Statistical Analysis Plan

Protocol Number: MT-8554-A01

A randomized, double-blind, placebo-controlled study to assess the effect of MT-8554 on the frequency and severity of vasomotor symptoms in postmenopausal women

Version: Final 1.0
Date: 1 Nov 2018

NCT number: NCT03291067

Statistical Analysis Plan

Protocol MT-8554-A01

A Randomized, Double-Blind, Placebo-Controlled Study to Assess the Effect of MT-8554 on the Frequency and Severity of Vasomotor Symptoms in Postmenopausal Women

PREPARED BY:

NAME: [redacted]
POSITION: Statistician
VERSION: Final 1.0
DATE: 1 Nov 2018

REVIEWED BY:

NAME: [redacted]
POSITION: Statistician
DATE: 1 Nov 2018

APPROVAL FORM

STATISTICAL ANALYSIS PLAN

Protocol No. MT-8554-A01

Protocol Title **A randomized, double-blind, placebo-controlled study to assess the effect of MT-8554 on the frequency and severity of vasomotor symptoms in postmenopausal women**

Version Final 1.0

Date [redacted]

Author

Signature

APPROVED BY:

MTDA Statistics Approver

Name: [redacted] Date: [redacted] Signature: [redacted]

Position: Head of Data Science, MTDA

MTPC CP Approver

Name: [redacted] Date: [redacted] Signature: [redacted]

Position: Head of CP

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
AST	Aspartate aminotransferase
BDRM	Blind data review meeting
BLQ	Below the limit of quantification
BMI	Body mass index
BUN	Blood urea nitrogen
CI	Confidence interval
CPK	Creatine phosphokinase
CRF	Case Report Form
CRO	Contract Research Organization
CS	Clinically significant
CSR	Clinical study report
CV	Coefficient of variation
CV%	Coefficient of variation percentage
DMC	Data Monitoring Committee
DB	Double Blind
DP	Decimal places
E2	Estradiol
ECG	Electrocardiogram
eCRF	Electronic Data Collection Form
EOS	End of study
EOT	End of treatment
FDA	Food and Drug Administration
Fmi	Daily frequency of mild VMS
Fmo	Daily frequency of moderate VMS
Fse	Daily frequency of severe VMS
FSH	Follicle stimulating hormone

Abbreviation	Definition
GAD7	7-Item Generalized Anxiety Disorder questionnaire
GCP	Good Clinical Practice
IAO	International Agreed Order
ICF	Informed Consent Form
ICH	International Conference on Harmonization
IEC	Independent Ethics Committee
IRB	Institutional Review Board
IMP	Investigational medicinal product
IND	Investigational New Drug Application
ISI	Insomnia Severity Index
ITT	Intent-to-Treat
IVRS	Interactive Voice Response System
IWRS	Interactive Web-based Response System
LDH	Lactate dehydrogenase
LDL-C	Low density lipoprotein-cholesterol
LH	Luteinizing hormone
LLOQ	Lower limit of quantification
LOCF	Last observation carried forward
LS mean	Least squares mean
MCH	Mean corpuscular hemoglobin
MCHC	Mean corpuscular hemoglobin concentration
MCV	Mean corpuscular volume
MedDRA	Medical Dictionary for Regulatory Activities
MENQOL	Menopause-Specific Quality of Life questionnaire
MsFLASH	Menopausal Strategies: Finding Lasting Answers to Symptoms and Health
MTDA	Mitsubishi Tanabe Pharma Development America, Inc.
MTPC	Mitsubishi Tanabe Pharma Corporation, Inc.
NCS	Not clinically significant
PGIC	Patient Global Impression of Change
PHQ-8	8-Item Patient Health Questionnaire
PK	Pharmacokinetic(s)

Abbreviation	Definition
PP	Per-Protocol
PSQI	Pittsburgh Sleep Quality Index
PT	Preferred term
QC	Quality control
QOL	Quality of life
QTc	Corrected QT interval
QTcF	Corrected QT interval using Fridericia's formula
RBC	Red blood cells
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Standard deviation
SE	Standard error
SF-36	36-Item Short Form Health Survey
SNRI	Serotonin-norepinephrine reuptake inhibitors
SOC	System Organ Class
SSRI	Serotonin reuptake inhibitors
SUSAR	Suspected unexpected serious adverse event
TEAE	Treatment-emergent adverse event
THC	Tetrahydrocannabinol
ULN	Upper limit of normal
US	United States
VMS	Vasomotor symptoms
WBC	White blood cells
WHO	World Health Organization
WMA	World Medical Association

1. INTRODUCTION

This statistical analysis plan (SAP) is based on the MT-8554-A01 final protocol and amendment 5 dated 22 March 2018. The plan covers statistical analysis, tabulations and listings of efficacy, safety data to assess the efficacy, safety and PK of 3 doses of MT-8554 compared to placebo.

The SAP is prepared by MTDA data science and reviewed by MTDA clinical study team and MTPC data science. The statistical analyses and production of the outputs described in the SAP will be conducted and QCed by [redacted] using SAS version 9.4 or higher. The final analyses and outputs will be approved by MTPC/MTDA data science.

1.1 Study Objectives

Primary Objective:

- To assess the efficacy of MT-8554 on the frequency and severity of VMS in postmenopausal women.

Secondary Objectives:

- To assess the dose-response effect and the minimum effective dose of MT-8554 on the frequency and severity of VMS.
- To assess the effect of MT-8554 on subjective sleep quality as measured by diary or insomnia questionnaires.
- To assess the safety and tolerability of MT-8554.

1.2 Study Design

This is a Phase II randomized, double-blind, placebo-controlled study for dose selection in women with naturally or surgically induced menopause with moderate to severe VMS.

Subjects meeting eligibility criteria will be enrolled in a 2-week single-blind Placebo Run-in period. All eligible subjects will receive single-blind placebo once daily before bedtime. Following the Placebo Run-in period, subjects meeting eligibility criteria will enter the 12-week, placebo-controlled Double-blind Treatment period. The Double-blind Treatment period has 4 arms including: placebo and MT-8554 1, 5, and 10 mg. A single daily dose of study medication will be administered before bedtime. An End of Study (EOS) Follow-up visit will be conducted by phone for safety follow-up 30 days after the end of the Double-blind Treatment period. Total duration is 20 weeks, inclusive of the Screening and Follow-up periods (see Figure 1).

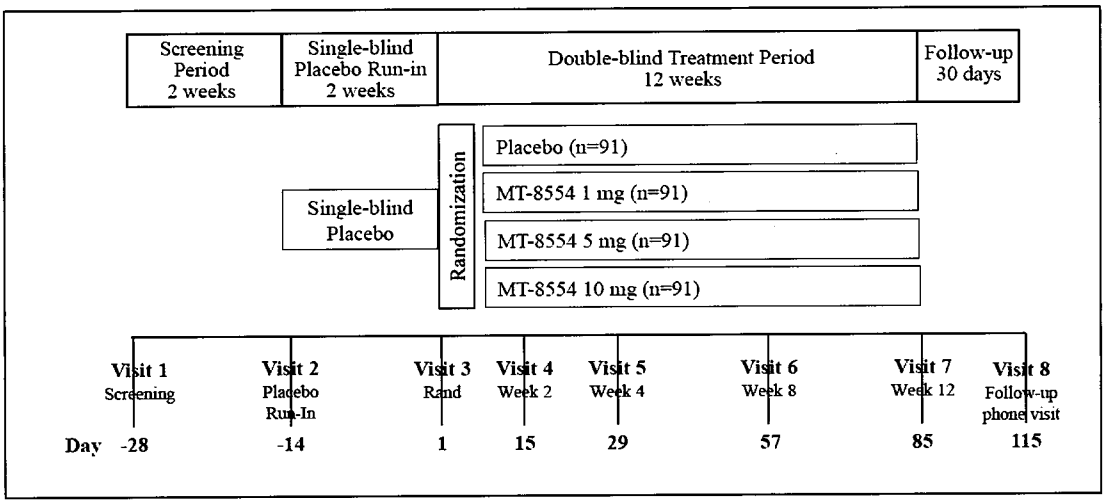
A planned interim assessment for safety, and a planned interim analysis for efficacy (futility) and safety will be conducted during the study; enrollment will proceed without interruption.

[redacted]

[redacted]

The primary endpoint will be evaluated at Weeks 4 and 12.

Figure 1 Study Design Schematic



1.3 Schedule of Study Procedures

Table 1 Schedule of Assessments

Study Period	Screening	Placebo Run-in	Double-blind Treatment					Follow-up
Visit Number	Visit 1	Visit 2	Visit 3 (Randomization)	Visit 4	Visit 5	Visit 6	Visit 7 (EOT)	Visit 8 (EOS)
Study Week	Week -5 to -4	Week -2		Week 2	Week 4	Week 8	Week 12[8]	Week 16[9]
Study Day ±Window	Day -35 to -28	Day -14±1	Day 1	Day 15 ±3	Day 29 ±3	Day 57 ±3	Day 85 ±3	Day 115 ±3
Informed consent	X							
Demography and height	X							
Medical/drug/smoking/alcohol/surgical history	X	X	X					
Inclusion/exclusion criteria	X[1]	X	X					
Physical examination[2] Complete (C) or Routine(R)	C	R	R	R	R	R	R	
Breast examination	X		X				X	
Body weight	X		X				X	
Vital signs	X	X	X	X	X	X	X	
12-lead ECG	X		X				X	
Transvaginal ultrasound[4]	X[10]						X	
Endometrial biopsy[4]	X[10]						X	
Routine lab tests	X[1]	X	X	X	X	X	X	
Reproductive hormones (FSH, LH, E2)	X[1]		X				X	
STOP-BANG Questionnaire	X							
VMS diary[5]	←						→	
ISI, MENQOL, PSQI, and SF-36		X	X		X		X	
PGIC					X		X	
PHQ-8 and GAD7		X	X	X	X	X	X	
Dispensing single-blind placebo[6]		X						

Study Period	Screening	Placebo Run-in	Double-blind Treatment					Follow-up
Visit Number	Visit 1	Visit 2	Visit 3 (Randomization)	Visit 4	Visit 5	Visit 6	Visit 7 (EOT)	Visit 8 (EOS)
Study Week	Week -5 to -4	Week -2		Week 2	Week 4	Week 8	Week 12[8]	Week 16[9]
Study Day ±Window	Day -35 to -28	Day -14±1	Day 1	Day 15 ±3	Day 29 ±3	Day 57 ±3	Day 85 ±3	Day 115 ±3
Dispensing MT-8554 /Placebo[6]			X	X	X	X		
Drug accountability			X	X	X	X	X	
Blood for MT-8554 PK[7]			X	X	X	X	X	
Adverse events	←							→
Concomitant medication	←							→
Compliance [redacted] Calls to Subjects[11]		←				→		

Abbreviations: E2=estradiol; ECG=Electrocardiogram; EOS=End of Study; EOT=End of Treatment Visit; FSH=Follicle stimulating hormone; FU=Follow-up Visit; ISI=Insomnia Severity Index; GAD7=Generalized Anxiety Disorder-7 questionnaire; LH=luteinizing hormone; MENQOL=Menopause-Specific Quality of Life; PGIC=Patient Global Impression of Change; PHQ-8=Eight-Item Patient Health Questionnaire; PK=Pharmacokinetics; PSQI=Pittsburgh Sleep Quality Index; SF-36=36-Item Short Form Health Survey; VMS=Vasomotor symptoms.

1. FSH and/or liver function labs may be retested once within the Screening window for subjects who are ineligible due solely to inclusion/exclusion criteria related to these tests.
2. Physical examination:
   a. Complete: abdominal, cardiovascular, general appearance, head, eyes, ears/nose/throat, lymph nodes, musculoskeletal, neck, neurological, dermatological, respiratory, and other.
   b. Routine: abdominal, cardiovascular, general appearance, respiratory, and other.
3. Vital signs (including sitting blood pressure, pulse rate, and tympanic body temperature) will be measured at all visits.
4. Transvaginal ultrasound and endometrial biopsy to be performed only on subjects who have a uterus.
5. VMS diary data will be collected for 14-days duration during Screening and for 14-days duration during the Placebo Run-in period for eligibility determination.
6. Investigator should instruct subject to administer dose at least 2 hours after starting the evening meal and approximately 30 minutes before bedtime.
7. Blood samples for MT-8554 PK will be collected at Visits 3 through 7. Date and time of most recent dose will be recorded at Visits 4 through 7.
8. The EOT should be performed for subjects who complete as well as those who withdraw from the study early.
9. The Follow-up visit will be conducted by phone 30 days after the EOT visit for subjects who complete as well as those who withdraw early from the study. The Follow-up visit should not be done for subjects enrolling in the long-term extension study, should it be implemented.
10. Baseline assessments for transvaginal ultrasound and endometrial biopsy at baseline may be scheduled and conducted at any time during the Screening period or the Placebo Run-in period, but at least 7 days prior to randomization, to accommodate site specific scheduling. Note: subjects qualifying based on transvaginal ultrasound will proceed to endometrial biopsy.
11. Site staff will make periodic calls to the subjects at a minimum during weeks where no in clinic visits occur, to confirm compliance with daily VMS diary data entry [redacted]

2. STUDY ENDPOINTS

2.1 Co-Primary Efficacy Endpoints

The co-primary efficacy endpoints are defined as:

- Change from baseline in the average daily frequency of moderate to severe VMS

  The average daily frequency of moderate to severe VMS defined as the sum of the number of moderate to severe VMS during 1 week divided by the number of days with data. The daily score here and below are average scores from a 7-day period. See Section 5.4 for more details.

- Change from baseline in the average daily severity score of mild to severe VMS Baseline VMS severity score is defined as (2xFmo + 3xFse)/(Fmo + Fse), and VMS severity score for a specific week during the double-blind treatment period is defined as (1xFmi + 2xFmo + 3xFse)/(Fmi + Fmo + Fse), where Fmi, Fmo, and Fse are the daily frequencies of mild, moderate, and severe VMS, respectively, during each applicable study week. See Section 5.4 for more details.

The Co-primary endpoints will be evaluated at Week 4 and Week 12 for primary analyses.

2.2 Secondary Efficacy Endpoints

Clinical response is defined as the subjects with cutoff number* or greater reduction in the average daily frequency of moderate to severe VMS from baseline.

*The cut-off number will be calculated using anchor-based method. The cutoff number is defined as numerical value to maximize the sum of sensitivity and specificity (i.e. maximize the Youden's index [Youden WJ (1950)], using Patient Global Impression of Change (PGIC) as the anchor.

- Proportion of responders at Weeks 4 and 12 (i.e., subjects with cutoff number or greater reduction in the average daily frequency of moderate and severe VMS compared to baseline)
- Time to response, defined as time (in weeks) from randomization to the first time the subject meets responder criteria (i.e., cutoff number or greater reduction from baseline in the average daily frequency of moderate and severe VMS)
- Change from baseline to Weeks 4 and 12 in the Insomnia Severity Index (ISI) total score

2.3 Other Efficacy Endpoints

- PGIC at Weeks 4 and 12
- Change from baseline to Weeks 4 and 12 in the Pittsburgh Sleep Quality Index (PSQI) total score
- Change from baseline to Weeks 4 and 12 in the total score and 4 domain scores of Menopause-Specific Quality of Life (MENQOL)
- Change from baseline to Weeks 4 and 12 in the domain scores of 36-Item Short Form Health Survey (SF-36)

- Change from baseline in the average daily severity score of moderate to severe VMS at Weeks 4 and 12, defined as 2xFmo + 3xFse

2.4 Safety Assessments

2.4.1 Adverse events (AEs)

An AE is any untoward medical occurrence in a subject or clinical investigation subject administered a pharmaceutical product that does not necessarily have a causal relationship with this treatment.

AEs will be coded using Medical Dictionary for Regulatory Activities (MedDRA) central coding dictionary, Version 20.0 or the later (to be updated at final version).

Treatment-emergent AEs (TEAEs) are defined as AEs that started or worsened in severity on or after the first dose of double-blind study medication. The treatment relationship with the AE is in 2 categories (reasonable possibility, no reasonable possibility). The detail is included in protocol.

2.4.2 Physical examination

The complete physical examination consists of a routine assessment of major body systems: abdominal, cardiovascular, general appearance, head, eyes, ears/nose/throat, lymph nodes, musculoskeletal, neck, neurological, dermatological, respiratory, and 'other'.

The routine physical examination consists of a routine assessment of the following body systems: abdominal, cardiovascular, general appearance, respiratory, and 'other'.

The complete physical examination is performed at Screening. The routine physical examination is performed at run-in period and Visits 3 to 7.

2.4.3 Breast examination

The breast examination consists of symmetry of breast shape, contour of breast, appearance of skin, nipple, areola, lymph nodes status, presence and characterization of lesions. The breast examination is performed at Screening, Visits 3 and 7.

2.4.4 Vital signs

- Sitting Systolic Blood Pressure (mmHg)
- Sitting Diastolic Blood Pressure (mmHg)
- Sitting Pulse Rate (bpm)
- Tympanic Body Temperature (^{0}C)
- Body Weight (kg)

2.4.5 ECG parameters

- Heart rate (bpm)
- PR (msec)

- RR (msec)
- QRS (msec)
- QT (msec)
- QTcF (msec)
- QTcB (msec)

2.4.6 Clinical laboratory assessments

The clinical lab tests are listed in the Table 2.

Table 2 Routine Laboratory Tests

Hematology:	
Hemoglobin	Mean corpuscular hemoglobin
Hematocrit	Mean corpuscular hemoglobin concentration
Platelet count	Mean corpuscular volume
Red blood cell count	White blood cell count and differential
Biochemistry:	
Alkaline phosphatase	Cholesterol
Aspartate aminotransferase	Triglycerides
Alanine aminotransferase	High density lipoprotein-cholesterol
Gamma-glutamyl transpeptidase	Low density lipoprotein-cholesterol
Potassium	Protein (total)
Sodium	Albumin
Chloride	Creatine kinase
Inorganic phosphate	Creatinine
Glucose	
Urea	
Bilirubin (direct and total)	
Coagulation:	
Prothrombin time	Activated partial thromboplastin time
International normalized ratio	
Urinalysis:	
Specific gravity, pH, protein, glucose, ketones, urobilinogen, blood	
Microscopic examination[1]	

[1] Performed only if required, based on urinalysis results.

2.4.7 Reproductive hormones

- Luteinizing hormone (LH)
- Follicle stimulating hormone (FSH)
- Estradiol (E2)

2.4.8 Endometrial safety

- Endometrial thickness, as measured by transvaginal ultrasound
- Incidence of endometrial hyperplasia, as measured by endometrial biopsy

2.4.9 Depression and Anxiety

Depression as measured by 8-Item Patient Health Questionnaire (PHQ-8). The total scores of PHQ-8 will be used as the safety endpoint.

Anxiety as measured by the 7-item Generalized Anxiety Disorder questionnaire (GAD7). The total scores of GAD7 will be used as the safety endpoint.

2.5 Pharmacokinetics (PK) Evaluations

MT-8554 concentrations in plasma will be measured at 5 visits during the Double-blind Treatment period. Planned time points for pharmacokinetics (PK) evaluation are as follows: Visit 3 (Day 1), Visit 4 (Week 2), Visit 5 (Week 4), Visit 6 (Week 8) and Visit 7 (Week 12).

Population PK analysis will be performed using the plasma concentrations of MT-8554 obtained in this study alone and/or in combination with data obtained from other clinical studies. Details of the population PK analysis will be presented in a modelling plan. Population PK analysis results will be reported separately from the clinical study report.

2.6 Demographic and Other

2.6.1 Demographic and Baseline Characteristic

The following subject demographic and baseline characteristic are collected.

- Age (years)
- Sex
- Race
- Ethnicity
- Weight (kg)
- Height (cm)
- BMI (kg/m^2)
- Menopausal status
- Antidepressant use
- Diabetic status
- Cardiovascular history
- Hypertension
- Heart Failure
- Baseline labs (creatinine, the estimated glomerular filtration rate [eGFR], haematocrit)
- Average daily VMS frequency
- Average daily VMS severity

2.6.2 Medical History

The subjects Medical and surgical history will be collected and recorded in the clinical database.

2.6.3 Prior or Concomitant Medications

Prior medications are defined as any medication taken prior to the double-blind treatment period. Any prior medication, including prescription and over-the-counter medications, taken within 1 month prior to Screening will be collected.

Concomitant medication is defined as any medication, other than study medication, which is taken on or after the start day of double blind treatment.

The prior and concomitant medication data will be coded using World Health Organization Drug Dictionary (WHO Herbal Dictionary).

3. PLANNED ANALYSES

Detail of what Listing, Table and Figure will be created at each analysis term is described in Section 11 List of Listings, Tables and Figures.

3.1 Data Monitoring Committee (DMC)

An independent DMC, composed of experts in the management of subjects with the disease under study will be established for MT-8554-A01 according US FDA guidance. An independent biostatistician will be assigned to the DMC to maintain the integrity of the study blind and provide safety data at regular predefined intervals during the study and provide the interim analysis to the DMC.

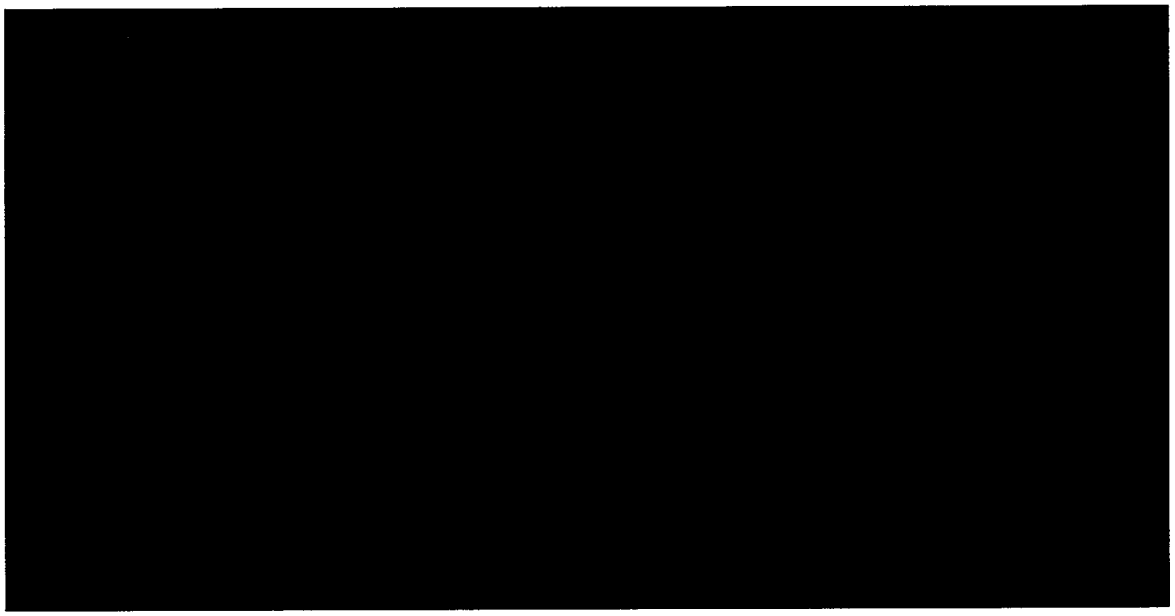
The primary purpose of this committee will be to review safety data for the protection of subject safety. In addition the DMC will review efficacy data for the interim analysis and make recommendations to the Sponsor. When approximately 50% of subjects complete the Week 12 visit, the planned interim analysis will be conducted to determine whether the study will be stopped due to futility, safety concerns, or both.

A DMC charter will outline the scope and key responsibilities, timing of reviews, communications between the DMC and the Sponsor, and recommendations and action rules for the study.

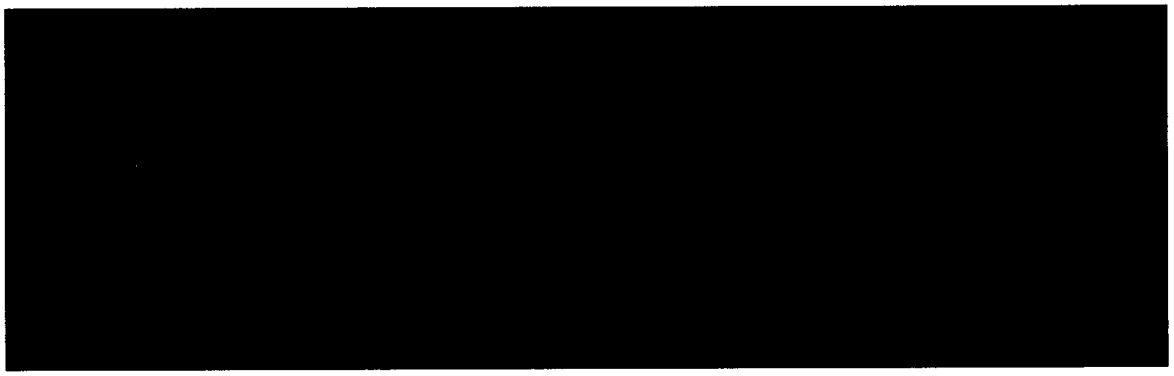
3.2 Interim Analyses

In addition to on-going safety monitoring, there will be one planned interim assessment for safety and one planned interim analysis for efficacy (futility) and safety.

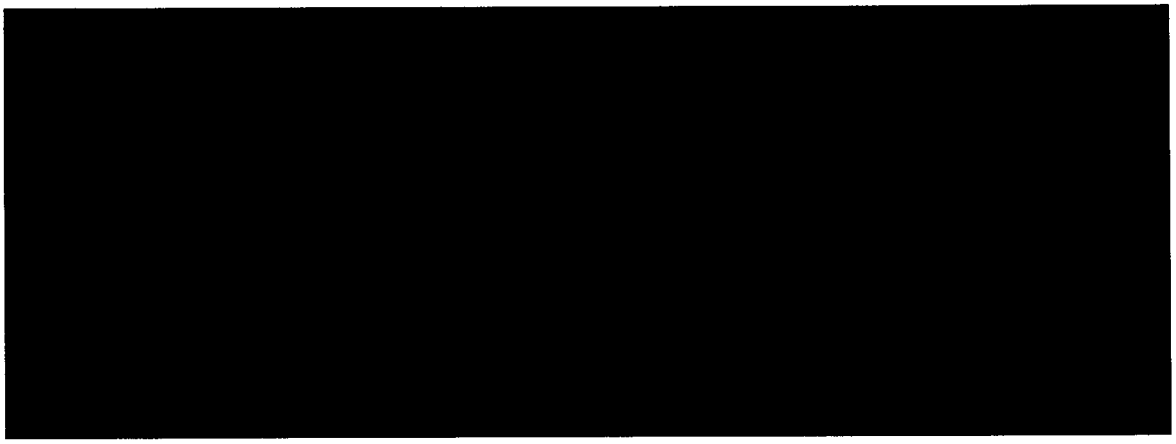
3.2.1 Interim Safety assessment



3.2.2 Interim Analysis for Safety and efficacy



Then DMC will evaluate the interim analysis results. The DMC will review the data and provide recommendations for the study to the Sponsor.



[redacted]

3.3 Final Analysis

The final analysis will be conducted with unblinded data after database lock.

All the planned efficacy and safety analyses in this SAP will be performed by [redacted]

3.4 Other Analysis

There is no plan for analyses other than what is defined in this SAP.

4. ANALYSIS POPULATION(S)

4.1 Randomized Population

Randomized Population includes all subjects who are randomized, regardless of whether they receive any study drug. Data from the all randomized subjects analysis set will be analyzed by the treatment group assigned at the time of randomization, even if the subject does not receive the correct treatment, is not compliant with the protocol or does not follow through with the study until completion.

4.2 Safety (SAF) Population

Safety (SAF) Population includes all randomized subjects who received at least 1 dose of study medication. The safety population will be used for all safety analyses. The subject exact treatment received will be used for safety analyses.

4.3 Intent-to-treat (ITT) population

Intent-to-treat (ITT) population: includes all randomized subjects who have at least 1 post-baseline value for primary efficacy endpoint. The ITT population will be used for all efficacy analyses. The subject randomized treatment will be used for efficacy analyses.

4.4 Per-protocol (PP) population

Per-protocol (PP) population: includes all ITT subjects who do not have any major protocol deviations. The PP population will be used for primary efficacy analyses as well. The subject randomized treatment will be used for efficacy analyses.

4.5 Pharmacokinetic (PK) population

PK population includes all randomized subjects who receive at least 1 dose of MT-8554 and who have at least 1 post-dose value for plasma concentration without important protocol deviations which may affect the PK of MT-8554.

5. GENERAL CONSIDERATIONS

The statistical evaluation will be performed using SAS® Version 9.4 or later.

Efficacy summaries will be performed on the ITT. As part of sensitivity analysis will be performed on the PP population. Safety summaries will be performed on the SAF. PK summaries will be performed on the PK population.

5.1 Number of digits to report

Table 3 Number of decimal places (DP) or significant digits (SD)

Statistic	Specification	Apply to
Minimum, maximum	same number of DPs as the data provided in the datasets	All original, i.e. non-derived, data provided in the datasets
mean, median, , confidence intervals, lsmean	one more DP than the raw data	All
SD, SE	two more DP than the raw data	All
Percentages	1 DP	All
p-values	3 DP	All
Odds Ratio	2DP	All
Hazard Ratio	3DP	All

5.2 Significance level and confidence interval

The statistical tests will be performed as two-sided with significance level of 5%. The confidence intervals will be determined with a confidence level of 95%.

5.3 Descriptive statistics values to calculate

Where appropriate, variables will be summarized descriptively (frequency and percent will be summarized for categorical variables; mean, standard deviation [SD], median, minimum, and maximum will be presented for continuous variables) by study visit and by treatment group.

The denominator for the percentages will be the total number of subjects in the treatment group and Analysis Set being presented, unless otherwise specified (e.g. on some occasions, percentages maybe calculated out of the total number of subjects with available data at a particular visit and/or time point).

Summary statistics required for PK or related data are described in relevant sections later.

5.4 Derived variables

The subjects' diary data will be used to derive the primary and some secondary efficacy endpoints. Subjects can use paper diary as contingent back up when they have problem with their ediary device. The paper diary data will be used only for the days that subjects have paper diary without ediary data. This means the paper diary data is not used when ediary data is available.

5.4.1 Definition of the average daily frequency of moderate to severe VMS

The average daily frequency of moderate to severe VMS at a time point (Baseline, Week 1, 2, 3, 4, ..., 12) is the average of the frequency of moderate to severe VMS of available diary days in a 7-day window. It will be rounded off the second decimal place and reported the value to the first decimal place. The nominal day for each time point relative to the first dose day will be used. The analysis windows are defined in the table below. The other efficacy endpoints based on average daily VMS diary data will be derived using these windows accordingly.

Table 4 Analysis visit windows of diary days used for average daily frequency of moderate to severe VMS

	Nominal day	Window
Baseline		Day -7 to -1
Week 1	8	Day 2 to 8
Week 2	15	Day 9 to 15
Week 3	22	Day 16 to 22
Week 4	29	Day 23 to 29
Week 5	36	Day 30 to 36
Week 6	43	Day 37 to 43
Week 7	50	Day 44 to 50
Week 8	57	Day 51 to 57
Week 9	64	Day 58 to 64
Week 10	71	Day 65 to 71
Week 11	78	Day 72 to 78
Week 12	85	Day 79 to 88

VMS data on the day of Week 12 visit is not used for the calculation of average daily VMS frequency.

If the count of available VMS data in a specific week is less than four (4), the latest diaries in the previous window can be carried as needed to make 4 diaries available in this window. Then the average daily VMS frequency in the week is calculated.

5.4.2 Definition of the average daily severity score of mild to severe VMS

The daily VMS severity score is defined as (2xFmo + 3xFse)/(Fmo + Fse) for any day before the first dose day, and (1xFmi + 2xFmo + 3xFse)/(Fmi + Fmo + Fse) for any day on or after the first dose day, where Fmi, Fmo, and Fse are the daily frequencies of mild, moderate, and severe VMS, respectively. It will be rounded off the third decimal place and reported the value to the second decimal place.

The avarage daily VMS severity of mild to severe VMS at a time point (Baseline, Week 1, 2, 3, 4, ..., 12) is the average of the daily severity of available diary days in the corresponding 7-day window defined in Table 4.

A daily severity score of zero (0) is assigned if it is reported that no VMS occurred on that day. VMS data on the day of Week 12 visit is not used for the calculation of average daily VMS severity. If the count of available VMS data in a specific week is less than four (4), latest diaries in the previous window can be carried as needed to make 4 diaries available in this window. Then the average daily VMS severity in the week is calculated.

5.4.3 Definition of clinical responder

A clinical responder will be defined as a subject with a cutoff number or greater reduction in the average daily frequency of moderate and severe VMS compared to baseline. The cutoff number of responder criteria will be calculated using anchor-based method, using PGIC as the anchor. The satisfied subjects will be defined as those whose PGIC response is either 6 or 7. The cutoff number is defined as numerical value to maximize the sum of sensitivity and specificity (i.e. maximaize the Youden's index[1]) for change from baseline in average daily frequency of moderate to severe VMS. This calculation will use pooled data from all four study arms and both Week 4 and 12 time points.

5.4.4 Definition(s) of baseline(s)

Unless otherwise specified, the baseline values are the last available assessment before the first dose of randomized treatment.

5.4.5 Definition(s) of (percent) change from baseline(s)

For numerical variables, change from baseline will be calculated as the post-baseline value minus the baseline value. If percent change from baseline is required, then percent change from baseline will be calculated as the change from baseline divided by the baseline value, multiplied by 100. If baseline value cannot be determined for a particular variable, the change from baseline and percent change from baseline will not be calculated.

5.4.6 Derived visits

The derived analysis visit windows are outlined in Table below.

Table 5 Derived Analysis Visit Windows

	Nominal day	Window
Baseline	NA	NA
Day 1	Day 1 (First dose day)	NA
Week 2	Day 15	Day 7 to 21
Week 4	29	Day 22 to 42
Week 8	57	Day 43 to 70
Week 12	85	Day 71 to 98
Week 16 (FU)	115	NA

The analysis visits are derived according to the following criteria:

- The unscheduled visits are not used for the deriving.
- If a study visit is the only one in an analysis window, this study visit becomes the derived analysis visit.
- If there are multiple visits in a visit window, the closest visit to the nominal day becomes the analysis visit. In the event that two visits are the closest visits to the nominal day and equally distanced to the nominal day, the visit after the nominal day becomes the derived visit.
- The follow-up visit will not be derived. The study visit will be used.
- Unscheduled visits and retests (same visit number assigned), will not be displayed in by-visit summary tables, but will be included in the data listings. All data will be listed. Listings will include treatment, scheduled, unscheduled, retest and early discontinuation data.
- The screening and run-in visits are carried to derived visits
- The efficacy endpoints based on diary data use different windows. See Section 5.4.1.

5.4.7 BMI

BMI (kg/m^2) = Body weight (kg)/{Body height (m)}^{2}

It will be rounded off the second decimal place and reported the value to the first decimal place.

5.4.8 Medical history and Adverse events

All medical history and adverse events will be coded from the actual term using the MedDRA version 20.0 or the later version.

5.4.9 Adverse reactions

Adverse reactions are defined as adverse events that are determined to have a "reasonably possible" causal relationship to the study drug.

5.4.10 Duration of the AE and time to the AE

Duration of the AE and time to the AE occurrence from start of study medication will be calculated and presented in days, where AE duration = AE stop date – AE start date + 1 and the time to the AE occurrence = AE start date – first DB dose date + 1.

5.4.11 Prior and Concomitant medications

All prior and concomitant medications will be coded using the WHO drug dictionary (WHO Herbal Dictionary) (Version SEP 2017).

5.4.12 Duration of exposure and Cumulative Dosing Count

The duration of exposure is calculated as the total number of days that the subject has been treated with double-blind study medication—that is, from the treatment start date to the date of their EOT/withdrawal visit. For subjects lost to follow up, the treatment end date is taken to be the date of their last visit. Interruptions, compliance, and dose changes are not taken into account for duration of exposure.

Duration of exposure (days) = date of end of treatment visit – date of first dose of double-blinded study medication; Cumulative Dosing days= Duration of exposure – number of days subject miss the medication

5.4.13 Treatment Compliance

Compliance with double-blind study medication—based on the drug accountability data—will be calculated as the number of days subjects take study medication divided by the duration of exposure in days, expressed as a percentage.

Study medication compliance will be calculated as follows:

$$\frac{\text{Cumulative Dosing days}}{(\text{date of end of treatment visit} - \text{date of randmization})} \times 100\%$$

The subjects should take one dose daily per protocol.

5.4.14 The Estimated Glomerular Filtration Rate (eGFR)[5]

$eGFR = 141 \times \min(S_{Cr}/\kappa, 1)^{\alpha} \times \max(S_{Cr}/\kappa, 1)^{-1.209} \times 0.993^{Age} \times 1.018$ [if female] x 1.159 [if Black]

S_{Cr} (standardized serum creatinine) = mg/dL

$\kappa = 0.7$ (females) or 0.9 (males)

$\alpha = -0.329$ (females) or -0.411 (males)

min = indicates the minimum of S_{Cr}/κ or 1

max = indicates the maximum of S_{Cr}/κ or 1

Age = years

It will be rounded off the second decimal place and reported the value to the first decimal place.

5.4.15 MENQOL scores deriving

The MENQOL is self-administered and consists of a total of 29 items in a Likert-scale format. Each item assesses the impact of one of four domains of menopausal symptoms, as experienced over the last month:

- vasomotor (items 1–3)
- psychosocial (items 4–10)
- physical (items 11–26)

- sexual (items 27–29)

Items pertaining to a specific symptom are rated as present or not present, and if present, how bothersome on a zero (not bothersome) to six (extremely bothersome) scale. Means are computed for each subscale by dividing the sum of the domain's items by the number of items within that domain. Non-endorsement of an item is scored a "1" and endorsement a "2," plus the number of the particular rating, so that the possible score on any item ranges from one to eight. (Hilditch JR, et al. 1996) Total MENQOL score will be calculated as the average of each domain score.

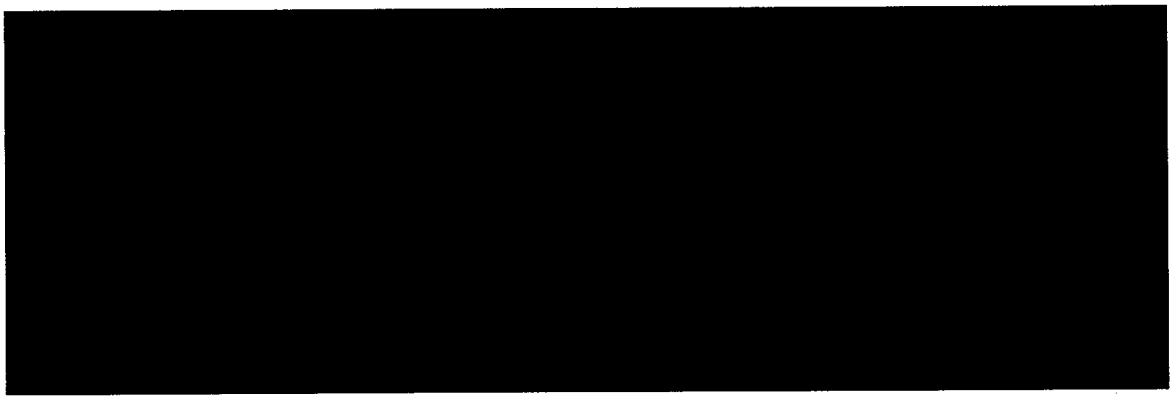
5.4.16 Lab values below/above the limit of quantitation deriving

The lab values below the limit of quantification (BLQ) are collected in the form like "< 2" in the clinical data base. In this case the numeric value of the lab is missing value. The cut-off value '2' will be used to impute the missing value for this case. Similarly, for the lab values above the limit of quantification, the cut-off value will be used for the numeric value of the lab.

Table 6 Imputation of lab values below/above the limit of quantification

Lab Test	Characteristic value	Derived Numeric value	
Bilirubin	<2	2	
Direct Bilirubin	<2	2	
Estradiol	<61	61	
Luteinizing Hormone	<0.1	0.1	
Specific Gravity	>1.045	1.045	

6. SAMPLE SIZE AND POWER CONSIDERATIONS



7. STATISTICAL METHODOLOGY

7.1 Blinded Data Review

Prior to database lock, a blinded data review meeting (BDRM) will be conducted. Protocol deviations, protocol defined analysis populations will be confirmed during the meeting. Additional data handling rules may be introduced as results of data review. Should additional

data handling rules be confirmed during the meeting, they will be incorporated into the analyses planned in this SAP.

PK data that are considered "invalid" will be flagged in the listing. The PK data handling will be assessed during BDRM prior to database lock and in the investigation of PK data handling assessment after unblinding. A separate PK data handling document will be produced to cover both pre- and post- unblinding decisions. If PK sample handling errors or other factors identified after data unblinding and these errors have led to unexpected erroneous data, then these erroneous data will be regarded as "invalid".

7.2 Disposition of Subjects

The number of subjects who enter screening will be summarized, and the percentage of these subjects who fail to meet entry criteria will be reported for total subjects.

Screen failures will be summarized in total and by each reason for screen failure.

The number of subjects who enter run in will be summarized. Run in failures will be summarized in total and by each reason for run in failure. (Table 14.1.1.2)

Subject disposition will be summarized for subjects completion status and discontinue reasons in tables for each treatment group, MT-8554 overall and total subjects.

The number and percentage of subjects in each analysis population will be summarized for each treatment group, MT-8554 overall and total subjects. (Table 14.1.1.1)

The data listing for Subject disposition will be generated.

7.3 Demographic and Other Baseline Characteristics

The subject demographic and baseline characteristic will be summarized by treatment in table and presented in data listing for the safety population. (Table 14.1.2)

7.4 Medical History

The subjects medical and surgical history will be summarized by treatment, SOC and PT in table and presented in data listing for the safety population. (Table 14.1.3)

The count and percentage of subjects who had at least one medical history or surgical history will be presented. For each medical history or surgical history term, the count and percentage of subjects will be presented.

In the table, SOC is sorted by International order; then within SOC, PT is sorted by descending counts under MT-8554 Total column, then descending counts under Placebo column, then alphabetic order for PTs with the same count.

7.5 Prior or Concomitant Medications

The prior and concomitant medications will be summarized in table by treatment group and presented in data listing for the SAF population. (Table 14.1.4.1, Table 14.1.4.2) The data will be coded using World Health Organization Drug Dictionary (WHO Herbal Dictionary). The summaries will be by ATC level 2 and preferred name.

The count and percentage of subjects who had at least one prior or concomitant medication will be presented. For each prior or concomitant medication term, the count and percentage of subjects will be presented.

Prior medications are those which stopped prior to first dose of double-blind study medication. Concomitant medications are medications that started prior to, on or after date of first dose of double-blind study medication and ended on or after date of first dose of double-blind study medication. This includes medications deemed as ongoing at the end of study.

7.6 Study Medication Exposure

Duration of exposure to double-blind study medication in days will be summarized in table and presented in data listing for the safety population. (Table 14.1.5)

7.7 Treatment Compliance

Compliance to Double-Blind study medication will be presented for the ITT population in table by treatment. (Table 14.1.6)

All study medication administration and accountability data will be listed by subject.

7.8 Protocol Deviations

The major protocol deviations will be presented in data listings.

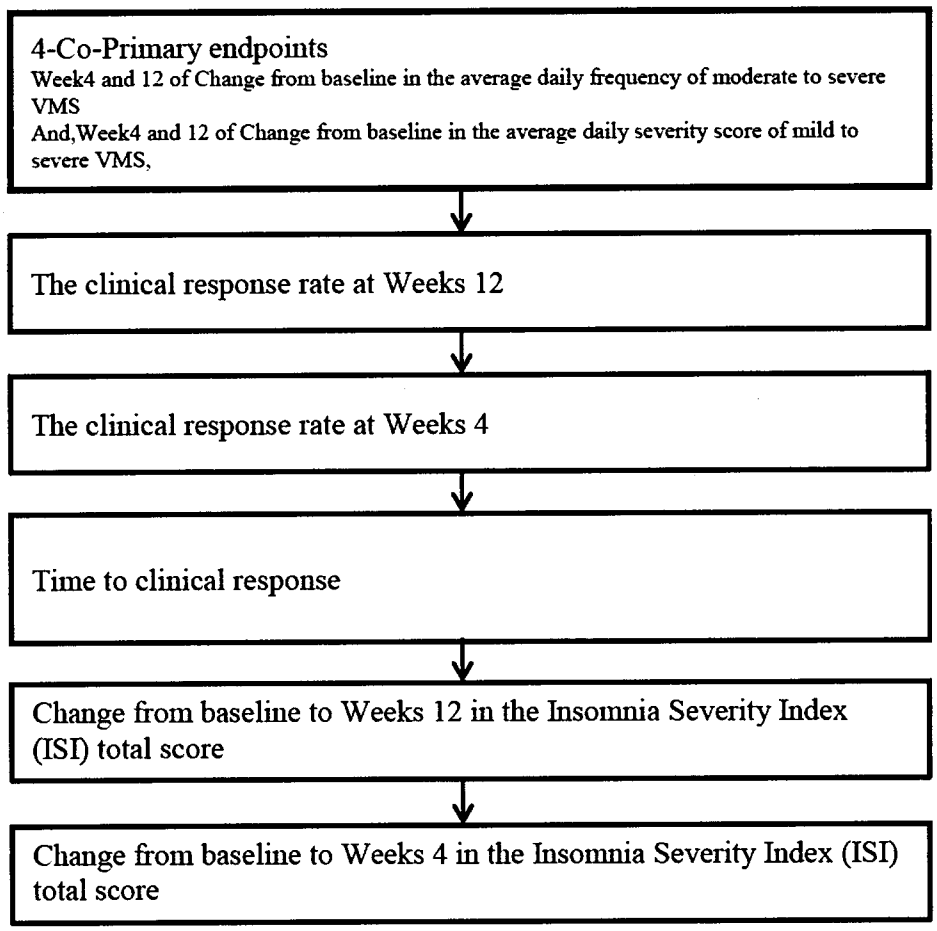
7.9 Efficacy Assessments

All Efficacy Tests data will be listed.

7.9.1 Primary Efficacy Analysis

Unless otherwise noted, these analyses are performed using ITT population.

In order to deal with multi-endpoints multiplicity, the primary efficacy analyses of primary and secondary endpoints will be judged to statistical significance in the order shown below.



7.9.1.1 Efficacy Analysis for Co-Primary Efficacy Endpoints

Change from baseline to Week 4 and Week 12 in the average daily frequency of moderate to severe VMS will be analyzed using repeated measures ANCOVA model with an unstructured covariance structure. The analysis model will include baseline value of the endpoint as a covariate and visits, treatment, interaction between visit and treatment, interaction between baseline and visit, antidepressant using (Y,N) , and interaction between antidepressant using (Y,N) and treatment as fixed effect. Baseline, weeks 1, 2, 3, 4, ..., 12 will be used for this analysis. (Table 14.2.1.1.A)

[redacted]

[redacted]

Similar analyses will be conducted for VMS severity score. (Table 14.2.1.2.A). The primary efficacy analysis will also be performed using PP population. (Table 14.2.1.1.B, Table 14.2.1.2.B)

If the normality assumption for the model is not met, the non-parametric method ranked ANOVA will be performed.

The scatter Plot of Mean change from baseline in VMS frequency and severity by treatment will be generated. (Figure 14.2.3.1, Figure 14.2.3.2). The scatter Plot of observed mean values of VMS frequency and severity by treatment will be generated. (Figure 14.2.4.1, Figure 14.2.4.2). The bar graph of Mean change from baseline in VMS frequency and severity at Weeks 4 and 12 by treatment will be generated. (Figure 14.2.5.1, Figure 14.2.5.2) The box plot of change from baseline in VMS frequency and severity by treatment will be generated as well. (Figure 14.2.6.1, Figure 14.2.6.2)

7.9.1.2 Efficacy Analysis for the clinical response rate

The clinical response rate will be analyzed using Logistic regression with treatment as fixed effects at Weeks 4 and 12. (Table 14.2.5)

The ROC curve will be created and cutoff number for responder will be calculated. (Figure 14.2.1)

7.9.1.3 Efficacy Analysis for the time to clinical response

The Log-rank test will be used to analyze the time to clinical response. (Table 14.2.6) And Kaplan-Meier curves will be created. (Figure 14.2.2) For subjects who do not have reached to clinical response, treat their last visits as censoring.

7.9.1.4 Efficacy Analysis for the Insomnia Severity Index

The change from baseline to Weeks 4 and 12 in the ISI total score will be analyzed using ANCOVA model with treatment as fixed effects and baseline value as covariate (Table 14.2.7.1). In addition, ISI Total Score in categories will be summarized for each visit (Table 14.2.7.2). The following categories will be used for the analysis;

- No clinical significant insomnia: ISI total score 0 – 7
- Subthreshold insomnia: ISI total score = 8 – 14
- Clinical insomnia (moderate): ISI total score = 15 – 21
- Clinical insomnia (severe): ISI total score = 22 – 28

7.9.2 Secondary or other exploratory Efficacy analysis

7.9.2.1 Efficacy Analysis for paired comparison as Secondary analysis

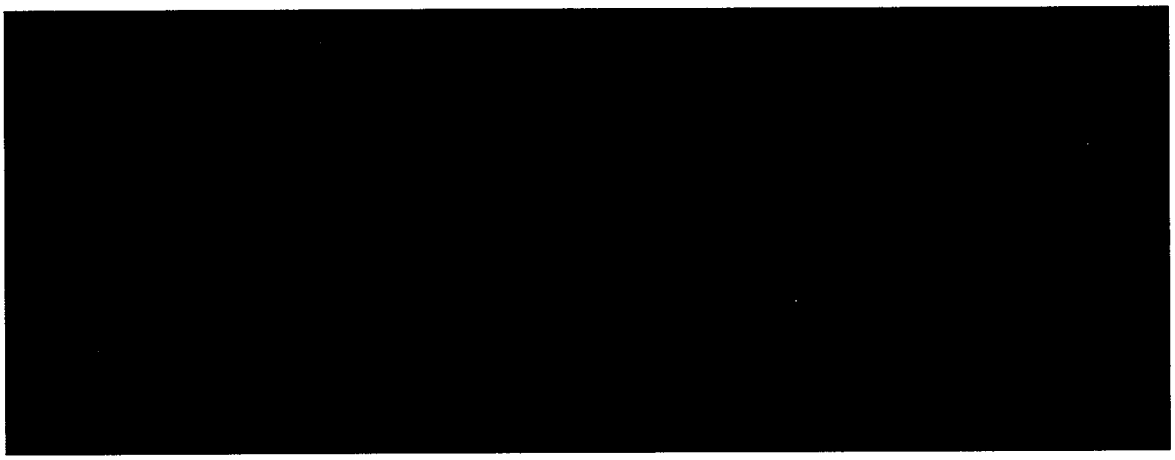
For all the efficacy endpoint that treatment effect is significant in the Section 7.9.1.1 Primary Efficacy Analysis, the paired comparison of the placebo vs each active dose will be performed using the same model.

Point estimates and 95% CIs for the difference between each active dose and placebo will be generated as well.

7.9.2.2 Dose Response Analysis with Contrasts as exploratory analysis

The contrasts analyses will be performed for evaluating the dose effect. For change from baseline in the average daily frequency of moderate to severe VMS, the analyses with the following contrasts will be conducted. (Table 14.2.4) The corresponding 4 arms are placebo, MT-8554 1 mg, 5 mg, and 10 mg.

C1=(-3,-1,1,3) : dose response is linear
C2=(-5,-1,3,3) : dose response is reaching plateau at medium dose
C3=(-3,1,1,1) : dose response is reaching plateau at low dose
C4=(-1,0,1,0) : dose response goes downturn at high dose
C5=(-3,-3,1,5) : dose response is ineffective at low dose



7.9.3 Analysis of Other Efficacy Endpoints

The PGIC score will be analyzed with ANOVA model at Weeks 4 and 12. (Table 14.2.8.1)

The response rate based on PGIC score (i.e. either 6 or 7) will be analyzed with logistic regression at Weeks 4 and 12. (Table 14.2.8.2)

The change from baseline to Weeks 4 and 12 in PSQI total and each domain score will be analyzed with ANCOVA model. The analysis model will include baseline value as a covariate and treatment as fixed effect. (Table 14.2.9.1) In addition, PSQI total score in categories will be summarized for each visit (Table 14.2.9.2). The following categories will be used for the analysis;

- Good sleep quality: PSQI total score ≤ 5

- Poor sleep quality: PSQI total score > 5

The ANCOVA model will be applied to analyze the change from baseline to Weeks 4 and 12 in total and each domain score of the MENQOL. The analysis model will include baseline value as a covariate and treatment as fixed effect. (Table 14.2.10)

The change from baseline to Weeks 4 and 12 in each domain i.e. Physical Functioning, Role Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role Emotional, Mental Health, Physical Component Summary [PCS], and Mental Component Summary [MCS]) score of the SF-36 will be analyzed using ANCOVA model with treatment as fixed effects and baseline value as covariate. (Table 14.2.11)

The ANCOVA model will be applied to analyze the change from baseline in the average daily severity score of moderate to severe VMS at Weeks 4 and 12, defined as 2xFmo + 3xFse. The analysis model will include baseline value as a covariate and treatment as fixed effect. (Table 14.2.12)

7.9.4 Sensitivity Analysis of Efficacy Endpoints

Sensitivity analyses for Efficacy Endpoints as following.

The sensitivity analysis will be conducted using primary analysis model with applicable covariates for ITT subjects without using antidepressants. (Table 14.2.2.1.A, Table 14.2.2.1.B. Table 14.2.2.2.A, Table 14.2.2.2.B)

The primary endpoints will also be analyzed using an ANCOVA model at week 4 and 12. The analysis model will include baseline value of the endpoint as a covariate and treatment as fixed effect. (Table 14.2.3.1, Table 14.2.3.2)

[redacted]

7.10 Safety Assessments

The safety evaluation will be purely descriptive using descriptive statistics (N, mean, SD, median minimum and maximum) or frequency tables where appropriate using SAF population, unless otherwise specified. No imputation will be made in case of missing values.

7.10.1 Adverse Events

The frequency and incidence of TEAEs will be summarized by System Organ Class (SOC) and Preferred Term (PT) by treatment and overall. The SOC is sorted by International order; then within SOC, PT is sorted by descending counts under MT-8554 Total column, then descending counts under placebo column, then alphabetic order for PTs with the same count.

Following summaries will be presented:

- Overall summary of TEAE (Table 14.3.1.1)

- TEAEs by SOC and PT (Table 14.3.1.2)
- TEAEs by SOC, PT and time of onset (Table 14.3.1.3)
- TEAEs by SOC, PT and severity (Table 14.3.1.4)
- TEAEs leading to discontinuation of study medication by SOC and PT (Table 14.3.1.5)
- Serious TEAEs by SOC and PT (Table 14.3.1.6)
- TEAEs by SOC, PT and drug relationship (Table 14.3.1.7)
- [redacted]
- Adverse reactions by SOC and PT (Table 14.3.1.9)
- Serious Treatment Adverse reactions by SOC, PT (Table 14.3.1.10)
- TEAEs by SOC and PT for AEs with frequency >= 3% (or 5%) in total MT-8554 group (Table 14.3.1.11)
- [redacted]

For each of the summaries will be done at the subject level - multiple occurrences of the same event within a subject will be counted once in the summaries by SOC and PT; multiple occurrences of the same event within a subject will be counted once in the maximum intensity category (severe > moderate > mild) and/or maximum drug relationship category (reasonable possibility, no reasonable possibility) and/or the earliest occurrence. If intensity or relationship is found to be missing the most severe occurrence will be imputed for that particular summary.

All AEs for each subject, including multiple occurrences of the same event, will be presented in full in a comprehensive listing including subject number, treatment, severity, seriousness, action taken, outcome, relationship to treatment, onset/stop date and duration.

Deaths that occur during the study will be listed in a data listing. The data listings for serious TEAE and TEAE leading to discontinuation will be generated as well.

[redacted]

7.10.2 Laboratory Tests

For all continuous laboratory parameters, the lab values and change from baseline to post baseline visits are summarized by treatment group and visit.

For all discrete laboratory parameters, the lab values are summarized with count and percentage by treatment group and visit. (Table 14.3.2.1, Table 14.3.2.2, Table 14.3.2.3, Table 14.3.2.4, Table 14.3.2.5)

The lab tables will be generated for each Lab categories (Hematology, Coagulation, Biochemistry and Hormones and Urinalysis).

All laboratory data will be listed with clinically relevant values flagged (L=Lower than lower limit of normal range or H=Higher than upper limit of normal range).

In addition, following clinical relevant ranges (and flags) will also be considered for summary table. The count and percent of subjects meeting the criteria will be presented. (Table 14.3.2.6)

- ALT ≥ 3 × Upper Limit of Normal Range (ULN), 5× ULN, 8× ULN, 10× ULN
- AST and/or ALT ≥ 3 × ULN, 5× ULN, 8× ULN, 10× ULN
- AST and/or ALT ≥ 3 × ULN with Total bilirubin ≥ 2 × ULN
- Total bilirubin ≥ 2 × ULN
- Creatinine >= 2 × ULN

7.10.3 Vital Signs

Vital signs data will be summarized descriptively in tables by treatment and scheduled visit. (Table 14.3.3.1) All vital sign data will be listed.

7.10.4 ECGs

ECG parameter values and changes from baseline will be summarized descriptively by treatment and scheduled visit. (Table 14.3.3.2.1)

For ECGs, number and percentage of subjects meeting the criteria listed below will be presented in tables (Table 14.3.3.2.2):

- QTc > 500ms
- QTc > 480ms
- QTc > 450ms
- Change from baseline in QTc > 30 msec
- Change from baseline in QTc > 60 msec

These criteria will be applied to both QTcB and QTcF.

Shift tables will present the changes in clinically relevant categories from baseline to EOT/Early discontinuation visit. The categories are; "QTcB<=450msec", "QTcB >450 to <=480msec", "QTcB >480 to <=500msec", "QTcB >500msec". (Table 14.3.3.2.3)

Similar shift table will be also created for QTcF. The categories are; "QTcF<=450msec", "QTcF >450 to <=480msec", "QTcF >480 to <=500msec", "QTcF >500msec". (Table 14.3.3.2.4)

All ECG parameters and findings will be listed.

7.10.5 Physical Examinations

Physical examination data will be summarized descriptively in tables by treatment and visit (Table 14.3.4).

All physical examination data will be listed.

7.10.6 Breast Examination

The breast examination data will be summarized in tables by treatment and visit (Table 14.3.5). All breast examination data will be listed.

7.10.7 Endometrial Safety

7.10.7.1 Endometrial Thickness Measured by Transvaginal Ultrasound (TVU)

The endometrial thickness as measured by TVU will be summarized in the tables by treatment and visit as below. (Table 14.3.6.1, Table 14.3.6.2)

1. Summary table presenting mean and SD for baseline, post baseline visits and change from baseline
2. Summary table of endometrial thickness by category (<5, 5 to 8, >8)
3. Summary table for category of change from baseline (>3 mm, >5 mm)

The central reading data will be used for the analysis.

7.10.7.2 Endometrial Histology (Endometrial Biopsy)

The Endometrial biopsy data will be summarized in the tables by treatment and visit as below (Table 14.3.7.1, Table 14.3.7.2).

1. Summary tables of proportion of subjects in abnormal* category
2. Shift table for change in abnormal category from baseline to week 12

*Note: The histological categories are based on 2003 draft FDA guidance [redacted]

The data from adjudication committee will be used for the analysis.

7.10.8 Other Safety Assessments

Total score of PHQ-8 and total score of GAD7 will be summarized descriptively in tables by treatment and scheduled visit. (Table 14.3.8, Table 14.3.9) The data listings will be generated.

7.11 Pharmacokinetics Assessment

These analyses will be performed on the PK population.

Changes to the procedures or events, which may impact the quality of the PK data will be considered significant protocol deviations or events. These changes will include any circumstances that will alter the evaluation of the PK. In the case of a significant protocol deviation or event, PK data collected during the affected visit will be flagged.

7.11.1 Listings, and Descriptive Statistics

Plasma MT-8554 data will be listed for each subject and scheduled visit and treatment period with the same precision as provided by the bioanalytical laboratory. PK blood sample collection times, most recent dosing times, as well as derived actual sampling time relative to the most recent dose will be provided in a listing. The actual sampling time relative to the most recent dose will be calculated in hours and rounded to 2 DP.

7.12 Statistical/Analytical issues

7.12.1 Adjustments of covariates

In the efficacy analysis, repeated measures ANCOVA model with an unstructured covariance structure or ANCOVA will be performed with baseline as a covariate in order to consider the impact of the baseline on the change in each variable. More details were described in each analysis.

7.12.2 Data Handling of Dropouts or Missing

Missing data will not be imputed in the summary at each time point. The missing data will not be imputed for efficacy analyses. For AE start and/or end date missing or partial missing, the AE will be treated as TEAE if it cannot be determined a non-TEAE.

7.12.3 Interim Analyses and Data Monitoring

The details are described in Section 3.2 Interim Analysis.

7.12.4 Multiple Comparison/Multiplicity

In order to deal with multi-endpoint multiplicity, the efficacy analyses of primary and secondary endpoints will be conducted in the order fixed method. The details are described in Section 7.9 Efficacy Assessments.

The secondary analysis of primary variable and the analysis of other efficacy endpoints are performed from the exploratory perspectives and therefore no adjustment will be made for the multiplicity among variables and time points.

7.12.5 Subgroup Analysis

For Co-Primary endpoints will be summarized by treatment group for the following subgroup, as subgroup analyses. (Table 14.2.2.3.1, Table 14.2.2.3.2)

- Age group (age < 65, age ≥ 65)
- Race (White, Black or African American, Asian, others)
- BMI (BMI <30, BMI ≥ 30 kg/m^2)
- Menopause status (Natural, surgical)
- Average daily VMS frequency (<Medn, ≥Medn), where Medn = Median of VMS frequency of ITT at baseline

8. CHANGES FROM THE PROTOCOL

Subgroup analysis of antidepressant using (Y, N) will not be performed because there are not enough subjects with using antidepressant for the analysis.

Analysis for PP Population does not apply to secondary endpoints. These analyzes were considered to be excessive.

9. VALIDATIONS

The tables, figures and listings (TFLs) planned in this SAP will be produced by [redacted] using SAS software version 9.4 (or above). The TFLs will be quality checked by statistics team in MTDA and [redacted] using SAS software version 9.3 (or above). The TFLs will be approved by MTDA.

[redacted]

10. PROGRAMMING AND DATA PRESENTATION CONVENTIONS

Listings will be presented in treatment, subject, visit (where applicable) and date (where applicable) order. Listings will be produced (landscape in MS Word) using PROC REPORT in SAS monospace font and pitch 8.

Summary tabulations will be presented by treatment group (and overall if appropriate), scheduled visit order (if appropriate). Continuous data summaries will present (unless stated otherwise) number of observations, mean, standard deviation, median, minimum and maximum. Categorical data summaries will present the number of observations and the corresponding percentage.

13. BIBLIOGRAPHY

1. Youden WJ. Index for rating diagnostic tests. Cancer. 3: 32–35. 1950.