehan disability Scale

Abbreviation	Full Form
SI	International System of Units
SIGH-A	Structured Interview Guide for the HAM-A
SIGMA	Structured Interview Guide for the MADRS
SOC	System organ class
TEAE	Treatment-emergent adverse event
tmax	Time to maximum concentration
UDS	Urine drug screen
ULN	Upper limit of normal
URS	User Requirements Specifications
USP	United States Pharmacopeia
VAS	Visual analog scale
WBC	White blood cells
Y-BOSC-BE	Yale-Brown Obsessive Compulsive Scale Modified for Binge Eating

1. Introduction

This statistical analysis plan (SAP) contains the definitions of analysis populations, derived variables, and statistical methods for the analyses of efficacy and safety data from study Dasotraline SEP360-321. It outlines the analyses to be performed for data collected both through the Week 12 visit and through withdrawal period. It is developed based on the version of the study protocol Version 3.00 dated Nov 17, 2017. This SAP is developed and finalized prior to database lock for unblinding of the clinical database for study SEP360-321.

If additional analyses are required to supplement the planned analyses described in the SAP after the database lock, they may be performed and will be identified in Clinical Study Report (CSR).

2. Study Objectives

2.1. Primary Objective

The primary objective of the study is to evaluate the efficacy of 2 doses of dasotraline (4 and 6 mg/day) compared with placebo in adults with moderate to severe binge eating disorder (BED) as measured by the number of binge days per week. .

2.2. Key Secondary Objective

To evaluate the efficacy of 2 doses of dasotraline (4 and 6 mg/day) compared with placebo in adults with moderate to severe BED as measured by:

- Binge-eating Clinical Global Impression-Severity (BE-CGI-S)
- 4-Week cessation from binge eating defined as a 100% reduction for at least 28 consecutive days in the number of binge eating episodes
- Yale-Brown Obsessive Compulsive Scale Modified for Binge Eating (Y-BOCS-BE)

2.3. Other Secondary Objective

The other secondary objectives of the study are:

- To evaluate the efficacy of 2 doses of dasotraline (4 and 6 mg/day) compared with placebo in adults with moderate to severe BED as measured by:
  - Number of binge episodes per week
  - Proportion of binge eating responders at Week 12 who show ≥ 75% reduction in the number of binge eating episodes
  - Eating Disorder Examination Questionnaire (EDE-Q) modified
  - Sheehan Disability Scale (SDS)
  - Montgomery-Asberg Depression Rating Scale (MADRS)
  - Hamilton Anxiety Rating Scale (HAM-A)
- To evaluate the safety and tolerability of 2 doses of dasotraline (4 and 6 mg/day) using physical examinations, 12-lead electrocardiograms (ECG), vital signs, adverse event (AE) reports, clinical laboratory results, body weight, body mass index (BMI), and Columbia – Suicide Severity Rating Scale (C-SSRS)
- To assess potential withdrawal effects upon abrupt discontinuation of dasotraline after 12 weeks of continuous daily treatment using the following assessments (administered during the withdrawal period):
  - Cocaine Selective Severity Assessment (CSSA)
  - Discontinuation-Emergent Signs and Symptoms (DESS) Scale
  - Symptoms of anxiety utilizing the HAM-A
  - Symptoms of depression utilizing the MADRS

- To assess the abuse potential of dasotraline utilizing a comprehensive abuse potential monitoring plan (APMP)
- To assess the relationship between dasotraline plasma concentration and the primary efficacy and selected secondary efficacy and safety endpoints

3. Study Design

This is a randomized, double-blind, parallel-group, multicenter, outpatient study evaluating the efficacy and safety of 2 doses of dasotraline in adults with BED using dasotraline (4 and 6 mg/day) versus placebo over a 12-week treatment period. This study is projected to randomize approximately 480 subjects to 3 treatment groups in a 1:1:1 ratio (4 mg/day dasotraline, 6 mg/day dasotraline and placebo). Subjects randomized to placebo receive placebo for the duration of the treatment throughout the study. Subjects randomized to 4 mg/day dasotraline receive 4 mg/day for the duration of the treatment period. Subjects randomized to 6 mg/day dasotraline are dosed with 4 mg/day dasotraline for the first 2 weeks of the treatment period and are increased to 6 mg/day at Week 2.

The study consists of 3 periods: Screening (up to 3 weeks), 12-weeks of treatment, and 3-week study drug withdrawal period. Subjects who complete the 12-week double-blind treatment period in this study may be eligible to enroll and continue treatment for an additional 12 months in an open-label extension study (Study SEP360-322). Subjects who do not enter the extension study complete the study drug withdrawal period in this study. See protocol Section 7.1 Figure 1 for study schematic.

Efficacy is evaluated at each visit using the subject binge eating diary and a clinician interview to assess the frequency of binge episodes and the number of binge days defined as days with at least one binge episode. Additional assessments include BE-CGI-S, Y-BOCS-BE, MADRS, HAM-A, SDS and EDE-Q modified. Safety and tolerability are monitored throughout the study by collection of physical examination results, ECGs, vital signs, AEs, clinical laboratory parameters, C-SSRS, body weight, and BMI. Subjects who have significant findings for suicidal ideation upon completion of the C-SSRS at any time during the study must be referred to the investigator for follow-up evaluation. Assessment of potential withdrawal effects is conducted via administration of the CSSA, DESS, HAM-A, and MADRS at the end of treatment and during the 3-week study medication withdrawal period.

Blood samples for pharmacokinetic (PK) and biomarkers associated with eating disorders to be determined are collected. Blood samples collected for PK assessment are analyzed for plasma concentrations of dasotraline. The relationship between dasotraline plasma concentration and the primary and selected secondary clinical outcome measures are evaluated. The relationship between biomarkers associated with eating disorders in relationship to the severity of BED and response to dasotraline will be explored at a later date in the development program. Plasma samples collected for PK and biomarker concentration analysis may also be used for the additional characterization of and/or bioanalytical method development for dasotraline.

A blood sample is collected from subjects who provide separate informed consent for pharmacogenomics (PGx) analysis for potential evaluation of associations between genetic polymorphisms such as, but not limited to Taq1DRD2, OPRM1, and DRD4-7R, and the severity

of BED, the clinical response to dasotraline, in addition to the safety, efficacy, and PK profiles of dasotraline to be conducted at a later date in the development program. Separate consent is required for collection of this specific blood sample and is obtained at Screening for subjects who agree to provide this sample (note: this separate consent is not required for participation in the study, it is required only for subjects who provide this sample).

A comprehensive Abuse Potential Monitoring Plan (APMP) for dasotraline designed to detect potential abuse of the compound and to more closely monitor AEs consistent with the known pharmacology for dasotraline is implemented.

An independent Data and Safety Monitoring Board (DSMB) will review safety data at regular intervals.

The schedule of assessments for this study is in Attachment 9.1; also see protocol Section 11.8 for study visits and assessments detail.

3.1. Determination of Sample Size

The sample size for this study was estimated based on 2 hypotheses associated with the primary efficacy endpoint (change from baseline in number of binge days per week at Week 12). A fixed sequence closed testing procedure was used to adjust for 2 comparisons of dasotraline doses vs placebo (ie, with dasotraline 6 mg vs placebo tested first and dasotraline 4 mg vs placebo tested after if the previous one is significant at the 0.05 level) for the sample size justification. Based on the Study SEP360-221 results, assuming a common standard deviation (SD) of 1.75 and a mean improvement of 0.9 (effect size 0.517) and 0.8 (effect size 0.457) over placebo for change from baseline in number of binge days per week at Week 12 for dasotraline 6 mg/day and 4 mg/day doses respectively, a sample size of 96 subjects per treatment group will provide at least 85% conjunctive power to reject both null hypotheses. An upward adjustment of approximately 40% is assumed to compensate for subjects who are randomized but discontinue from the study, thus, a total sample of 480 subjects (160 subjects per group) will be randomized with a ratio of 1:1:1 for placebo, and dasotraline 4 mg/day and 6 mg/day. The sample size calculation was based on Monte Carlo simulation from EAST 6.4.

3.2. Randomization, Stratification and Blinding

After successfully meeting study entry criteria, subjects are randomly assigned in a 1:1:1 ratio to 1 of 3 treatment arms:

- Dasotraline 4 mg/day
- Dasotraline 6 mg/day
- Placebo

Study protocol V1.00 (dated Feb 23, 2017) and V2.00 (dated May 18, 2017) state that the randomization was to be balanced using permuted blocks with 2 stratification criteria: (1) baseline number of binge eating days per week, which is defined as number of binge days/week

determined from the 2 weeks before the Baseline visit (3 – 4 binge eating days per week, > 4 binge eating days per week); and (2) regions in the United States, which are detailed in the User Requirements Specification (URS) (see Attachment 9.2.2).

The stratification process is handled by an Interactive Response Technology System (IRT), an integrated web-based subject and drug-management system. An IRT is used to manage randomization at Day -1 and, if necessary, for emergency unblinding of treatment assignment during the study.

Study medication is assigned by an IRT at Day -1 based on the randomization schedule. The IRT generates instructions on which medication number to assign to a subject. Each randomized subject is dispensed one 10-day blister pack per scheduled visit up to and including Visit 5 and two 10-day blister packs per scheduled visit thereafter through Visit 9.

This is a double-blind study. Subjects, Investigator staff, persons performing the assessments, clinical operations personnel, data analysts, and personnel at central laboratories remain blind to the identity of the treatment from the time of randomization until database lock (DBL) and unblinding, using the following methods: (1) randomization data are kept strictly confidential until the time of unblinding, and are not accessible by anyone else involved in the study with the following exceptions: external bioanalytical personnel involved in the analysis of PK samples, DSMB members involved in regular review of safety data, external statistician and programmer who prepare materials for DSMB review, and Clinical Trial Materials Management personnel and (2) the identity of the treatments are concealed by the use of study medication that are all identical in packaging, labeling, schedule of administration, and appearance.

During the study and prior to database lock, the treatment assignments for all subjects are provided to the Bioanalytical Contract Research Organization (CRO) to facilitate plasma PK sample handling and analysis. No PK samples are analyzed for placebo subjects.

Protocol and IRT Amendment for Stratification Factor of Region

For this US study, there is no concern of 'region' difference for treatment effect. Sponsor's initial intention of including stratification factor of region was to provide an optimization approach to ensure a balanced treatment assignment within each region. It should be noted that stratification by region was not implemented in study SEP360-221, which is also a US study with a similar study design.

The study team noticed on July 20, 2017 that the stratification by region was not implemented as planned, ie, the initiated sites in IRT were linked to incorrect regions (See Section 6.1.5.1 for details). As of July 20, 2017, among 50 planned sites of this study, a total of 44 sites were screening subjects; and 137 subjects from 35 sites had already been randomized. Due to the 'random' nature of the incorrect assignment of region according to the root cause of this incident, the study team decided to remove stratification criteria related to 'region' to minimize the impact of this mistake. Stratification criteria were amended in study protocol V3.00 and randomization is to be balanced using only permuted blocks with baseline number of binge eating days per

week (3 – 4 binge eating days per week, > 4 binge eating days per week).

4. Changes in Planned Analyses

There is not any change made in this SAP from the analyses planned in the protocol.

5. Efficacy and Safety Variables

Section 6.1.2 provides definitions of DB Baseline and Withdrawal Baseline. For the sake of simplicity, hereafter, Baseline and DB Baseline will be used in an interchangeable way.

5.1. Efficacy Variables

5.1.1. Variables Derived Per Binge Eating Diary

All binge episodes will be captured daily by the subject in a binge eating diary using paper format to record the date, duration (total hours of binging), type of binge episode (meal or non-meal), amount and type of food, and number of binges. The number of binge days per week and the number of binge episodes per week at each assessment timepoint will be derived based on clinician review of the binge eating diary collected in the time intervals as follow:

- Baseline: binge eating diary completed for 14 consecutive days immediately prior to the Baseline visit;
- Weeks 1, 2, 3, and 4 (ie, Visits 3, 4, 5, and 6): binge eating diary collected at each visit, which will span from the previous visit to the current visit and therefore cover Weeks 1, 2, 3, and 4, respectively; Particularly, Week 1 visit period will span from Day 1 (ie, 1st dose day) to one day before Week 1 visit date (inclusive).
- Weeks 6, 8, 10 and 12 (ie, Visits 7, 8, 9, and 10): binge eating diary collected at each visit, which will span from the previous visit to the current visit and therefore cover Weeks 5/6, 7/8, 9/10, and 11/12, respectively.
- Withdrawal Weeks 1, 2, 3 (ie, Week 13, 14, and15): binge eating diary collected at each withdrawal visit, which will span from the previous visit to the current visit and therefore cover Visits 11, 12, and 13, respectively.

Specifically, the diary information collected on the post-DB baseline visits will be counted into the corresponding week span; for an example, diary collected on Week 1 visit data will be counted into the diary data at Week 1; diary collected on Week 2 visit data will be counted into the diary data at Week 2; and so on.

Diary records reviewed at early termination visit will be mapped into scheduled visit. Mapping rule of early terminated visit for binge eating diary based on study day is provided in Section 6.1.4.1.

For very rare situations, a subject might miss a scheduled visit and come back at the next scheduled visit. For such a case, diary records are expected to be collected continuously without interruption. The collected diary will be reviewed on the next scheduled visit. During the double-blind treatment period, if a diary contains binge eating diary records for a missed visit, the diary records will be split into two parts by using the target day of the missed visit (see Table 2 in study protocol) as a data cut: part 1 starts from the previous scheduled visit date (inclusive)

to the target day of the missed visit (exclusive), and part 2 starts from the target day of the missed visit (inclusive) to the current visit (exclusive). Part 1 and Part 2 data will be used to derive binge diary related variables for previous scheduled visit and the missed visit, respectively.

5.1.1.1. Number of Binge Days, Number of Binge Episodes, and Number of Assessed Days During a Visit Period

Number of assessed days at baseline will be 14 days. The number of binge eating days per week at baseline is calculated as the number of binge eating days in the 14 days immediately prior to the Baseline visit multiplied by 7 and divided by the number of days in the Baseline period. No missing diary entry is expected. In case there is a missing diary, the non-missing diary days in 14 days immediately before randomization will be used to derive these variables.

For each post-baseline assessment visit defined in Section 5.1.1,

- Number of assessed days within the visit period is defined as the difference in study day between the current visit and the previous visit (ie, study day of current visit – study day of previous visit). If missing diary entries exist during the assessed period for a subject, the number of assessed days will be derived based on the days with non-missing diaries.
- Number of binge days is counted as the binge days that occur between the previous visit date (inclusive) and a day before the current visit date (inclusive).
- Number of binge episodes is counted in a similar way. If missing diary entries exist during the assessed period for a subject, number of binge days and number of binge episodes will be derived based on the days with non-missing diaries.

5.1.1.2. Number of Binge Days per Week and Number of Binge Episodes per Week for a Visit Period

Based on the number of binge days and number of assessed days (defined in Section 5.1.1.1) at each visit, the number of binge eating days per week is calculated as

(total number of binge days) x 7 / (total number of assessed days in the visit span);

the number of binge eating episodes per week is calculated as

(total number of binge episodes) x 7 / (total number of assessed days in the visit span).

5.1.1.3. Percent Reduction in the Number of Binge Eating Episodes from Baseline at Post-baseline Visits in DB Treatment Period

Percent change in the number of binge eating episodes from baseline at a post-baseline visit is calculated as:

(value at the post-baseline visit – value at baseline) / value at baseline for number of binge eating episodes per week;

Binary indicator for subjects who have ≥ 75% reduction in the number of binge-eating episodes from baseline will be derived for each post-baseline visit in DB treatment period, respectively.

5.1.1.4. 4-Week Cessation

4-Week cessation from binge eating is defined as a 100% reduction for at least 28 consecutive days in the number of binge eating episodes at Week 12/EOT visit. If a subject is early terminated in the double-blind treatment period prior to collection of 28 days of diary data or the subject has missing diary data in the last 28 days prior to Week12/EOT, the subject will be counted as no cessation. See below for specific details about how to derive the variable:

Count the number of days from the first study dose date to the Week 12/EOT visit date,

- If a subject has less than 28 days, 4-week cessation flag will be set as 'No';
- If a subject has ≥ 28 days,
  - If missing diary entries exist during the last 28 days or no missing diary exist in the last 28 days but reduction <100%, 4-week cessation flag will be set as 'No'.
  - If no missing diary entries exist in the last 28 days and reduction =100%, 4-week cessation flag will be set as 'Yes'.

5.1.2. Binge-eating Clinical Global Impression-Severity (BE-CGI-S)

The BE-CGI-S is a single value, clinician-rated assessment of illness severity, and 7-point scale with range from 1 = normal, not at all ill; 2 = borderline mentally ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; 7 = among the most extremely ill subjects. A higher score is associated with greater illness severity.

BE-CGI-S score is collected at Baseline, Weeks 2, 4, 6, 8, 10, and 12/EOT for randomized subjects, and at Week 15 for subjects who enter into the withdrawal period.

5.1.3. Yale-Brown Obsessive Compulsive Scale Modified for Binge Eating (Y-BOCS-BE)

Yale Brown Obsessive Compulsive Scale Modified for Binge Eating (Y-BOCS-BE) is a clinician-rated scale that measures the obsession of binge eating, thoughts and compulsiveness of binge eating behaviors. 10 questions are included in this assessment.

1. Time occupied by obsessive thoughts to binge eat
2. Interference due to thoughts to binge eat
3. Distress associated with thoughts to binge eat
4. Resistance of thoughts to binge eat

5. Degree of control over obsessive thoughts to binge eat
6. Time spent on compulsive behaviors to binge eat
7. Interference due to binge eating
8. Distress associated with binge eating
9. Resistance
10. Degree of control over binge eating

The 10-item scale, is rated from 0 (no symptoms) to 4 (extreme symptoms). Total scores range from 0 to 40. A score of 0-7 is sub-clinical; 8-15 is mild; 16-23 is moderate; 24-31 is severe; and 32-40 is extreme.

Y-BOCS-BE is divided into 2 subscales: obsessions and compulsions. The Y-BOCS-BE obsessions subscale is calculated as the sum of raw scores from obsessions questions (questions 1 to 5); and the subscale score ranges from 0 to 20.The compulsions subscale is calculated as the sum of raw scores from compulsions questions (questions 6 to 10); and the subscale score ranges from 0 to 20. Higher values of Y-BOCS-BE score and subscale scores represent greater severity of illness.

If there are any missing responses on the YBOCS-BE, the total score and corresponding subscale score are set as missing. Y-BOCS-BE score is collected at Baseline, Weeks 2, 4, 6, 8, 10, and 12/EOT for randomized subjects, and at Week 15 for subjects who enter the withdrawal period.

5.1.4. Eating Disorder Examination Questionnaire (EDE-Q) and Eating Disorder Examination Questionnaire (EDE-Q) Modified

The Eating Disorder Examination Questionnaire (EDE-Q) is a self-report version of the eating disorder examination (EDE) (Fairburn-1994). Like the EDE, the EDE-Q measures eating-disorder psychopathology in the past 28 days and over longer intervals for diagnostic items. EDE-Q generates two types of data. First, it provides frequency data (ie, numerical data) on key behavioral features of eating disorders in term of number of episodes of the behavior and in some instance number of days which behavior has occurred; Second, they provide ratings (ie, categorical data), 4 subscale scores, and a EDE-Q global score reflecting the severity of aspects of the psychopathology of eating disorders. The 4 subscale scores are Restraint, Eating Concern, Shape Concern, and Weight Concern. Severity rating ranges 0- 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, 6=’every day’.

Both of the Eating Disorder Examination Questionnaire Brief Version (EDE-Q7) and EDE-Q modified (EDE-QM) are brief versions of the EDE-Q. EDE-Q7 was used in study SEP360-221 while EDE-QM is used in this study (SEP360-321).

EDE-Q7 comprises 7 items to generate a global score and 3 subscale scores (Restraint, Shape Concern, and Weight Concern). The EDE-Q7 along with 3 items (see below) to assess binge eating, including the number of binge eating days, is referred to as the EDE-Q modified.

- **Item 4:** Over the past 28 days, how many **times** have you eaten what other people would regard as an **unusually large amount of food** (given the circumstances)?
- **Item 5:** On how many of these times did you have a sense of having lost control over your eating (at the time that you were eating)?
- **Item 6:** Over the past 28 days, on how many **DAYS** have such episodes of overeating occurred (i.e., you have eaten an unusually large amount of food **and** have had a sense of loss of control at the time)?

See Table 1 below for definition of rating data for EDE-Q, EDE-Q7, EDE-Q modified, and their relationship. Table 2 provides the definition of frequency data for EDE-Q, EDE-Q modified, and their mapping relationship.

Table 1: Rating Data of EDE-Q , EDE-Q7, and EDE-Q modified Items

EDE-Q Modified Item	EDE-Q7 Item	EDE-Q Item	Abbreviated Content	Subscale
1	**1**	**1**	**Restraint over eating**	Restraint
		2	Avoidance of eating	
2	**2**	**3**	**Food avoidance**	
3	**3**	**4**	**Dietary Rules**	
		5	Empty stomach	
		7	Preoccupation with food, eating or calories	Eating Concern
		9	Fear of losing control over eating	
		19	Eating in secret	
		21	Social eating	
		20	Guilt about eating	
		6	Flat stomach	Shape Concern
		8	Preoccupation with shape or weight	
8	**5**	**23**	**Importance of shape**	
		10	Fear of weight gain	
10	**7**	**26**	**Dissatisfaction with shape**	
		27	Discomfort seeing body	
		28	Avoidance of exposure	
		11	Feelings of fatness	
7	**4**	**22**	**Importance of weight**	Weight Concern
		24	Reaction to prescribed weighing	
		8	Preoccupation with shape or weight	
9	**6**	**25**	**Dissatisfaction with weight**	
		12	Desire to lose weight	

Reference: Fairburn C.G., 2014, EDE Edition 17.D, Cognitive Behavior Therapy and Eating Disorders[1]

Table 2: Frequency Data of EDE-Q and EDE-Q modified Items

EDE-Q Modified Item	EDE-Q Item	Abbreviated Content
4	13	Times of unusually large amount of food
5	14	Times of having lost control eating
6	15	Days of overeating episodes
	16	Times of having made of sick for controlling shape or weight
	17	Times of having taken laxatives for controlling shape or weight
	18	Times of having exercised in a "driven" or "compulsive" for controlling shape or weight

For EDE-Q, EDE-Q Modified, and EDE-Q7, the subscale scores are calculated as average of the corresponding individual scores. If any individual score of the subscale is missing, the corresponding subscale score will be set as missing and no imputation will be done. An EDE-Q global score is calculated as average of 4 EDE-Q subscale scores. An EDE-Q7 global score is calculated as average of 3 EDE-Q7 subscale scores, and an EDE-Q modified global score is calculated as average of 3 EDE-Q modified subscale scores. The EDE-Q global score, EDE-Q modified global score, and EDE-Q7 global score will be missing if any of corresponding subscale score is missing. For EDE-Q, EDE-Q modified, and EDE-Q7, higher scores of the global score and subscale scores indicate greater levels of symptomatology. The global scores and subscale scores of EDE-Q modified and EDE-Q7 are equivalent.

EDE-Q is collected only at screening visit. EDE-QM is collected at Baseline and Week 12/EOT. Both EDE-Q and EDE-QM are designed to collect the data in the past 28 days; for a subject who is early terminated from the treatment period, these questionnaires may not be appropriate since the subject may have much less days since last visit compared to the targeted 28 days. In order to collect more accurate information for early terminated subjects, a modified version of EDE-QM 28 days version, named as EDE-QM 14-days look-up version, having a very similar rating scale as the regular EDE-QM, will be applied on early termination subjects who discontinue between 12 and 24 days, inclusive, following the previous administration of the EDE-QM. For those who discontinue more than 24 days following the previous visit, the regular EDE-QM (28 day lookback version) is administered. Data collected by using EDE-QM 14-days look-up version and regular version (ie, 28 days look-up version) will be summarized together.

5.1.5. Sheehan Disability Scale (SDS)

The Sheehan Disability Scale (SDS) is to assess functional impairment in 3 domains: (1) work/school; (2) social life; and (3) home life. The subject rates the extent to which work/school, social life and home life have been impaired by his/her symptoms on a 10-point anchored visual analog scale (VAS). The subject's responses for each domain are rated on the following scale: 0 = not at all; 1-3 = mildly; 4-6 = moderately; 7-9 =markedly; 10 = extremely. The 3 items can be combined into a single global measure of impairment (SDS total score) that ranges from 0 (unimpaired) to 30 (highly impaired). If any of the 3 SDS subscale scores is missing at a visit,

the SDS total score for that visit is also missing. A higher subscale score and total score are associated with greater illness severity.

The SDS is collected at Baseline, Weeks 6, and 12/EOT for randomized subjects, and at Week 15 for subjects who enter the withdrawal period.

5.1.6. Montgomery-Asberg Depression Rating Scale (MADRS)

The MADRS is a clinician-rated assessment of the subject's level of depression. This scale 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, with higher scores indicating increased depressive symptoms. The score for each question will be summed to compute a total score ranging from 0 to 60. If the response to any question is missing, the total score will be missing.

The MADRS is collected at screening, Baseline and Week 12/EOT for randomized subjects, and at Weeks 13, 14 and 15 for subjects who enter the withdrawal period.

5.1.7. Hamilton Anxiety Rating Scale (HAM-A)

The HAM-A is a rating scale developed to quantify the severity of anxiety symptomatology. It consists of 14 items, each defined by a series of symptoms. The 14 items are: anxious mood, tension, fears, insomnia, intellectual, depressed mood, somatic (muscular), somatic (sensory), cardiovascular symptoms, respiratory symptoms, gastrointestinal symptoms, genitourinary symptoms, autonomic symptoms, and behavior at interview. Each item is rated on a 5-point scale, ranging from 0 (not present) to 4 (severe/disabling). The score for each question are summed to compute a total score ranging from 0 to 56. If the response to any question is missing, the total score is missing. A higher score is associated with a greater degree of anxiety. A score of ≥ 14 has been suggested to indicate clinically significant anxiety.

The HAM-A is collected at Baseline and Week 12/EOT for randomized subjects, and at Weeks 13, 14 and 15 for subjects who enter the withdrawal period.

5.2. Safety Assessment

The safety assessments/endpoints are

- adverse event (AE) reports, including the incidence of overall AEs, serious AEs (SAEs), and AEs (or SAEs) leading to discontinuations
- Clinical laboratory evaluations (serum chemistry, hematology, and urinalysis)
- Clinical evaluations (vital signs including orthostatic effects, and 12-lead ECGs)
- weight and body mass index (BMI)
- concomitant medication reports
- physical examinations
- Columbia – Suicide Severity Rating Scale (C-SSRS)

The following safety measurements will also be collected to assess the occurrence of symptoms of withdrawal from dasotraline:

- Cocaine Selective Severity Assessment (CSSA)
- Discontinuation-Emergent Signs and Symptoms (DESS) Scale
- Symptoms of anxiety utilizing the HAM-A
- Symptoms of depression utilizing the MADRS

5.2.1. Columbia – Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is an instrument designed to systematically assess and track suicidal behavior and suicidal ideation for life time, one month prior to screening and throughout the trial. The strength of this suicide classification system is in its ability to comprehensively identify suicidal events while limiting the over-identification of suicidal behavior. The C-SSRS includes the following four sections: Suicidal Ideation, Ideation Intensity, Suicidal Behavior and Actual Suicide Attempts. The C-SSRS is administered by a trained rater at the site.

Severity of suicidal ideation is rated on a 6-point scale from 0=’No ideation present’ to 5=’Active ideation with plan and intent’. A score of 4 or 5 on this scale indicates serious suicidal ideation. Subjects with Type 4 (active suicidal ideation with some intent to act, without specific plan) or Type 5 (active suicidal ideation with specific plan and intent) suicidal ideation during the study will be discontinued from the study and referred to a mental health professional. Suicidal ideation score is defined as the most severe suicidal ideation score (1-5 on the CSSRS) present at the assessment; assign a score of 0 if no ideation is present. The ideation intensity total score is calculated as summation of the five items from the Ideation Intensity scale: frequency, duration, controllability, deterrents, and reasons for ideation. If one or more of these 5 items are missing at a visit, no imputation will be made and the corresponding ideation intensity total score will be set as missing. Items corresponding to frequency and duration are each rated on a scale from 1 to 5 and items corresponding to controllability, deterrents, and reasons for ideation are each rated on a scale from 0 to 5. If the patient did not endorse any suicidal ideation, set the intensity rating to 0. Thus, the possible range for the intensity total score is 0 to 25.

Suicidal behavior is collected as presence/absence of actual attempts, non-suicidal self-injurious behavior, interrupted attempts, aborted attempts, preparatory acts or behavior, and any suicidal behavior. In addition, the number of actual attempts, interrupted attempts, and aborted attempts is captured. Any attempt will be defined as suicidal behavior.

The lethality associated with actual attempts is rated on a 6-point scale from 0=’No physical damage or very minor physical damage’ to 5=’Death’. Potential lethality of attempts is rated on a 3-point scale from 0=’Behavior not likely to result in injury’ to 2=’Behavior likely to result in death despite available medical care’.

The C-SSRS is collected at screening, Baseline, Weeks 1, 2, 3, 4, 6, 8, 10 and 12/EOT for randomized subjects, and at Weeks 13, 14 and 15 for subjects who enter the withdrawal period.

This study utilizes 2 versions of the C-SSRS. At the screening visit, the "Baseline/Screening" version will be completed. C-SSRS at 'Baseline/Screening' consists of two parts: part 1 – life time; part 2 – past 1 month. For all subsequent visits, the "Since Last Visit" version of the C-SSRS is administered.

5.2.2. Withdrawal Assessment for Physical Dependence and Withdrawal Symptoms

Subjects who either are early terminated in treatment period or complete 12-week treatment period but not enter into extension study need to complete the study medication withdrawal period assessments and visits.

Assessment of potential withdrawal effects will be conducted via administration of the CSSA, DESS, HAM-A, and MADRS at Week 12 (End of Treatment), then again at Weeks 13, 14, and 15 during the 3-week study medication withdrawal period. The CSSA and DESS will also be completed during the 3-week study medication withdrawal period on non-clinic visit days. Clinical site staff will call the subject every other day during the study medication withdrawal period, beginning the second day after the last dose of study drug, unless a clinic visit is scheduled for that day. Clinical site staff will record the responses in the subject's source information and in the CRF with the contact date and time. A window of ± 1 day is allowed for each weekly clinic visit and telephone call. If the subject cannot be contacted on a given day, during the next contact the study site staff will retrospectively collect the missed CSSA and DESS, in addition to the current day's CSSA and DESS.

5.2.2.1. Cocaine Selective Severity Assessment (CSSA)

The CSSA is a clinician-administered scale designed to evaluate withdrawal signs and symptoms related to stimulants over the past 24 hours. Included in the CSSA are those symptoms most often associated with early cocaine abstinence, including change in appetite, depression, fatigue, anhedonia, anxiety, irritability, sleep disturbance, and inability to concentrate, paranoia, carbohydrate craving, bradycardia, and suicidality. The CSSA scale contains 18 items. All items except item 14, are scored 0 to 7 according to instructions generally with 0 = no symptoms and 7 = maximum score on any individual item. Item 14 is scored 0 to 8 with 0 = no symptoms and 8 = maximum score. The score for each question will be summed to compute a total score ranging from 0 to 127 with higher score indicates severe withdrawal signs and symptoms. If the response to any question is missing, the total score will be missing.

The CSSA is collected at clinic visits: Week 12/EOT, and Weeks 13, 14 and 15 for subjects who enter the withdrawal period, as well as at phone contacts during the withdrawal period, beginning the second day after the last dose of study drug.

5.2.2.2. Discontinuation-Emergent Signs and Symptoms (DESS) Scale

The DESS Scale is a clinician-rated instrument of 43 items used to evaluate signs and symptoms associated with discontinuation or interruption of monoamine reuptake inhibitor treatment. Each of the 43 signs and symptoms is assessed and placed in one of the following categories: new symptom, old symptom but worse, old symptom but improved, old symptom but unchanged, and symptom not present. Each sign and symptom that is placed in the category new symptom or in the category old symptom but worse receives 1 point. All other categories receive 0 points. The total DESS score is the sum of the number of points, which corresponds to the number of symptoms rated as either a new symptom or as an old symptom but worse. The score ranges from 0 to 43, where higher scores indicate more discontinuation-emergent signs and symptoms.

The DESS is collected at clinic visits: Week 12/EOT, and Weeks 13, 14 and 15 for subjects who enter the withdrawal period, as well as at phone contacts during the withdrawal period, beginning the second day after the last dose of study drug.

5.2.2.3. HAM-A

See Section 5.1.7

5.2.2.4. MADRS

See Section 5.1.6

5.3. Other Assessment

Pharmacokinetic (PK) endpoints include dasotraline plasma concentration which is collected at Weeks 2, 6 and 12/EOT and biomarkers concentration which is collected at Baseline and Week 12/EOT visit.

PK analysis will be based on the PK population. Subjects will be analyzed based on the last treatment prior to the collection of PK sample. See Section 6.1.7 for definition of PK population.

The relationship between SEP-225289 plasma concentration and the primary and selected secondary clinical outcome measures of SEP-225289 using population PK methods will not be done by INC Research and the analysis is not included in the SAP.

6. Statistical Methods

6.1. General Analysis Definition

6.1.1. Analysis Period and Study Day

Two scopes of data for a subject are considered in this study. Each scope of data consists of a specific portion (subset) of all study data for that subject.

6.1.1.1. Double-Blind Treatment Period (On Treatment)

The double-blind treatment period will begin with the day when the first dose of double-blind study medication is taken, which will be denoted as study Day 1. The end of the double blind phase is defined as the maximum of the last date of administration of study medication and the withdrawal completion date of the treatment period (ie, Week 12/EOT visit date). If the end date of the medication intake is unknown (eg, subject is lost to follow-up), the end of the double blind period is defined as the withdrawal date provided by investigator site. If a subject does not take any study medication, start and end of the treatment period are set to missing.

The start and end of treatment period are given below:

- **Start of treatment period** = Date/time of the first study medication administration during the double-blind treatment period
- **End of treatment period** = maximum of the last date of administration of study medication and the withdrawal/ completion date (ie, Week 12/EOT visit date)

Each subject's assessment at each visit in pre-treatment and treatment period will be assigned a study day with respect to start date of the treatment period.

- Study Day = Assessment date − start date of the treatment period + 1; if assessment date ≥ start date of the study period
- Study Day = Assessment date − start date of the treatment period; if assessment date < start date of the study period

6.1.1.2. Withdrawal Period (Off Treatment)

For subjects who either are early terminated from DB period or complete DB period but do not enter the extension, withdrawal period analyses will be generated; no withdrawal period analyses will be produced for subjects entering the extension study.

The withdrawal period will begin with immediately after the last study dose. Study day relative to the last study dose date will be denoted as Day 0. The end of the withdrawal period is defined as the maximum of last study visit (scheduled or unscheduled) and last phone contact after the

last study dose. If the last study dose date is missing (eg, lost follow-up subjects), start and end days will be set as missing.

Each subject's assessment at each visit (site or phone contact) in the withdrawal period will be assigned a study day with respect to start date of the withdrawal period.

- Study Day = Assessment date − start date of the withdrawal period if assessment date ≥ start date of the withdrawal period

6.1.1.3. Overall Study Period

Overall study period is defined as treatment period for subjects who complete the 12-week treatment period and enter the extension study, and a combination of treatment and withdrawal periods for subjects who either are early terminated during the treatment period or complete the 12-week treatment period but do not enter into extension study. See Table 3 below for summary of definition of all analysis periods and study day used for this study.

Table 3: Definition of Analysis Study Period

Analysis Study Period	Start Date	End Date	Applicable Scope
Double-Blind Treatment Period	Day 1*	EOT [a]	All subjects
Withdrawal Period	Day 0 **	EOS [b]	Subjects who either are early terminated in treatment period or complete 12-week treatment period but do not enter into extension study
Overall Study Period	Day 1* of treatment period	EOT	Subjects who complete 12-week treatment period and enter extension study.
		EOS	Subjects who either are early terminated in treatment period or complete 12-week treatment period but do not enter into extension study

Note: *: the day of the first dose of double-blind study medication.
[a] : Week 12 visit/ early withdrawal visit
**: the day of the last dose of double-blind study medication.
[b] : last scheduled or unscheduled (site/phone contact) visit in withdrawal period

Except for adverse event data, summaries and analyses of all data on double-blind treatment are based on the portion of study data that fall in the treatment period (defined in Section 6.1.1.1). For adverse events, see Section 7.4.2 for how a study subject is considered to be on treatment during the treatment period (ie, last study medication administration + 7 days for AE or +14 days for Serious AE).

6.1.2 Baseline

Two types of Baseline will be defined: DB Baseline and Withdrawal Baseline. DB baseline is also referred as Baseline by default in this document.

6.1.2.1 DB Baseline for Assessment Based on Diary Data

See Section 5.1.1 for definition of Baseline for eating diary based data.

6.1.2.2 DB Baseline for Other Data

Baseline is defined as the pre-dosing measure on Day 1. If it is missing, unless otherwise specified, it will be imputed by the latest non-missing measurement prior to Day 1.

6.1.2.3 Withdrawal Baseline for Assessment Based on Diary Data

Withdrawal Baseline is defined as the derived value at Week 12/EOT. See Section 5.1.1

6.1.2.4 Withdrawal Baseline for Other Data

See Section 6.1.4.2 for the definition of Withdrawal Baseline. No imputation will be made if the withdrawal Baseline value is missing.

6.1.3 Endpoint

6.1.3.1 Week 12 Endpoint

Week 12 Endpoint is defined as the last post-baseline non-missing assessment in the treatment period. Hereafter, for the sake of simplicity, Week 12 Endpoint will be referred as Endpoint by default.

6.1.3.2 Withdrawal Endpoint

Withdrawal Endpoint is defined as the last non-missing post-withdrawal baseline assessment in the withdrawal period (regardless of clinic or non-clinic visits).

6.1.4 Visit Windows

6.1.4.1 Early Termination (ET) Visit in Double-Blind Treatment Period

All data are organized and analyzed according to the scheduled times as outlined in the protocol and by the visit denoted on the electronic case report form (eCRF), except for early termination (ET) visits during the treatment period. If a subject terminates early during the DB treatment period, the early termination visit (ie, EOT visit) will be re-mapped using the study day of the visit in Table 4, which applies on both efficacy and safety data per a specific rule associated with that measurements except of the variables derived from binge eating diary. Table 5 describes the remapping rules applied to diary data at ET visits.

Table 4: Visit Window for Early Termination Visit in Treatment Period for Non-Binge Eating Diary Data

Last Scheduled Visit	Time Interval (Day)[a]	Analysis Visit (label on output)
Baseline *	≥ DLSV+1	Week 1
Week 1	≥ DLSV+1	Week 2
Week 2	≥ DLSV+1	Week 3
Week 3	≥DLSV+1	Week 4
Week 4	≥ DLSV+1	Week 6
Week 6	≥ DLSV+1	Week 8
Week 8	≥ DLSV+1	Week 10
Week 10	DLSV+1 – EOT	Week 12
[a] Relative to the day of the first dose of double-blind study medication DLSV=study day of last scheduled visit relative to the first dose of study medication *: for the study day of the ET visit > Day1 Note: Applicable scope of Table 4: vital sign, weight and BMI, C-SSRS, HAM-A, MADRS, EDE-QM, lab, ECG, PK, neurological exam, and breath alcohol test.		
BE-CGI-S and Y-BOCS-BE: if the last scheduled visit is Baseline or Week 1, the ET visit will be mapped to Week 2; if the last scheduled visit is Week 2 or Week 3, the ET visit will be mapped to Week 4; if the last scheduled visit is Week 4 or after, the ET visit will be mapped same as Table 4 **SDS**: if the last scheduled visit is prior to Week 6, the ET visit will be mapped to Week 6; if the last scheduled visit is at/after Week 6, the ET visit will be mapped to Week 12.		

Only protocol planned visits for each measurement (ie, Baseline, Weeks 1, 2, 3, 4, 6, 8, 10, and 12 as appropriate) are included in the by-visit analyses. That is to say, the unscheduled visits will not be included in the by-visit summary; however, measurements from this/these additional visit(s) will be included in the safety analyses (eg, markedly abnormal value in treatment period, the worst case analysis etc.) to assure all on-treatment measurements are included.

Table 5: Visit Window for Early Termination Visit in Treatment Period for Binge Eating Diary Data

Last Scheduled Visit	Time Interval (Day)[a]	Analysis Visit (label on output)
Baseline *	DLSV – DLSV+30	Week 1
Week 1	DLSV – DLSV+30	Week 2
Week 2	DLSV – DLSV+30	Week 3
Week 3	DLSV – DLSV+30	Week 4
Week 4	DLSV – DLSV+40	Week 6

Week 6	DLSV – DLSV+40	Week 8
Week 8	DLSV – DLSV+40	Week 10
Week 10	DLSV – DLSV+40	Week 12

[a] Relative to the day of the first dose of double-blind study medication
DLSV=study day of last scheduled visit relative to the first dose of study medication
*: for the study day of the ET visit > Day 1

6.1.4.2 Visit Window for Visits in Withdrawal Period

Visits within withdrawal period will be mapped per visit window defined in Table 6.1 below for efficacy and safety data collected in withdrawal period except of binge eating diary data. For diary data, see Section 5.1.1 for data derivation rule.

Table 6.1: Visit Window for Clinic Visits in Withdrawal Period for Non-Binge Eating Diary Data

Variables to be Applied	Time Interval (Day)[a]	Target Time Point (Day)	Analysis Visit (label on output)
CSSA, DESS	–5 to 1 [(b)]	**0**	Withdrawal Baseline
	2 – 10 [(c)]	7	Withdrawal Week 1
	11 – 17[(c)]	14	Withdrawal Week 2
	18 – EOS[(c)]	21	Withdrawal Week 3
Other	–5 to 3	**0**	Withdrawal Baseline
	4 – 10	7	Withdrawal Week 1
	11 – 17	14	Withdrawal Week 2
	18 – EOS	21	Withdrawal Week 3

[a] Relative to the day of the last dose of double-blind study medication
[b] per phone contacts or clinic visits
[c] for clinic visits only.

If a subject has 2 or more visits in one analysis visit window, the visit closest to the targeted day will be used as the study visit for that analysis visit window. The other additional visit(s) will be denoted as an unscheduled visit and not included in the by-visit descriptive summary analyses. However, measurements from this/these other additional visit(s) may be included in the other safety analyses to assure all off-treatment measurements are included. If the 2 study visits occur with the same number of days from the target day within the same visit window, the later one will be considered as the study visit for that target day.

Non-clinic phone contact visits within withdrawal period will be mapped per visit window defined in Table 6.2 below for CSSA and DESS specifically.

Table 6.2: Visit Window for Non-Clinic/Phone Contact Visits in Withdrawal Period

Variables to be Applied	Time Interval (Day)[a]	Target Time Point (Day)	Analysis Visit (label on output)
CSSA, DESS	1 – 2	**1**	Day 1-2
	3 – 4	3	Day 3-4
	5 – 6	5	Day 5-6
	7 – 8	7	Day 7-8
	9 – 10	9	Day 9-10
	11 – 12	11	Day 11-12
	13 – 14	13	Day 13-14
	15 – 16	15	Day 15-16
	17 – 18	17	Day 17-18
	19 – 20	19	Day 19-20
	21 – 22	21	Day 21-22
	23 – 24	23	Day 23-24
	25 – EOS[(b)]	25	≥ Day 25

[a] Relative to the day of the last dose of open-label study medication
[b] per phone contacts visits only.

6.1.5 Protocol Amendment for Region Stratification

6.1.5.1 Stratification Error of 'Region'

The original protocol (version 1.0) and amendment 1 (version 2.0) stated that the randomization was to be balanced using permuted blocks with 2 stratification criteria: (1) baseline number of binge eating days per week and (2) regions in the United States, which were defined in the Interactive Response Technology System (IRT) User Requirements Specification and is provided in Attachment 9.2.2. For sake of convenience, hereafter in this document, terms of IXRS and IRT are used interchangeably.

In total 50 sites participated in this study. The study team noticed on July 20, 2017 that stratification by region was not implemented as planned, that is to say, the initiated sites in IXRS were linked to incorrect regions. As of July 20, 2017, a total of 44 sites were screening subjects and 137 subjects from 35 sites had already been randomized. A full investigation was immediately requested by Sponsor to IRT vendor (Oracle). According to the incident report (Incident Number: INC000004665171) provided by Oracle, the root cause of this mistake was with a third party software tool used to develop the IXRS system; more specifically, the mistake was caused by the software's component related to a dropdown list for "region" selection . Region field should have been populated with the set of initial sites that Sponsor provided to Oracle and then should have been populated by Sponsor for all other sites. Once a site was added, the "region" selection should have been disabled. Due to the unexpected behavior of the software component, the "region" value was sometimes overwritten. Based on this investigation

report, in one scenario, a consistent pattern for region selection was identified: when a new site was added in "Manage Sites", the "region" dropdown was loaded as a blank value, and eventually the "region" value assignment would be moved up by one row in the database. In other scenarios, a consistent pattern for region selection was not identified. IXRS vendor noted that there was no way for them to reproduce how the data was assigned in this trial.

6.1.5.2 IXRS Update

Although unexpected region assignment of sites may not be truly random in a statistical sense, region allocation appears inconsistent and no clear pattern was found. To minimize the impact of the region strata error on data interpretation, the study team decided to amend the IXRS and study protocol to remove the regional stratification factor due to the following considerations.

(1) Once a site was initiated in the IXRS system, its associated region was determined and could not be changed. When the mistake was found, 44 of the 50 planned sites had been incorrectly linked to incorrect regions; by the time that IXRS eventually could have been fixed, 47 out the 50 sites would have been linked to incorrect regions. This means that about 85%~90% randomized subjects were to have been linked to incorrect region. Therefore, the regional stratification of subjects would no longer have been meaningful.
(2) Incorrectly linked regions of the study sites show no overall consistent patterns according to the investigation report issued by the IXRS vendor. Two-dimensional distribution of sites by initial planned region in IRT User Requirements Specifications (URS) vs. the region actually assigned by IXRS (see Attachment 9.2.3) seems to support that regions were assigned nearly randomly before IXRS amendment.
(3) Initial purpose of stratification by region was solely to balance the treatment allocation within a region for this United States study. Differences in treatment response among US regions were not expected, nor was the sponsor's intention to detect potential US regional differences. The stratification strategy was chosen for optimization purposes and should not be treated as necessary.

IXRS URS was amended and signed off on Aug 09, 2017 (V7.0). The updated IXRS went alive on Aug 23, 2017. By removing the stratification factor 'Region', study SEP360-221 essentially adopted the same stratification criterion as study SEP360-221. Data analysis methods in protocol amendment 2 (V3.0) of this study are similar to what was planned and performed for study SEP 360-221.

6.1.6 Pooling Strategy for Analysis of Centers

Study SEP360-321 is a multicenter study conducted in the United States. In principle, all centers with 18 or fewer subjects are pooled based on geographic proximity if necessary, with the intention that no pooled center contains more than 36 randomized subjects. The list of pooled centers for the study based on the pooling strategy described above is provided in Attachment 9.2.1 and finalized prior to database lock of the study.

Similarly, planned region and actual region (defined in Section 6.1.7) will be pooled based on geographic proximity as well if a region has 18 or fewer subjects. The list of pooled regions is provided Table 6.3 in Section 6.1.7.

6.1.7 Definition of Planned Region Group and Actual Region Group ('Modified' Region Strata)

To assess the robustness of primary and key efficacy analyses, two region-related exploratory efficacy analyses for primary and key secondary endpoints will be conducted. For each randomized subject, two region variables are defined as below:

- **Planned region group**: unless otherwise specified, it is defined per URS of original IRT (see Attachment 9.2.2).
- **Actual region group (**or equivalently**,** 'Modified' region strata): defined as

Case 1: for subjects randomized PRIOR to the IXRS update:

Actual region is equal to the region strata value.

Case 2: for subjects randomized AFTER the IXRS update and from an 'old' site, where an old site is defined as the one with at least one subject randomized prior to the IXRS update (see Attachment 9.2.3):

Actual region is same as the corresponding region strata value of the site per IXRS data.

Case 3: for subjects randomized AFTER the IXRS update and from a 'new' site, where a new site is defined as the one without any subject randomized prior to the IXRS update but having at least one subject randomized after the IXRS update (see Attachment 9.2.3):

Actual region will be assigned with the corresponding planned region value per URS of original IXRS if location of the new site (ie, State) was listed in the URS of original IXRS (see Attachment 9.2.2); otherwise, the planned region and actual region will be assigned as 'Undefined' since the new site is not listed in URS of original IXRS.

Numbers of subjects within each planned and actual region are listed in Table 6.3 as below. As of the completion of study enrollment, there is one site (site 067), which has 3 randomized subjects, from a location that was not listed in original URS of IXRS. In addition, the actual

region 'Central 2' is small and has only 3 subjects. To make sure exploratory analyses related to the planned region and actual region are feasible, small planned and actual regions will be pooled for analysis purposes. Specifically, for planned region, site '067' will be pooled with the initial planned region 'Central 1' to generate a pooled planned region still named as 'Central 1'; no pooling is needed for other planned regions; for actual region, site '067' and initial actual region 'Central 2' will be pooled with the initial actual region 'Central 1' to generate a pooled actual region named as 'Central1/Central 2'; no pooling is needed for other actual regions. Numbers of subjects within each pooled planned and actual region are listed in Table 6.3 as well and are finalized prior to database lock of the study. The exploratory efficacy analyses are to be performed as needed by replacing term of 'pooled center' with pooled actual region or pooled planned region group in the corresponding statistical analysis to further corroborate primary and key secondary efficacy results.

Table 6.3 Pooled Planned and Actual Region Groups Used for Exploratory Efficacy Analyses

Planned Region Group		**Pooled Planned Region Group**		**Actual Region Group**		**Pooled Actual Region Group**	
Region Name	**Number of Subjects**	**Region Name**	**Number of Subjects**	**Region Name**	**Number of Subjects**	**Region Name**	**Number of Subjects**
Northeast	59	Northeast	59	Northeast	127	Northeast	127
East	75	East	75	East	37	East	37
Central 1	50	**Central 1** [(a)]	**53**	Central 1	138	**Central 1 /Central 2**[(b)]	**144**
Central 2	69	Central 2	69	Central 2	3		
Southeast 1	55	Southeast 1	55	Southeast 1	62	Southeast 1	62
Southeast 2	22	Southeast 2	22	Southeast 2	17	Southeast 2	17
West 1	88	West 1	88	West 1	24	West 1	24
West 2	70	West 2	70	West 2	80	West 2	80
Undefined	3			Undefined	3		

Note: (a): Site 067 from Michigan (undefined in URS of IXRS) is pooled with initial region 'Central 1'.
(b): Site 067 and actual region 'Central 2' are pooled with actual region 'Central 1'.

6.1.8 Adjustments for Covariates

MMRM model and ANCOVA model contain baseline as a covariate to adjust baseline. Other covariates may be included for modeling purposes, as specified.

6.1.9 Analysis Populations

The Intent-to-Treat (ITT) Population: All randomized subjects who receive at least one dose of study medication and have at least 1 post-baseline result in any efficacy evaluation. The ITT population is the primary population for efficacy analyses, ie, unless otherwise specified, all efficacy analyses are based on ITT population.

The Per Protocol (PP) Population: All ITT subjects who have no major protocol violations that may affect the interpretation of the primary efficacy endpoint. The PP population will be used to assess robustness of the primary efficacy analysis. Major protocol violations and the PP population will be determined through a blinded data review meeting (BDRM) and identified prior to the database lock.

Subjects in the ITT and PP populations will be analyzed based on the treatment to which they are randomized.

The Double-blind (DB) Safety Population: All randomized subjects who receive at least one dose of study medication. The DB safety population mainly is used for the data analyses of the safety data collected through Week 12 or during the treatment period. Subjects are analyzed based on the actual treatment received; specifically, for subjects who are randomized to dasotraline 6 mg group and dosed with dasotraline 4 mg but never titrated to dasotraline 6 mg (ie, early terminated on or prior to Week 2 visit), the actual dose will be dasotraline 4 mg; otherwise, actual dose will be set as dasotraline 6 mg; In the event that a subject receives a treatment other than the one to which he/she is randomized or expected, the actual treatment is defined as the treatment to which the subject is exposed for the greatest duration during the treatment period. The actual treatment is generally the same as the randomized treatment group, unless the subject takes incorrect/unexpected dose during the entire study. Hereafter, DB safety population will be referred as safety population by default.

The Withdrawal Population: The withdrawal population includes all randomized subjects who receive at least one dose of study medication, and either discontinues the study drug during the treatment period before the Week 12 visit or completes the 12-week treatment period but do not enter the extension study (Study SEP360-322), and have at least 1 assessment after the last study drug for any safety or efficacy evaluation. The withdrawal population mainly is used to summarize the safety/efficacy assessments collected in the withdrawal period except of adverse event data.

The Modified Withdrawal Population: all subjects in the withdrawal population plus the subjects who do not enter the extension study and do not enter the withdrawal period, but the last study dose dates are at least 2 days earlier than the last study visit. The modified withdrawal population will be mainly used for adverse event that could be retrospectively collected.

Withdrawal analyses will be presented by planned treatment group for efficacy data and by actual treatment group for safety data.

Pharmacokinetics (PK) Population: The PK population is defined as all subjects who are randomized, receive at least 1 dose of study drug, and have any post-baseline plasma concentrations sample collection of dasotraline.

6.1.10 Stratification Factor 'Baseline Number of Binge Days' vs. Baseline Number of Binge Days Group

Group of Baseline number of binge days is defined as: Baseline number of binge days ≤ 4, Baseline number of binge days >4; this group should be essentially equivalent to the stratification factor unless there is a subject stratified into the incorrect stratum. The stratification factor (ie, Baseline binge days category) will be consistently used in the statistical models, when applicable, for efficacy analysis as described in this document. Compared with the stratification factor, group of Baseline number of binge days will be used in the other scenarios, including the statistical models based analyses for safety variables.

6.1.11 Handling of Dropouts or Missing Data

6.1.11.1 Missing Diary Entry

Missing diary on certain assessed days in an assessment period will not be imputed, see Section 5.1.1.1.

6.1.11.2 Unknown or Unconfirmed Last Study Dose Date

If the end date of the medication intake is unknown or 'unconfirmed' (including but not limited to, eg, subject is lost to follow-up), it will be imputed as the earlier of the disposition date (ie, date of the last contact with subject) and 7 days from the date that the last medication kit is dispensed.

6.1.11.3 Composite Score

Unless otherwise specified, any individual missing item in any scale won't be imputed. When calculating a total score, subscale score, composite score, or any assessment with more than one item (e.g. Y-BOCS-BE total score and 2 subscale scores, EDE-Q global score and 4 subscale scores, EDE-QM global score and 3 subscale scores, SDS total score and 3 subscale scores, MADRS total score, HAM-A total score, CSSA total score, DESS total score), if one or more items are missing at a visit, then the associated score (ie, total score or subscale score) will be set to missing.

In the event of missing data at Week 12 for efficacy variables, the last post-baseline observation carried forward (LOCF) approach is used to impute missing values in the ITT LOCF efficacy analyses. Unscheduled visits are included in the imputation.

For analyses of the change from baseline for a given efficacy parameter, only subjects who have both baseline and at least one post-baseline measurement are included.

6.1.11.4 Incomplete/Missing Dates

6.1.11.4.1 Start/End Dates for Adverse Event

Treatment-emergent adverse events (TEAEs) are those events with an onset date on or after the start of double-blind study medication, through 7 days after study drug discontinuation (14 days for serious adverse events and deaths) for subjects who complete or discontinue this study but do not enter into the extension study), or through the last study day of the double-blind treatment period for subjects continuing into the extension study. A conservative approach will be used to handle the missing dates for AE onset date and end date to identify the treatment-emergent AE (ie, if the available incomplete date cannot determine treatment emergent status, the AE will be considered as a treatment-emergent AE). No imputation is done to calculate AE duration.

6.1.11.4.2 Start/End Dates for Prior/Concomitant Medication

Partial dates of non-study medication start/end date need to be imputed in order to classify the medications as prior or concomitant medications, or post-treatment medication (ie, medication initiated after the last dose of study medication). A conservative approach will be used to handle the incomplete date(s) for medication start/end date (ie, if the available incomplete date(s) cannot determine the concomitant relationship relative to the double-blind phase, the medication is considered as a concomitant medication).

6.1.11.4.3 Psychotherapy Dates /Psychiatric History Dates

Partial dates of initial symptoms of BED, initial diagnosis of BED, and psychotherapy will be imputed as follows: (1) If year and month are known, and if the year or month is previous to Screening, use the 15th of the month; otherwise, if the month is the month of Screening, use the 1st of the month. (2) If only year is known, and it is previous to the year of Screening, use June 30th of that year; If only year is known and it is same as the year of Screening, use Jan 1. This imputation rule will be applied to calculation of duration of initial symptoms of BED, initial diagnosis of BED, and psychotherapy.

6.1.11.4.4 Lab Data

Where applicable, the upper limit of quantification (ULOQ) will be imputed for chemistry, hematology, urinalysis, and urine microscopic values that are greater than the ULOQ. Similarly, if the value is below the lower limit of quantification (LLOQ), then the LLOQ will be imputed.

6.2. Methods of Analysis

6.2.1 General Methodology

All statistical inference analyses will be performed with 2-sided tests at a significance level of 0.05, and 2-sided 95% confidence intervals (CIs) will be calculated whenever appropriate. All

data will be summarized by treatment group and visit as appropriate. All subject data will be presented in data listings by subject.

Data for unscheduled assessments will be included when deriving baseline values and the worst case analysis during an assessed period. Unscheduled visits will also be included in treatment endpoint (Week 12 LOCF), which is the last available non-missing treatment endpoint carried forward to Week 12.

6.2.2 Statistical Hypotheses for Trial Objectives

6.2.2.1 Hypotheses for Primary Objective

In adults with moderate to severe BED, after 12 weeks of treatment,

- Dasotraline 6 mg/day reduces change from baseline in the number of binge days per week relative to placebo
- Dasotraline 4 mg/day reduces change from baseline in the number of binge days per week relative to placebo

6.2.2.2 Hypotheses for Key Secondary Objectives

In adults with moderate to severe BED, after 12 weeks of treatment, 6 mg/day dasotraline or both doses of dasotraline:

- Reduce change from baseline in BE-CGI-S score relative to placebo
- Provide a greater proportion of subjects with a 4-week cessation from binge eating at Week 12 relative to placebo
- Reduce change from baseline in Y-BOCS-BE total score relative to placebo

6.2.3 Interim Analysis

Not Applicable.

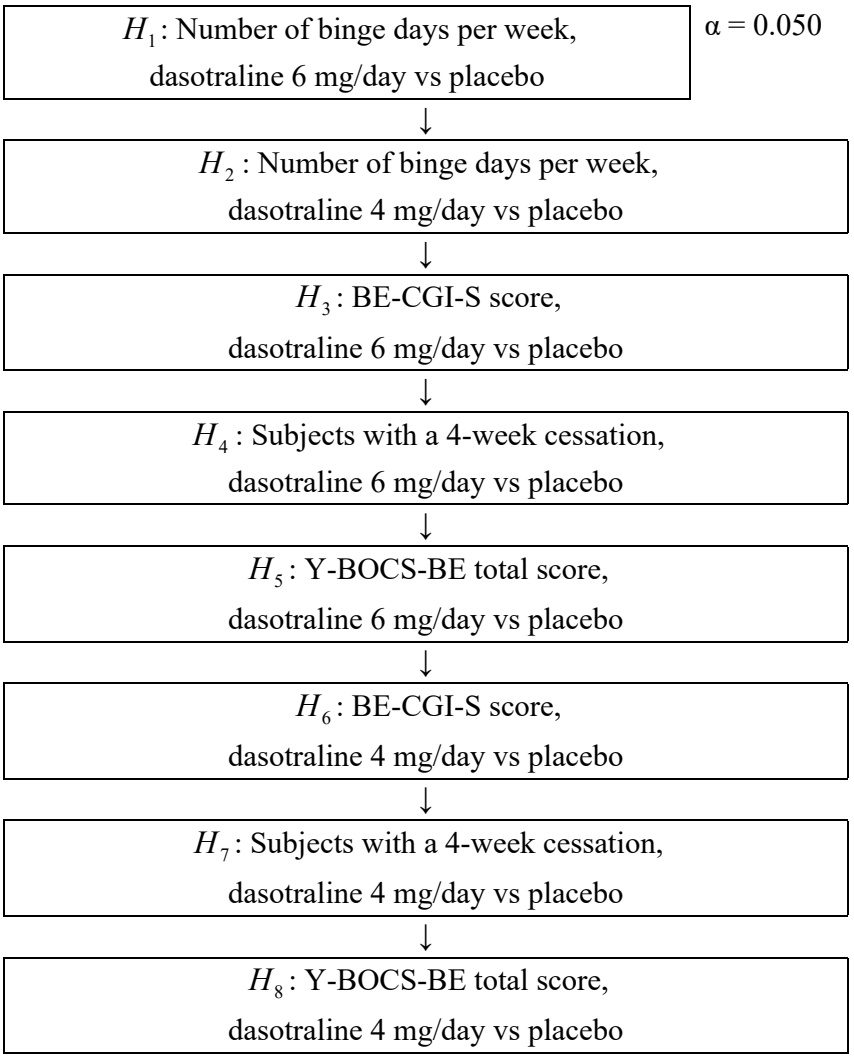
6.2.4 Data Monitoring

Data and Safety Monitoring Board (DSMB) meetings are planned to be held at regular intervals. Data to be presented to the DSMB are described in its charter.

6.2.5 Multiple Comparisons/Multiplicity

To control the overall type I error rate strongly at 5% for the primary and key secondary endpoints, for the hypotheses to be tested, a sequential testing strategy will be used. Following the fixed sequence closed testing procedure in Figure 1, testing will only proceed conditional on the statistical significance of the test(s) of prior level(s) at a 2-sided 5% significance level.

Figure 1: Fixed Sequence of Hypothesis Testing for Primary and Key Secondary Efficacy Endpoints



Denote raw *p*-values associated with hypotheses H_1 to H_8 as p_1 to p_8. Adjusted *p*-value of raw *p*-value of each testing hypothesis is calculated as below.

$$p_{1\_adj} = p_1,$$

$$p_{i\_adj} = \max(p_i, p_{i-1\_adj}), \quad i = 2, ..., 8$$

7. Statistical Analyses

All analyses and summaries will be produced using SAS® version 9.3 (or higher) for windows. Unless otherwise specified, continuous variables will be summarized with descriptive statistics including: the number of observations (n), mean, standard deviation (SD), median, range (minimum and maximum values). For change from baseline, 95% CI will be calculated whenever appropriate. All categorical data will be presented using frequency counts and percentages. Shift tables will be produced for some assessments that contain counts and percentages of subjects in each cross-classification level of baseline versus post-baseline assessment. The by-visit summary of continuous and categorical data will be displayed at protocol planned visits (weeks) and endpoint. Data for unscheduled assessments will be included when determining baseline values, selecting worst result for shift analyses, and for summary of normal/abnormal values. The listings, except for a listing for screen failures, will include all randomized subjects in the database. In general, the subject listings will be sorted by randomized treatment group, subject number and assessment visit and date (and time, if applicable).

7.1 Subject Information

Unless otherwise specified, subject information will be presented by treatment group for the ITT population. The treatment groups will include all Dasotraline (ie, pool two dasotraline groups together) and overall (ie, pooled three treatment groups together). If the safety population differs from the ITT population, similar descriptive statistics will be provided by treatment group, all Dasotraline, and overall for the safety population.

7.1.1 Demographic and Baseline Characteristics

Demographic and baseline characteristics will be summarized by treatment group, and overall for the ITT and PP populations, respectively.

Baseline descriptive statistics (N, mean, SD, median, and range) of demographic characteristics will be provided by treatment group for age at study entry (in years), body weight, height, body mass index (BMI, calculated as weight (kg)/ [height $(m)^2$]), and waist circumference.

The number and percentage of subjects in each of the following categories (and sub-categories) at baseline will also be summarized by treatment group:

- Age group: <40, ≥40 years old;
- Gender: Male, Female;
- Race: white, black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Other;
- Ethnicity: Hispanic or Latino, Not Hispanic or Latino;
- Baseline BMI category: Underweight/Normal (<25), Overweight (25 to <30), Obesity Class I (30 to <35), Obesity Class II (35 to <40), and Obesity Class III (≥ 40) kg/m^2;

Baseline descriptive statistics (N, mean, SD, median, and range) of disease characteristics will be provided by treatment group for number of binge eating days per weeks, number of binge eating episodes per weeks, BE-CGI-S, and Y-BOCE-BE total score.

The number and percentage of subjects in each of the following categories (and sub-categories) at baseline will also be summarized by treatment group:

- Baseline number of binge eating days category: 3-4, > 4 to 5, >5 to 6, >6 to 7;
- Baseline severity of BED (number of episodes/week): Mild (3), Moderate (4-7), Severe (8-13), Extreme (≥14)

7.1.2 Psychiatric History

Descriptive statistics (N, mean, SD, median, and range) of baseline disease characteristics will be provided by treatment group, and overall for both the ITT and PP populations for

- Age at the first onset of initial symptoms of BED (years);
- Duration of initial onset of initial symptoms of BED to screening (in years);
- Age at the first onset of initial diagnosis of BED (years);
- Duration of initial onset of initial diagnosis of BED to screening (in years);
- the average number of binge eating days per week in the preceding six months
- the least number of binge eating days per week in the preceding six months
- the maximum number of binge eating days per week in the preceding six months
- the average number of binge episodes per week in the preceding six months

Frequency distribution of the following variables will be summarized by treatment

- the least number of binge eating days per week in the preceding six months
- the maximum number of binge eating days per week in the preceding six months

The number and percentage of subjects in each of the following categories (and sub-categories) at baseline will also be summarized by treatment group

- the average number of binge eating days per week in the preceding six months: 2 to <3, 3 to < 4, 4 to <5, 5 to <6, and 6 to 7;
- the average number of binge episodes per week in the preceding six months: 2 to <4, >=4 to <=7, >7 to <14, and >=14.

Summaries of any other current psychiatric disorders present at screening include DSM-V codes for the ITT and PP populations. If a subject had 1 or more history events more than once, the subject will be counted only once under any given DSM-V code.

7.1.2.1 Psychiatric Family History

The number and percentage of subjects whose biological relatives (mother, father, siblings, children, extended family) have/had each of pre-specified psychiatric history will be summarized by treatment group for both the ITT and PP populations.

Listings of psychiatric history and psychiatric family history are provided for all randomized subjects.

7.1.3 Disposition Information

Subject disposition for all randomized subjects will be summarized by treatment group using the following categories:

- Subjects who are randomized;
- Subjects who are randomized, but not dosed;
- Subjects in the ITT population;
- Subjects in the Safety population;
- Subjects in the PP population;
- Subjects in the ITT who complete the Week 12 visit;
- Subjects in the ITT who complete the Week 12 visit and enter into the open-label extension study;
- Subjects who discontinue during the DB period.

- Subjects in the withdrawal population;
- Subjects who complete the withdrawal period;
- Subjects who discontinue during the withdrawal period

For subjects who discontinue from the study after randomization, the corresponding reasons for discontinuation will be summarized for DB period and withdrawal period, respectively. For subjects who are excluded from the PP population, the corresponding reasons will be summarized by treatment.

For subjects who consented but were not randomized (eg, screen failures) the corresponding reasons for discontinuation will be summarized.

Kaplan-Meier plot of the time to early discontinuation in the DB period will be presented by treatment group. Subjects who complete 12-weeks DB treatment period will have their time to treatment discontinuation censored on the study day of end of the treatment period.

7.1.4 Treatment Compliance

At each post-baseline (Weeks 1 − 12) visit, prior to dispensing study medication, previously dispensed study medication will be retrieved and assessed by tablet count.

Unless otherwise specified, compliance in the double-blind phase will be calculated for each post-baseline visit and overall as follows:

- Percent compliance = (number of tablets taken / number of tablets should have taken) x 100.
- Number of tablets taken = number of tablets dispensed – number of tablets returned.
- Number of tablets should have taken = (number of tablets supposed to take in a day) x (number of exposure days).
- Number of exposure days = last dose date – first dose date + 1.

For each post-baseline visit, if medication kit usage information is incomplete (ie, meet one of following conditions: any dispensed medication kit at the visit has missing number of tables returned, or has missing dosing start or ending date), the corresponding compliance will not be derived and will be set as missing.

For an assessment period that involves multiple study visits (ie, overall double-blind treatment period), if at some visit(s) the drug usage information is incomplete, then the other visits with complete information of kit usage will be used to derive the compliance rate. Number and percentage of subjects with incomplete medicate kit usage information will be summarized by study visit and treatment group.

Non-compliance is defined as less than 75% or more than 125% non-missing compliance with the study medication. Subjects with missing compliance will not be classified as non-compliant.

Descriptive summary statistics (N, mean, median, and range) will be provided by treatment group and overall for percent compliance. In addition, the number and percentage of subjects in the categories of the following variables will also be summarized by treatment group through the Week 12 visit:

- Compliant;
- Non-compliant (<75% or >125%);
- Non-compliant: <75%;
- Non-complaint: >125%;
- Complete and Incomplete medication kit usage information: complete, incomplete for 1-2 weeks, 3-4 weeks, and > 4 weeks visit period.

7.1.5 Extent of Exposure

Treatment duration is defined as the amount of elapsed time between the first and the last day that study medication was taken (inclusive). It will be calculated (in days) in terms of the difference in relative study days between the last and first dose of study medication, plus one

day. See Section 6.1.9.2 for imputation rule if the end date of the medication intake is unknown or 'confirmed'.

Descriptive statistics (N, mean, standard deviation, median, and range) will be presented by treatment group for treatment duration. Number of subjects with duration in each of following categories: ≥ 7, ≥ 14, ≥ 21, ≥ 28, ≥ 42, ≥56, ≥70, and ≥ 83 days;1-6, 7-13, 14-20, 21-27, 28-41, 42-55, 56-69, 70-82, and ≥ 83 days will also be presented by treatment group and overall. Return of the blister card at any given visit will not be a prerequisite for inclusion in exposure summary statistics if corresponding dose start and end dates are reported.

Note that exposure will be shown only up through the Week 12 visit or, for subjects with premature discontinuation, the last day of dosing of double-blind study drug for the subject.

7.1.6 Protocol Deviations

Deviations from the protocol, as adjudicated by Medical Monitor review, may have an impact on the efficacy analysis.

Protocol deviations that may impact on efficacy analysis include, but will not be limited to the following:

- Subject received any incorrect study treatment or was not dosed in DB period
- Overall missing diary rate in double-blind treatment period is > 5%
- Subject does not have a baseline efficacy measurement or at least one post-baseline efficacy measurement for primary efficacy variable
- Subject was unblinded during the double-blind treatment period
- Subject does not have 28 days or more of exposure
- Subject that does NOT belong to one of the following 2 cases:

  Case 1: there is no medication kit usage with incomplete information, and overall medication compliance is between 75% and 125%;

  Case 2: if there is 1 or 2 medication kits with incomplete usage information, overall medication compliance per the worst case imputation is between 75% and 125%,
- Subject violated inclusion/exclusion criteria which have any potential impact on efficacy results
- Subject took >1 dose of a prohibited medication or more than one day of a prohibited dose of an allowed medication or received prohibited psychotherapy during the double-blind treatment period, that were judged to have a potential impact on the primary efficacy analysis based upon blinded medical review.

- If the subject tests positive for substance of abuse at any post-baseline visit during the double-blind period (through 10 days after the Early Termination Visit for subjects who discontinue prematurely or for subjects who complete the study but do not continue into the extension study; and, through the Week 12 Visit date for subjects continuing into the extension study), without a verified prescription for use of a drug that could have caused a positive test result (confirmed by medical review)

A complete list of subjects who met with at least one of above protocol deviation criteria, will be identified prior to database lock.

Subjects in ITT population regarded as protocol deviators per above criteria will not be included in the PP population. A summary of all these protocol deviations will be presented by treatment group and overall for all randomized subjects. Subjects will be counted once within each deviation category, regardless of how many deviations they have in that category. A listing of protocol deviations will be provided for all randomized subjects. A listing will also be provided for subjects who take any incorrect study drug.

In addition, regardless of potential efficacy impact, protocol deviations are collected during monitoring visits. Protocol deviations will be placed into the following categories: concomitant medications, dosing, enrollment criteria, laboratory, non-compliance, visit schedule, visit/procedure requirement, and other. Each instance of a protocol deviation will be reviewed and determined to be major or minor prior to database lock of the study.

Number and percentage of subjects with at least one major protocol deviation of each deviation category will be summarized by treatment group. A listing that presents all protocol deviations will be also provided.

7.1.7 Study Medication and Prior and Concomitant Medications

7.1.7.1 Prior and Concomitant Medications

Prior (ie, medications taken before initiation) and concomitant medications (ie, medications taken on or after initiation of double-blind study medication that were either started prior or after initiation of double-blind study medication) will be coded using World Health Organization (WHO) Drug Dictionary and Anatomical Therapeutic Chemical (ATC) codes.

Number and percentage of subjects taking prior and concomitant medications will be provided by level 3 ATC classification and preferred name by treatment group and overall. Since medications are coded to ATC classification by indication, preferred names may appear under multiple ATC classifications. Medications started after last dose of study drug will be included in the prior and concomitant medication listing only.

7.1.8 Medical History

Events reported on the medical history page of the eCRF at the Screening visit will be coded according to MedDRA®. The number and percentage of subjects having each system organ class (SOC) and preferred term will be summarized by treatment group and overall.

7.1.9 Psychotherapy

A listing of CRF collected psychotherapy will be provided for all safety subjects.

7.2 Efficacy

The primary efficacy analysis will be performed based on the ITT population (Section 6.1.7) Assessment of the primary efficacy endpoint on the PP population will be conducted as supportive analyses.

7.2.1 Analysis Specifications

7.2.1.1 Level of Significance

Unless otherwise specified, all statistical tests will be interpreted at a 2-sided significance level of 5% and all confidence intervals (CIs) will be presented at a 2-sided confidence level of 95%.

The overall type I error rate for testing dasotraline 6 mg/day group versus placebo and dasotraline 4 mg/day group versus placebo will be strongly controlled at the 5% level across the primary endpoint and key secondary endpoints, which is detailed in Section 6.2.5.

7.2.2 Presentation of Analysis Result from Statistical Model

Whenever MMRM or ANCOVA analysis are conducted, LS mean of treatment differences, associated 95% CIs and p-values (raw) will be generated per corresponding statistical model. For each efficacy variable, effect sizes for within-group and between-group are presented at each post-baseline visit using a same approach in Section 7.2.3.2.1.

7.2.3 Primary Efficacy Endpoint

7.2.3.1 Definition

The primary efficacy endpoint is the change from baseline in the number of binge days per week at Week 12. See Section 5.1.1 for details of data definition and derivation at each assessed time point.

7.2.3.2 Analysis Methods

7.2.3.2.1 Primary Efficacy Analysis

The primary efficacy analyses of primary efficacy endpoint (the change from baseline in the number of binge days per week at Week 12) will be performed using a likelihood-based mixed model for repeated measures (MMRM). The response (dependent) variable is the change from baseline in the number of binge days per week assessed at Weeks 1, 2, 3, 4, 6, 8, 10, and 12. Specifically, the MMRM model includes fixed effects terms for treatment, visit (as a categorical variable), pooled center, baseline binge days category (stratification factor), the number of binge days per week at baseline, and treatment-by-visit interaction. Restricted maximum likelihood estimation method will be applied using unstructured covariance model. The Kenward Roger approximation will be used to estimate denominator degrees of freedom. In case of a failure of convergence with the unstructured covariance, the following structures may be assessed in a sequential order: heterogeneous Toeplitz, Toeplitz, and spatial exponential covariance pattern model. Of the above 3 covariance structures, the first covariance structure yielding convergence in the MMRM model will be used for the MMRM analysis. The spatial exponential model is selected for the analysis of data with unequally spaced time points. The LS mean treatment differences (each dasotraline group minus placebo group) of change from baseline at Week 12, their 2-sided 95% CIs, and the associated p-values will be calculated based on this model. At each post-baseline visit, effect size for within-group and between-group will be presented. Within-group effect size is calculated as mean change from baseline divided by the SD. Based on the MMRM, between-group effect size at a visit is calculated as the LS mean difference from placebo divided by the model estimate of the pooled SD at the visit, which is obtained from the square root of the diagonal element, associated at the visit, from the covariance matrix.

In addition, descriptive statistics (mean, SD, 95% CI) will be provided for the number of binge days per week and change from baseline by study visit for each treatment group. The likelihood based mixed effects model can accommodate incomplete data under the assumption of ignorable attrition.

The primary efficacy analysis will be repeated for the PP population to examine the impact of premature dropouts and/or protocol deviations.

7.2.3.2.2 Assessment of Model Assumptions in the Primary Efficacy Analysis

The normality and homoscedasticity assumptions underlying the primary MMRM model will be assessed graphically. Conditional studentized and scaled residuals will be plotted against the predicted values, respectively, and Q-Q (quantile-quantile) plots of these residuals versus the expected quantiles of the standard normal distribution will be presented to provide a graphical view of similarity and difference in the two distributions. Cumulative distribution function (CDF) of the primary endpoint will be also provided by treatment group. Considering that the validity of likelihood-based analyses also hinge on correct specification of the correlation

structure, an unstructured correlation matrix for modelling the within-patient errors is used (see Section 7.2.3.2.1) as a means of avoiding misspecification. However, this approach does not guarantee validity. If unstructured fails to converge, a more parsimonious model is likely necessary, but not necessarily correct; for such case, MMRM based on structures more general than unstructured one, ie, separate unstructured matrices by treatment, will be checked.

If there are evidences of deviations from the model assumption(s), the degree and nature of such deviation(s) will be explored to better understand the potential impact on interpretation of the primary efficacy analysis.

7.2.3.2.3 Supportive Analysis

The primary efficacy endpoint will be analyzed using an analysis of covariance (ANCOVA) model and the LOCF approach, as a supportive analysis. The model include terms for treatment, pooled center, and baseline binge days category (stratification factor) as factors, and the number of binge days per week at baseline as covariates. The LS mean of treatment differences (each dasotraline group minus placebo group), their 2-sided 95% CIs, and the associated p-values will be obtained from the model.

Effect sizes for within-group and between-group will be presented at each post-baseline visit. Within-group effect size is calculated as mean change from baseline divided by the SD. Based on the ANCOVA model, between-group effect size is calculated as the LS mean difference from placebo divided by the model estimate of the pooled SD, which is calculated as the SE of the LS mean difference divided by the square root of the sum of inverse treatment group sizes.

7.2.3.2.4 Sensitivity Analyses

To address early dropouts under the assumption of missing not at random (MNAR), a pattern-mixture model (PMM) using a placebo-based multiple imputation method (Attachment 9.3.1) and a pattern mixture model using multiple imputations with penalties **(**ie**,** tipping point analysis by deflating the individually estimated treatment effect size by known factors) (Attachment 9.3.2) will be performed as sensitivity analyses to explore the robustness of the MMRM results for the primary analysis based on the ITT population.

The PMM using a placebo-based multiple imputation method is a method based on methodology proposed by Ratitch and O'Kelly ([5]). The assumption that efficacy profiles of dropouts after discontinuation are similar to those of placebo subjects is considered very conservative because this methodology tends to minimize the difference between each dasotraline group and placebo group.

The PMM using multiple imputations with penalties by deflating the individually estimated treatment effect size by known factors provides a way to assess plausible deviations from MAR [8]. The tipping point, which is defined as the value of the factor where statistical significance of

treatment effect is lost, will be evaluated. If the tipping point is unrealistically high, treatment effect is robust. This approach will generate a serial of conservative estimates and provide the extent of robustness of primary efficacy result in a stepwise way.

In case of a deviation from the assumptions required for the primary analysis (Section 7.2.3.2.2), to confirm the robustness of the primary analysis result, two additional sensitivity analyses, permutation test and Generalized Linear Mixed Model (GLMM) analysis (Attachment 9.3.3) will be performed.

Permutation test is to fit a large number of datasets (ie, 10000) based on a same MMRM for the primary analysis with randomly assigning pseudo-treatment group designations. The empirical *p*-value will be obtained from the permutation test. The non-parametric based permutation test provides a conservative way to assess the primary efficacy endpoint.

GLMM analysis will fit count data over time (ie, number of binge days among number of assessed days at each visit period) based on a binomial distribution. This approach may better address the potential unequal variances assumption among subjects due to different number of assessed days (either because of assessment schedule, early dropout, or missing diary) among subjects and therefore is expected to better reflect the true distribution of the primary efficacy endpoint.

7.2.3.2.5 Exploratory Analyses Related to Planned Region and Actual Region Group

For primary efficacy variable, following exploratory analyses will be performed using the same predictor terms as described above (Sections 7.2.3.2.1, 7.2.3.2.3, and 7.2.3.2.4) by replacing the fixed term 'pooled center' with either the planned region or actual region group (pooled ones when applicable), respectively, to evaluate the impact of mis-stratification of region on treatment effect and further corroborate the primary efficacy result.

- the MMRM model,
- the ANCOVA model
- the PMM using a placebo-based multiple imputation method,
- the PMM using multiple imputations with penalties by deflating treatment effect size,

7.2.3.2.6 Assessment of Dropout and Dropout Profiles

To assess the impact of the drop-outs on the primary efficacy variable, a plot of cumulative distribution of the primary efficacy endpoint (change from baseline in number of binge days per week at Week 12) will be presented by treatment using the worst imputation for dropouts (i.e., missing value at a visit will be imputed by the most extreme increase in binge days at the visit per treatment group).

Subjects will be grouped by the visit group at which they have their last number of binge days per week. This will result in five categories for subjects discontinuing: Weeks 1/2 dropouts, Weeks 3/4 dropouts, Weeks 6/8 dropouts, Weeks 10/12 dropouts, and Completers. Specifically, for an example, subjects who discontinue after Week 2 visit but prior to or at Week 4 visit will be classified as Weeks 3/4 dropouts. The means change from baseline in number of binge days per week and BE-CGI-S score over time will be plotted by dropout category for each treatment (prior to and including the termination visit).

A similar plot will be generated using the grouped termination reasons defined in Table 7 as below. Summaries involving dropout category and termination reason will be based on the ITT population. The reason of withdrawal by subject will be reviewed prior to DBL to better understand the sub-category it belongs (due to efficacy, due to adverse reaction, due to other).

Table 7 Reasons for Discontinuation and Mapping for Termination Reason

eCRF Term	Grouped Term
Adverse event	Adverse event
Lack of efficacy	Lack of efficacy
Lost to follow-up	Lost to follow-up or withdrew consent
Withdrawal by subject	
Non-compliance with study drug	Non-compliance and other
Protocol Violation	
Pregnancy	
Other	

7.2.3.2.7 Missingness of Diary Entry

Summary statistics (ie, N, mean, median, range) of number of days with missing diary entries and missing diary rate (calculated as 100*the number of days with missing diary divided by the total number of days in the assessed visit period), as well as number and percentage of subjects of each missing entries category (ie, missing 1 days only, 2 days only, 3 days only, 4 days only, and >4 days) will be summarized by visit and overall DB treatment period for each treatment. A listing of missing diary entry by subject will be provided as well.

7.2.4 Key Secondary Endpoints Analyses

7.2.4.1 Definition

The key secondary efficacy endpoints involved in the hypothesis testing of each dasotraline dose (4 mg/day and 6 mg/day) versus placebo include change from baseline at Week 12 in BE-CGI-S score and Y-BOCS-BE total score, and percent of subjects with a 4-week cessation from binge eating prior to Week 12/EOT. See Section 5.1.1.4 for details.

7.2.4.2 Analysis Methods

Change in BE-CGI-S score and Y-BOCS-BE total score

Continuous key secondary variables (ie, change in BE-CGI-S score and Y-BOCS-BE total score) will be analyzed with an MMRM model similar to the primary efficacy analysis (Section 7.2.3.2.1) by adjusting the corresponding baseline. The LS mean treatment differences (each dasotraline group minus placebo group) of change from baseline at Week 12, their 2-sided 95% CIs, and the associated p-values will be calculated based on the model. Effect sizes for within-group and between-group will be presented at each post-baseline visit using a same approach in Section 7.2.3.2.1.

As a supportive analysis of the above MMRM analyses, the change from baseline in BE-CGI-S and Y-BOCS-BE total score at Week 12 will be analyzed using an ANCOVA model and the LOCF approach for the ITT population, which includes terms for treatment, pooled center, baseline binge days category (stratification factor), and corresponding baseline score as covariate; Effect sizes for within-group and between-group will be presented at each post-baseline visit by a similar approach described in Section 7.2.3.2.3.

To address early dropouts or potential deviation from the model assumption(s), sensitivity analyses similar to Section 7.2.3.2.4, including PMM using a placebo-based multiple imputation method, PMM using multiple imputations with penalties by deflating the individually estimated treatment effect size, and permutation test may be conducted for the key secondary endpoint for BE-CGI-S; PMM using a placebo-based multiple imputation method may be conducted for Y-BOCS-BE total score.

For above key secondary efficacy variables, following exploratory analyses will be performed using the same predictor terms described above by replacing the fixed term 'pooled center' with the planned region or actual region group (pooled one when applicable), respectively, to evaluate the impact of mis-stratification of region on treatment effect and further corroborate these key secondary efficacy result.

- the MMRM model
- the PMM using a placebo-based multiple imputation method
- the PMM using multiple imputations with penalties by deflating treatment effect size (for BE-CGI-S only)

Percent of subjects with a 4-week cessation from binge eating

The categorical key secondary efficacy endpoint of percent of subjects with a 4-week cessation from binge eating will be analyzed using a logistic regression model with treatment, baseline binge days category (stratification factor), and baseline number of binge days per week as covariate using a LOCF approach based on ITT population. The odds ratios, their 2-sided 95%

CIs, and the associated p-values for each dasotraline group over the placebo group will be derived from the model.

As a supportive analysis, percent of subjects with a 4-week cessation from binge eating will be analyzed by an inverse probability weighting (IPW) model; See Attachment 9.3.4 for details about IPW method ([6]). A sensitivity analysis per quasi-likelihood approach will be conducted by fitting same fixed terms as listed above with random-residual based on binary distribution with logit link, which will effectively lift the constraints of the model parameter. The above approach will be implemented using GLIMMIX procedure in SAS.

For above logistic regression model, two exploratory analyses will be performed using same predictor terms with an additional fixed term of planned or actual region group (pooled one when applicable), respectively, to evaluate the impact of mis-stratification of region on treatment effect and further corroborate key secondary efficacy results.

7.2.5 Multiplicity Adjustment

To control the overall type I error rate strongly at 5% for the primary and key secondary endpoints, for the hypotheses to be tested, a sequential testing strategy will be used. Following the fixed sequence closed testing procedure in Figure 1, testing will only proceed conditional on the statistical significance of the test(s) of prior level(s) at a 2-sided 5% significance level. See Section 6.2.5 for details.

7.2.6 Other Secondary Endpoints Analyses

7.2.6.1 Definition

The other secondary endpoints of interest are:

- Change from baseline in number of binge days per week at Weeks 1, 2, 3, 4, 6, 8, and 10
- Change from baseline in number of binge episodes per week at Weeks 1, 2, 3, 4, 6, 8, 10, and 12
- Change from baseline in BE-CGI-S score at Weeks 2, 4, 6, 8, and 10
- Change from baseline in Y-BOCS-BE total score at Weeks 2, 4, 6, 8, and 10 and subscale scores (obsessions and compulsions) at Weeks 2, 4, 6, 8, 10, and 12
- Change from baseline in MADRS total score at Week 12
- Change from baseline in HAM-A total score at Week 12
- Change from baseline in SDS total score and subscale scores (school/work disability, social life disability, and family life disability) at Weeks 6 and 12
- Proportion of binge eating responders (ie, ≥ 75% reduction in the number of binge eating episodes) at Week 12

- Change from baseline in EDE-Q modified including EDE-Q modified global score and 3 subscale scores (dietary restraint, shape/weight overvaluation, and body dissatisfaction) at Week 12

7.2.6.2 Analysis Methods

The analyses of continuous variables associated with the primary efficacy endpoint and the key secondary endpoints (ie, the change from baseline in number of binge days per week, BE-CGI-S score and Y-BOCS-BE total score) have been described on Sections 7.2.3 and 7.2.4 of this SAP.

Continuous variables related to change in number of binge episodes per week, Y-BOCS-BE subscale scores , and SDS-total score and subscale scores will be analyzed using an MMRM model similar to the primary efficacy analysis (Section 7.2.3.2.1) with adjusting the corresponding baseline; other continuous variables with one scheduled post-baseline assessment (ie, change in MADRS total score, HAM-A total score, EDE-QM global score and subscale scores) will be analyzed using an ANCOVA model similar to the supportive analysis for the primary efficacy endpoint (Section 7.2.3.2.3) with adjusting the corresponding baseline as a covariate. The treatment differences and 95% CIs for each dasotraline group relative to placebo in the change from baseline will be obtained from this model.

The proportion of binge eating responders at Week 12 will be analyzed using a logistic regression model and LOCF approach with factors of treatment, pooled center, baseline binge days category (stratification factor), and baseline number of binge episodes per week as covariate. For subjects who are discontinued, binge eating responders at Week 12 LOCF is defined per the last post-baseline time period during the treatment period that binge diary is assessed. To ensure that the logistic regression model is feasible, pooled region may be removed from the analysis model if some pooled center has no responder either because there is a small number of subjects in the pooled center or subjects experience little change (for example, no subject achieving at least 75% reduction in number of binge episodes per week). In addition, for each dasotraline group (vs. Placebo), effect size and the number needed to treat (NNT) will be presented as well. The effect size will be calculated as log(odds ratio)*sqrt(3)/π, where odds ratio is based on 2 by 2 contingency table and the NNT is derived as 1/Risk Reduction, where Risk Reduction (RR) = (each dasotraline response rate – placebo response rate). The 95% CI of NNT will be obtained similarly by using the 95% CI of the RR when both lower and upper confidence limits are positive. The NNT and 95% CI results will be provided in whole number format as any fractional values are rounded up to the nearest whole number (the lower confidence limit is rounded down and the upper confidence limit is rounded up.)

As a supportive analysis, the proportion of binge eating responders for achieving 75% reduction over time will be analyzed, respectively, by an inverse probability weighting (IPW) Generalized Estimating Equations (GEE) model. See Attachment 9.3.4 ([6]) for details.

In addition, the number and percentage of subjects in the following categories for reduction in the number of binge eating episodes from baseline will be summarized by treatment and by visit:

- cessation (defined as 100% reduction)
- 99% to 75% reduction
- 74% to 50% reduction
- other

At each assessment time point, for each of above continuous secondary efficacy variables, summary statistics (N, mean, standard deviation, median, and range) will be presented for the actual values and the changes from baseline. The 95% CIs of the changes from baseline will also be presented.

Plots of the mean actual values as well as MMRM LS means (±SE) of change in number of binge days over time by treatment group will be presented for the ITT and PP populations; plot for ANCOVA LS means (±SE) of change will be also presented over time including Week 12 LOCF endpoint. Similar plots of the key secondary endpoints such as BE-CGI-S score, Y-BOCS-BE total score, and number of binge episodes over time will also be provided based on the ITT population.

7.2.7 Subgroup Analysis

The primary efficacy variable (change from baseline in number of binge days per week at Week 12) and key secondary variable (change from baseline in BE-CGI-S at Week 12) will be examined to explore the consistency of the treatment effect across certain subgroups listed below (if there are at least 45 subjects in the subgroup).

The subgroup analysis will be conducted based on a similar MMRM model to the corresponding primary efficacy analysis with three additional terms: subgroup, treatment-by-subgroup interaction, and 3-way interaction of treatment*subgroup*visit.

Descriptive statistics (N, mean, standard deviation), LS Means and SE, the treatment differences vs. placebo and 95% CIs at Week 12 from the MMRM model will be provided for each category of the subgroups by treatment group:

- Gender: male or female
- Race group: White, Non-white
- Age group: <40, ≥40 years old
- Ethnicity: Hispanic, non-Hispanic
- Baseline BMI (kg/m^2): Non-obese (<30), Obese (≥30 kg/m^2)
- Baseline severity (number of episodes/week): Moderate (<=7), Severe (>7)

P-value for treatment-by-subgroup interaction at Week 12 from the MMRM model will be provided. For each category of above subgroups and for overall subjects, LS mean differences (vs. placebo) and their 95% CIs in the change from baseline at week 12 in the number of binge

days/week will be plotted on a same forest plot. A similar forest plot will be provided for BE-CGI-S as well.

All subgroup analyses will be based on the ITT population.

7.3 Pharmacokinetics (PK) Analysis

Plasma concentrations for dasotraline will be summarized at Week 2, 6, and 12 by the most recent dasotraline dose level for the PK population. The descriptive statistics (n, mean, SD, coefficient of variation [CV%], median, range) will be provided by most recent dasotraline dose group. A listing of PK concentration for all samples will be provided.

7.4 Safety

7.4.1 Analysis Specifications

Safety reporting includes all safety data reported during the 12-weeks double-blind treatment period, as well as the all safety data reported in post-treatment withdrawal period, if a subject does not enter the extension study.

The safety data will be analyzed by treatment group based on the safety population for data up to the end of treatment period and on the withdrawal population for data collected during the withdrawal period, separately, as appropriate. Specifically, the data, which are used to assess physical dependence and withdrawal symptoms, will be analyzed by treatment for data collected in the withdrawal period based on the modified withdrawal population for AE, CSSA and DESS, and based on withdrawal population for MADRS and HAM-A, respectively. Unless otherwise specified, for continuous data collected in the treatment period, the actual value and change from baseline (ie, DB baseline) will be summarized by treatment and visit; for continuous data collected in the withdrawal period, the actual value, change from DB baseline (when applicable), and change from withdrawal baseline will be summarized by treatment and visit.

There will be no imputation of missing values for clinical laboratory test results, vital sign measurements, and ECG evaluations in the by-visit analyses. All analyses of safety data are descriptive. No inferential analyses of safety data are performed.

7.4.2 Adverse Events

All AEs are coded using Medical Dictionary for Regulatory Activities (MedDRA) version 18.0 or higher.

Treatment-Emergent Adverse Event (TEAE)

A treatment-emergent adverse event (TEAE) is defined as an AE with a start date on or after the date of first dose through 7 days after study drug discontinuation (14 days for serious adverse events and deaths) for subjects who complete or discontinue this study but do not enter into the extension study), or through the last study day of the double-blind treatment period for subjects

continuing into the extension study (ie, subjects in safety population). AEs with a start date prior to initiation of double-blind study medication which are reported to have an increase in intensity, or which are reported to have a attribution in relationship to study medication (ie, attribution to possible related, probably related, related) after the initiation of double-blind study medication are considered as TEAEs.

Discontinuation-emergent adverse event (DEAE) is defined as an AE with a start date after the last dose day of study medication through the last study visit in the withdrawal period.

Table summary of overall AEs (and SAEs) will be limited to TEAEs and DEAEs. All AE data will be displayed in data listings. For sake of simplicity, hereafter TEAE is referred as AE in this section.

The overall incidence (ie, number and percent of subjects with one or more AE in each category) of AEs, serious AEs, deaths, AEs leading to discontinuation, study drug-related AEs, serious study drug-related AEs, study drug-related AEs leading to discontinuation, and serious AEs leading to discontinuation, serious study drug-related AEs, will be summarized by treatment group.

The AEs also will be summarized by system organ class (SOC) and preferred term (PT) by presenting the number and percentage of subjects with each AE category. The incidence of AEs (by preferred term, grouped by SOC, and presented by treatment group) will also be summarized by severity, by the relationship to study medication, by the action taken regarding the study medication, as well as by the outcome. If a subject has the same AE on multiple occasions, the highest severity (severe, moderate, mild) recorded for the event and the highest drug relationship (unrelated, unlikely to be related, possibly related, probably related, and definitely related) will be presented, respectively. The relationship to study medication will be presented by 2 categories: related (which includes possibly related, probably related, and related, as determined by investigators), and not related (which includes unrelated and unlikely to be related, as determined by investigators). Furthermore, for AE summary by SOC and PT or PT only, the following combined preferred terms will be summarized as well:

- Insomnia: Insomnia, Initial insomnia, Middle insomnia, Terminal insomnia.
- Hallucination: Hallucination; Hallucination, auditory; Hallucination, gustatory; Hallucination, olfactory; Hallucination, synaesthetic; Hallucination, tactile; Hallucination, visual; Hallucinations, mixed; Hypnagogic hallucination; Hypnopompic hallucination; Somatic hallucination.

A Kaplan-Meier plot of the time to the first onset of treatment emergent insomnia (combined term) will be presented by treatment group. Time to the first onset (in days) is defined as the earliest onset date of insomnia – start date of DB treatment period + 1. Subjects who do not experience any insomnia on or before the last dose date will be censored on the day of last dose.

Prevalence rate of insomnia (combined term) at a study day in the DB treatment period is defined as number of subjects experiencing insomnia on the day divided by number of subjects remaining on DB treatment period at that time point. It is notable that subjects may experience multiple insomnia episodes during the study. Prevalence rate over time will be presented graphically by treatment group.

AE maximum relationship

Each AE will be classified as 'related', 'not related', and 'not specified'. TEAEs with a missing AE relationship will be classified as 'not specified'. AE summary by the relationship to the study medication will be based on the highest relationship within an SOC and PT. Response values will be ranked in order from minimum to maximum as 'not specified' < 'not related' < 'related'.

AE maximum severity

AEs with a missing severity will be classified as 'not specified'. AE summary by the severity will be based on the maximum severity within an SOC and PT. Response severity values will be ranked in order from minimum to maximum as 'not specified' < 'mild' < 'moderate' < 'severe'.

AE maximum action taken

AEs with a missing action taken will be classified as 'unknown'. AE summary by the action taken will be based on the maximum action taken within an SOC and PT. Response action values will be ranked in order from minimum to maximum as 'not applicable'< 'unknown' < 'dose increased' < 'unchanged'< 'dose reduced' < 'drug interrupted'< 'drug withdrawn'.

AE maximum outcome

AEs with a missing outcome will be classified as 'unknown''. AE summary by the outcome will be based on the maximum outcome within an SOC and PT. Response outcome_values will be ranked in order from minimum to maximum as 'unknown' < 'recovered/resolved' < 'recovered/resolved with sequelae' < 'recovering/resolving' < 'not recovered/not resolved'< 'fatal'.

AEs with Special Interest

A list of preferred terms that are to be combined for the assessment of each of the pre-specified adverse events (including cardiovascular-related AEs, metabolic related AEs, neuropsychiatric-related AEs, and suicide and self-injury related AEs) is provided in Attachment 9.4. These AE preferred terms are determined and verified by medical review of coded AE terms during BDRM and finalized prior to database lock. The incidence of cardiovascular-related AEs, metabolic related AEs (weight-related and non-weight-related), neuropsychiatric-related AEs (including a defined subset of psychosis and mania-related AEs) and suicide and self-injury related AEs will be summarized by treatment, respectively.

Listings will also be generated for deaths, SAEs, and discontinuations due to AEs. All adverse events regardless of treatment emergent status will be provided in the AE data listing.

Discontinuation-Emergent Adverse Event (DEAE) (ie, AEs Onset After Last Study Dose)

All summary of DEAEs will be based on a 'modified' withdrawal population (see Section 6.1.7).

The overall incidence of DEAEs, DEAEs by SOC and PT (by presenting the number and percentage of subjects with each AE category), and DEAEs by PT will be summarized by treatment group.

In addition, incidence of DEAEs by SOC and PT will be also summarized using time intervals beginning with the earliest AE concurrent withdrawal duration in the withdrawal period, where denominators in percentage calculation for withdrawal duration intervals (in days) will be based on the number of subjects whose withdrawal duration are no less than the first day of the interval. Withdrawal duration (in days) will be derived as the last withdrawal visit date minus the last dose date +1 and withdrawal duration categories to be assessed are 1-7 days, 8-14 days, and ≥ 15 days.

7.4.3 Laboratory Measurements

Laboratory data will be summarized for each laboratory test listed in Attachment 9.5.1.

7.4.3.1 Treatment Period

Descriptive statistics (N, mean, standard deviation, median, and range) will be reported for each continuous laboratory analyte at baseline, each scheduled post-baseline time point, and endpoint for observed value as well as change from baseline. Categorical results (eg, urinalysis tests) will be summarized at each visit by treatment group using frequency and percentage. If a subject has more than one measurement at any time point, the one closest to the scheduled visit will be used in the by-visit analysis. All lab values will be presented in the data listings.

Results for glucose and lipid tests (HDL cholesterol, LDL cholesterol, total cholesterol, and triglycerides) will be presented separately by fasting status, which includes fasting only and overall (fasting, non-fasting, or unknown). The change from baseline value at endpoint for these selected laboratory parameters (ie, lipid tests, glucose, HbA1c, and insulin), will be evaluated using a nonparametric rank ANCOVA. For the comparison of each dasotraline group versus placebo group, the change from baseline at endpoint and baseline value is ranked. A linear regression will be conducted on the change from baseline ranks and baseline value rank as independent variable to produce regression residuals. Using the values of the residuals as scores, Mantel-Haenszel row mean score tests will be produced for each dasotraline group versus placebo group.

In addition, the change from baseline value at Weeks 6, 12, and endpoint for these selected laboratory parameters (ie, lipid tests, glucose, HbA1c, and insulin), will be further evaluated

using an ANCOVA model which includes terms for treatment, baseline binge days group, and the corresponding baseline as covariate. The LS means of treatment differences (each dasotraline group minus placebo group) and 2-sided 95% CIs will be obtained from the model.

The normal reference ranges for laboratory tests will be used to determine whether the laboratory test value is below, within, or above the normal range. Shifts from baseline to endpoint will be produced by treatment group to show the percentage of subjects with laboratory test values below, within, and above the normal range.

Criteria for Markedly Abnormal Post-baseline Laboratory Values (MAPLV) per International System of Units (SI) of selected laboratory parameters can be found in Table 8. Equivalent MAPLV criteria for conventional unit are provided in Attachment 9.5.2. The criterion of MAPLV is satisfied if a value falls into the markedly abnormal range at an assessed visit or if subjects have experienced at least once MAPLV in the study period. The number and percentage of subjects with a particular MAPLV at each post-baseline visit, endpoint, and during the post-baseline double-blind phase (up to and including endpoint), regardless of baseline value, will be presented by treatment group.

Table 8 Criteria for Markedly Abnormal Post-baseline Laboratory Values (MAPLV)

(SI Unit)

	Criteria for Markedly Abnormal Status	
Laboratory Test	Low	High
Hematology		
Hematocrit	Male: ≤ 0.37 Female: ≤ 0.32	Male: ≥ 0.60 Female: ≥ 0.54
Hemoglobin	Male: ≤ 115 g/L Female: ≤ 95 g/L	Male: ≥ 190 g/L Female: ≥ 175 g/L
RBC	≤ 3.5 $x10^{12}$/L	≥ 6.4 x 10^{12}/L
Total WBC Count	≤ 2.8 $x10^{9}$/L	≥ 16 $x10^{9}$/L
Basophils	N/A	≥ 10 % ≥ 1.6 $x10^{9}$/L
Eosinophils	N/A	≥ 10 % ≥ 1.6 $x10^{9}$/L
Lymphocytes	N/A	≥ 75 % ≥ 12.0 $x10^{9}$/L
Monocytes	N/A	≥ 15 % ≥ 2.4 $x10^{9}$/L
Neutrophils	≤ 15 % ≤ 0.42 $x10^{9}$/L	≥ 85 % ≥ 13.6 $x10^{9}$/L
Platelet Count	≤ 75 $x10^{9}$/L	≥ 700 $x10^{9}$/L
Chemistry		
Albumin	≤ 25 g/L	N/A
ALP	N/A	≥ 3 x ULN
AST	N/A	≥ 3 x ULN

Laboratory Test	Criteria for Markedly Abnormal Status	
	Low	High
ALT	N/A	≥ 3 x ULN
Bicarbonate	≤ 16 mmol/L	≥ 35 mmol/L
Calcium	≤ 2.05 mmol/L	≥ 2.99 mmol/L
Chloride	≤ 90 mmol/L	≥ 118 mmol/L
Creatinine	N/A	≥ 176.8 µmol/L
Glucose (fasting)	≤ 2.22 mmol/L	≥ 9.71 mmol/L
Inorganic Phosphorus	≤ 0.5 mmol/L	≥ 1.7 mmol/L
Magnesium	≤ 0.6 mmol/L	≥ 1.15 mmol/L
Potassium	≤ 3 mmol/L	≥ 6 mmol/L
Sodium	≤ 126 mmol/L	≥ 156 mmol/L
Total Bilirubin	N/A	≥ 34.2 µmol/L
Total Protein	≤ 45 g/L	≥ 100 g/L
Uric Acid	Male: N/A Female: N/A	Male: ≥ 625 µmol/L Female: ≥ 506 µmol/L
Urea (BUN)	N/A	≥ 11 mmol/L
Lipids panel		
Total Cholesterol (fasting)	N/A	≥7.77 mmol/L
LDL Cholesterol (fasting)	N/A	≥5.18 mmol/L
HDL (fasting)	≤ 1.036 mmol/L	N/A
Triglycerides (fasting)	N/A	≥ 3.39 mmol/L
HbA1c	N/A	≥ 7.5%
Urinalysis		
Urine Blood	N/A	≥ 1+
Urine Glucose	N/A	≥ 3+
Urine Ketones	N/A	≥ 2+
Urine Protein	N/A	≥ 2+

Note: ULN = Upper Limit of Normal; N/A = Not Applicable

A listing of subject data for those who have experienced at least once post-baseline MAPLV during the double-blind treatment period will be generated. Urine drug screen results will be presented for each parameter by visit and at endpoint. Number and percentage of positive and negative results will be displayed. Summary statistics of laboratory outputs will be generated per conventional units and International System of Units (SI), separately.

7.4.3.2 Withdrawal Period

Laboratory data will be summarized by presenting descriptive statistics of actual values, change from baseline, and changes from withdrawal baseline by visit and treatment.

7.4.4 ECG Evaluations

As a part of each ECG, heart rate (HR), axis, and the following intervals will be measured: RR, PR, QRS, and QT. Corrected QT intervals (QT_cB and QT_cF) values are based on the following formula:

- Bazett: QT_cB (msec) = QT (msec) / $(RR/1000)^{1/2}$
- Fridericia: QT_cF (msec) = QT (msec) / $(RR/1000)^{1/3}$

where RR will be reported in milliseconds (msec).

7.4.4.1 Treatment Period

Descriptive statistics (eg, mean, SD, median, range) for observed value and change from baseline values by visit and endpoint will be displayed for continuous variables (ie, HR, QRS, QTcF, etc). In addition, 95% CIs of the mean change from baseline value will be provided for post-baseline visits.

The overall ECG assessment is conducted by central reading and a determination is made whether the reading is normal (or absent if applicable) or abnormal). Shifts from baseline to post-baseline visit and endpoint will be produced by treatment group.

Bazett and Fridericia corrected QT_c values will be classified as having QT_c prolongation according to Table 9.

Table 9: – QTc Prolongation

QTc Prolongation
> 450 msec for males > 470 msec for females
> 480 msec
> 500 msec
Increase from baseline QTc ≥ 30 msec
Increase from baseline QTc ≥ 60 msec

The number and percentage of subjects having a QT_c prolongation using any of the two correction methods (Bazett, Fridericia) by baseline, post-baseline visits, endpoint, and any post-baseline measurement (ie, the worst case in DB period) will be summarized by treatment group. A listing of ECG data for subjects with at least one prolonged QT_c will also be produced.

For other ECG parameters, subjects are classified as normal or abnormal using the limits specified in Table 10. For abnormal ECG parameter reporting, counts and percentages of subjects will be presented for baseline, each post-baseline visit, endpoint, and any post-baseline DB visit (the worst case) by treatment group.

Table 10: – Definition of Abnormal ECG Values by Parameter

ECG parameter	Abnormally Low	Abnormally High

(unit)		
HR (bpm)	≤50	≥100
PR interval (msec)	N/A	≥210
QRS interval (msec)	N/A	≥120
QT interval (msec)	N/A	> 500

Note: N/A = Not Applicable

7.4.4.2 Withdrawal Period

ECG data will be summarized by presenting descriptive statistics of observed values, change from baseline, and changes from withdrawal baseline by visit and treatment.

7.4.5 Vital Signs, Weight, and BMI

7.4.5.1 Treatment Period

At each visit and endpoint, descriptive statistics (mean, SD etc.) will be provided by treatment group for

- Observed and change from baseline values of each vital sign parameter, including orthostatic changes in systolic blood pressure, diastolic blood pressure, and pulse rate;
- Observed and change, and percent change from baseline values in weight and BMI.

A similar MMRM model for the primary efficacy variable by using corresponding baseline will be applied for the following variables at the specified time points.

- Change and percent change from baseline in body weight (kg) at Weeks 1, 2, 3, 4, 6, 8, 10, and 12.
- Change and percent change from baseline in BMI at Weeks 1, 2, 3, 4, 6, 8, 10, and 12.

The number and percentage of subjects with Markedly Abnormal Post-baseline Vital Signs (MAPVS), defined in Table 11.1, will be presented by visit, endpoint, or study period and by treatment group. Included are subjects having experienced at least once post-baseline MAPVS during the double-blind phase, up to and including endpoint.

Table 11.1: – Criteria for Markedly Abnormal Post-Baseline Vital Signs and Weight

Parameter (unit)	Markedly Low	Markedly High
SBP (supine, standing) (mmHg)	<90	≥180
	Value ≤ 90 and ≥ 20 decrease from baseline	Value ≥ 180 and ≥ 20 increase from baseline
DBP (supine, standing) (mmHg)	<50	≥105
	Value ≤ 50 and ≥ 15 decrease from baseline	Value ≥ 105 and ≥ 15 increase from baseline
Pulse rate (supine, standing) (bpm)	<50	≥120
	Value ≤ 50 and ≥ 15 decrease from	≥120 and increased >15 from

	baseline	baseline
Weight (kg)	≥ 7% decrease from baseline	≥ 7% increase from baseline
Temperature (˚C)	N/A	Value ≥ 38.3°C and ≥0.8°C increase from baseline
Orthostatic Criteria		
SBP orthostatic criteria (mmHg)	≥ 20 mmHg reduction from supine to standing position	N/A
DBP orthostatic criteria (mmHg)	≥ 10 mmHg reduction from supine to standing position	N/A
Orthostatic tachycardia (bpm)	N/A	≥ 20 bpm increase from supine to standing position in pulse rate .

BMI measurements will be used to classify a subject's weight status as underweight, normal, overweight or obese using Table 11.2 (Category 1) and as underweight/normal, overweight, obesity class I, obesity class II, and obesity class III using Table 11.3 (Category 2). Shifts from Baseline to Week 12 Endpoint will be produced by treatment group for the number and percentage of subjects that fall into each BMI category. In addition, the number and percentage of subjects with 3%, 5%, 7%, and 10% reduction in body weight will be presented by visit and treatment group for overall safety population as well as by baseline BMI groups (Category 2).

Table 11.2: – BMI Weight Status Category 1

BMI (kg/m^2)	Weight Status
< 18.5	Underweight
18.5 ≤ BMI < 25.0	Normal
25.0 ≤ BMI < 30	Overweight
≥ 30.0	Obese

Table 11.3: – BMI Weight Status Category 2

BMI (kg/m^2)	Weight Status
BMI < 25	Underweight/ Normal
25 ≤ BMI < 30	Overweight
30 ≤ BMI < 35	Obesity Class I
35 ≤ BMI < 40	Obesity Class II
40 ≤ BMI	Obesity Class III

A listing of subject data for those who satisfy MAPVS at least once post-baseline visit in the double-blind treatment period will be presented.

7.4.5.2 Withdrawal Period

Vital sign data will be summarized by presenting descriptive statistics of observed values, change from baseline, and changes from withdrawal baseline by visit and treatment.

7.4.6 Assessments of Withdrawal Effects (CSSA, DESS, HAM-A, and MADRS)

CSSA total score, DESS total score, MADRS total score, and HAM-A total score will be summarized by presenting descriptive statistics (N, mean, median, and range) of observed values and change from withdrawal baseline by treatment and visit (ie, both clinic and non-clinic visits) for the withdrawal period.

A similar analysis will be performed by gender as well.

7.4.7 Columbia Suicide Severity Rating Scale (C-SSRS)

According to FDA draft guidance for suicidal ideation and behavior ([7]), the following C-SSRS outcomes, which have binary responses (yes/no), will be re-ordered and numbered to facilitate the further definitions of the composite endpoint and comparative endpoint of C-SSRS as below.

Category 1 – Wish to be Dead
Category 2 – Non-specific Active Suicidal Thoughts
Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan
Category 5 – Active Suicidal Ideation with Specific Plan and Intent

Category 6 – Preparatory Acts or Behavior
Category 7 – Aborted Attempt
Category 8 – Interrupted Attempt
Category 9 – Actual Attempt
Category 10 – Completed Suicide
Category 11– Self-injurious behavior without suicidal intent

Unless otherwise specified, C-SSRS table outputs will be presented by treatment for following analysis time points: lifetime (when applicable), baseline, any post-baseline in DB period, and any post-withdrawal Baseline. Definitions of these analysis time points are given in Table 12. The analysis time point is applied to suicidal ideation score and ideation intensity total score.

Table 12: Definition of Analysis Timepoint

Analysis Timepoints	eCRF Visit(s) Used	C-SSRS Version	Derivation Rule
Baseline	Screening	1 month prior to screening	Most severe value
	Baseline	Since last visit	

Any post-baseline in DB period	Post-baseline visits up to and including Week 12/EOT	Since last visit	Most severe value
Any post- withdrawal Baseline visit	Post-withdrawal baseline visits up to and including Week15/EOS	Since last visit	Most severe value

At each analysis time points, the following binary variables will be derived.

- Any Suicidal Ideation (Categories 1-5)
- Any Suicidal Behavior (Categories 6-10)
- Any Suicidality (Categories 1-10)
- Individual Item for Category 6-11, respectively.

Comparative endpoints for 'Emergence' (outcomes that include events that first emerge) and 'Worsen' (outcomes that include events that worsen) will be derived as below. See Table 13.

Table 13: Comparative Endpoint of C-SSRS in DB Treatment Period

Comparative Endpoints (Relative to DB Baseline)	Derivation Rule
Emergence of Suicidal Ideation	No suicidal ideation at Baseline, and any type of suicidal ideation in any post-baseline in DB period
Emergence of Serious Suicidal Ideation	No suicidal ideation at baseline, and any serious suicidal ideation [ideation score of 4 or 5] in any post-baseline in DB period
Emergence of Suicidal Behavior	No suicidal behavior at baseline, and any type of suicidal behavior in any post-baseline in DB period
Emergence of Suicidality	No suicidality at baseline, and any suicidality in any post-baseline in DB period
Worsening of Suicidal Ideation	Most severe suicidal ideation in any post-baseline in DB period is more severe than the most severe value at baseline

7.4.7.1 Treatment Period

The number and percentage of subjects with following categories will be summarized by treatment for lifetime, baseline and any post-baseline in DB period, respectively.

- each type of suicidal ideation,
- any suicidal ideation,
- each type of suicidal behavior,
- any suicidal behavior,
- any having self-injurious behavior without suicidal intent,
- any suicidality

At any post-baseline visit in DB period, comparisons between each dasotraline group and placebo group for suicidality will be conducted using Fisher's exact test.

In addition, the number and percentage of subjects with following categories will also be summarized by treatment for any post-baseline in DB period.

- emergence of suicidal ideation,
- emergence of serious suicidal ideation,
- worsening of suicidal ideation,
- emergence of suicidal behavior,
- emergence of suicidality

Shift from baseline to any post-baseline in DB period in suicidal ideation score will be presented by treatment to show the percentage of subjects with values 0 ('No ideation present') to 5 ('Active ideation with plan and intent'). Descriptive statistics (N, mean, standard deviation, median, and range) will be reported for Ideation intensity total score at baseline, each scheduled post-baseline time point, and endpoint for observed value as well as change from baseline. Data collected on actual suicide attempts (lethality of actual attempts and potential lethality of attempts) will be presented in a data listing.

All summaries pertaining to the C-SSRS will be based on subjects in the safety population.

7.4.7.2 Withdrawal Period

The number and percentage of subjects with following categories will be summarized by treatment for any post-withdrawal baseline for withdrawal population.

- each type of suicidal ideation,
- any suicidal ideation,
- each type of suicidal behavior,
- any suicidal behavior,
- any having self-injurious behavior without suicidal intent,
- any suicidality

7.4.8 Other Safety Assessments

7.4.8.1 Breath Alcohol Test

The number and percentage of subjects at each assessment (negative/positive) will be summarized at each visit by treatment group.

7.4.8.2 Neurological Examination

Frequency distribution of each category of assessment (ie, Cranial Nerves, Motor System, sensory System, Reflexes, Coordination, Gait and Rhomberg's test) will be summarized by

treatment at each visit. A shift table of each assessment for baseline to the worst case during treatment period will also be presented by treatment.

7.4.8.3 Abuse Potential Monitoring Plan (APMP)

The Abuse Potential Monitoring Plan (APMP) for dasotraline has been designed to detect potential abuse of the compound and to more closely monitor AEs consistent with the pharmacology. The plan detects irregularities in the handling of dasotraline in clinical trials and identifies the misuse of dasotraline or other psychoactive substances. Moreover, the APMP provides a process by which events subject to additional monitoring are identified, processed, and reviewed. Events subject to additional monitoring are identified based on the pharmacology of dasotraline, as well as adverse event profiles of compounds with similar mechanisms of action.

APMP results will be provided by sites before DBL and provided in a listing.

7.4.8.3.1 Medication Handling Irregularities

Medication handling irregularities will be summarized using the following categories:

- Drug accounting errors, not involving suspected abuse or diversion by subject
- Non-compliance with study procedures, not involving suspected abuse or diversion by subject
- Other cases not involving suspected abuse or diversion by subject
- Suspected or known abuse or diversion of study drug
- Suspected abuse (nonmedical use) of alcohol, illicit substances, OTC drugs or prescription drugs obtained outside the study protocol

Instances that meet the threshold criteria for a medication handling irregularity will be summarized descriptively. Both the number of instances and the number (%) of subjects experiencing medical drug irregularities will be provided for each category and by treatment. Instances of medication handling irregularity will be listed for all randomized subjects.

7.4.8.3.2 Events Subject to Additional Monitoring (ESAMs)

In addition to monitoring for irregularities in medication handling, AEs that may be suggestive of a developing abuse potential, will receive special attention. The adverse experiences or related signs and symptoms that will require additional monitoring (ESAMs) are listed in Attachment 9.6. AEs identified as ESAMs will be summarized by ESAM type and PT by treatment for safety population (treatment period) and withdrawal population (withdrawal period), respectively. ESAMs will also be listed for all randomized subjects.