STATISTICAL ANALYSIS PLAN (52-Week)

(analysis for fezolinetant exposure up to 52 week)

Version 2.0, dated 27APR2021

A Phase 3, Randomized, Placebo-controlled, 12-week Double-blind Study, followed by a Non-Controlled Extension Treatment Period, to Assess the Efficacy and Safety of Fezolinetant in Women Suffering from Moderate to Severe Vasomotor Symptoms (Hot Flashes) Associated with Menopause

Skylight 2

ISN: 2693-CL-0302

IND number: 130277

EudraCT 2018-003529-27

NCT04003142

Sponsor: **Astellas Pharma Global Development, Inc. (APGD)**

1 Astellas Way

Northbrook, IL 60062

Table of Contents

I. LIST OF ABBREVIATIONS AND KEY TERMS

List of Abbreviations

Abbreviations	Description of abbreviations
ANCOVA	Analysis of Covariance
ASCM	Analysis Set Classification Meeting
CI	Confidence Intervals
CMH	Cochran Mantel Haenszel
CSR	Clinical Study Report
ED	Early Discontinuation
EOT	End of Treatment
EQ-5D-5L	European Quality of Life 5 Dimensions-5 Levels
FAS	Full Analysis Set
FAS_F	Full Analysis Set for Fezolinetant
FD	Fezolinetant Day (study day relative to first dose of Fezolinetant)
FUD	Follow-up Day (study day relative to last dose date of study treatment)
ICH	International Conference on Harmonization
IRT	Interactive Response Technology
MAR	Missing at Random
MENQOL	Menopause-specific Quality of Life Questionnaire
MMRM	Mixed Model Repeated Measures
PDAS	Pharmacodynamics Analysis Set
PDAS_F	Pharmacodynamics Analysis Set for Fezolinetant
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PKAS	Pharmacokinetics Analysis Set
PRO	Patient-Reported Outcome
PROMIS	Patient-Reported Outcomes Measurement Information System
SAF	Safety Analysis Set
SAF_F	Safety Analysis Set for Fezolinetant
SAP	Statistical Analysis Plan
SD	Standard Deviation
SD	PROMIS Sleep Disturbance
SRI	PROMIS Sleep Related Impairment
TLF	Tables, Listings and Figures
VMS	Vasomotor Symptoms
WPAI	Work Productivity and Activity Impairment

List of Key Terms

Terms	Definition of terms
Double-blind Treatment Period	The 12 weeks of treatment after the first administration of study drug and prior to first administration of study drug in the extension period. Only this period contains a placebo treatment.
Extension Treatment Period	Period of time after the first administration of study drug in the extension period after completing 12 weeks of placebo-controlled double-blind treatment until the last dose of the study drug (at Week-52 or early discontinuation).
Follow-up Period	Period of time after the last dose of study drug or last assessment of the protocol. Additional safety data and follow-up observations for sustained adverse events are conducted during this period.

1 INTRODUCTION

This Statistical Analysis Plan (SAP) contains technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes procedures for executing the statistical analysis to fulfil the objectives of the study.

This SAP describes the plans for data summary of the 52-week study including the 40-week extension treatment period after completing 12-week double-blind placebo-controlled treatment period. These analyses may be conducted prior to all subjects completing the full 52-week study, if needed. A separate SAP describes the analysis of the 12-week double-blind study data period.

The SAP v1.0 was approved prior to the database lock for the 12-week analysis. This updated SAP will be approved prior to the final study database lock. The SAP will be finalized before unblinding the subject treatment assignments.

Changes from the planned analyses in the finalized SAP that impact the statistical analyses will be documented and justified in the Clinical Study Report (CSR).

2 STUDY OBJECTIVE(S) AND DESIGN

2.1 Study Objective(s)

Note that the analyses for many of these objectives have been described in the SAP for the double-blind placebo-controlled study period for 12-week analysis. All protocol objectives are listed below but the specific endpoints in this SAP have been restricted to those extending beyond the double-blind treatment period. The objectives of the study beyond the double-blind treatment period are to describe efficacy and safety for subjects exposed to fezolinetant for up to 52 weeks.

Primary objective:

- To evaluate the efficacy of fezolinetant versus placebo on the frequency and severity of moderate to severe vasomotor symptoms (VMS).
  - The estimand of the primary objective will use a hypothetical strategy and compare patients as though they had continued on the assigned treatment.

Key secondary objective:

- To evaluate the efficacy of fezolinetant versus placebo on patient-reported sleep disturbance.

Secondary objectives:

- To evaluate the effect of fezolinetant versus placebo on the frequency and severity of moderate to severe VMS at weekly time points.
- To evaluate the safety and tolerability of fezolinetant.

Exploratory objectives:

- To evaluate population pharmacokinetics (popPK) of fezolinetant and metabolite ES259564.
- To evaluate the effect of fezolinetant on pharmacodynamic markers.
- To evaluate the efficacy of fezolinetant versus placebo on the frequency and severity of mild, moderate and severe VMS.
- To evaluate the short-term and sustained effects of fezolinetant versus placebo on patient-reported sleep disturbance.
- To evaluate the effect of fezolinetant versus placebo on the following patient-reported outcomes (PROs): global assessments of VMS and sleep disturbance, overall sleep-wake function, quality of life and work productivity.

2.2 Study Design

This is a randomized, 12-week double-blind, placebo-controlled, parallel group, multicenter clinical study to assess the efficacy and safety of fezolinetant in women suffering from moderate to severe VMS associated with menopause. Approximately 450 subjects will be enrolled into this study, 150 subjects per treatment arm. Duration of treatment is 52 weeks. After completing 12 weeks of treatment, subjects on placebo will be reassigned to 30 mg or 45 mg of active treatment in an extension treatment period through end of study. Subjects who were already randomized on an active arm will continue on their assigned dose for the remaining 40 weeks of treatment. Following the completion (or early discontinuation [ED]) of the treatment period (week 52), subjects will complete an end of treatment (EOT; or ED) visit and final safety follow-up visit 3 weeks after the last dose of study drug is administered (week 55).

Details of the schedule of clinical assessments are available in the protocol.

2.3 Randomization

The first 12 weeks of the study are double blind. Subjects will be randomized to receive fezolinetant or placebo in a blinded fashion such that neither the investigator, sponsor's study management team, clinical staff, nor the subject will know which agent is being administered. The randomization number will be assigned based on information obtained from the Interactive Response Technology (IRT).

Subjects will be randomized in a 1:1:1 ratio to a treatment arm according to the randomization schedules and stratified by smoking status (current smoker vs former/never) through IRT. Subjects who complete the 12-week treatment period and have received placebo will be re-randomized to either fezolinetant 30 mg or 45 mg for the extension period. The assigned treatment in the extension period will remain blinded to the site personnel and participating subjects. The site personnel will dispense the treatment according to the IRT system's assignment. Specific procedures for randomization through the IRT are contained in the IRT specifications manual.

3 SAMPLE SIZE

A total of 450 subjects are planned to be randomized; 150 subjects in each of the three treatment arms. The subjects for the analyses covered in this SAP are determined by the number of subjects who receive at least one dose of fezolinetant during the study. Subjects originally randomized to placebo will only be included if they complete the double-blind treatment phase, are re-randomized, and receive at least one dose of fezolinetant during the study, unless otherwise specified.

The sample size of 450 subjects was driven by the double-blind comparisons to placebo at week 4 and week 12. See the 12-week analysis SAP for sample size details.

4 ANALYSIS SETS

In accordance with International Conference on Harmonization (ICH) recommendations in guidelines E3 and E9, the following analysis sets will be used for the analyses.

4.1 Full Analysis Set

The full analysis set for the 52-week analysis will consist of all subjects who are randomized and receive at least 1 dose of fezolinetant during either the double-blind or the extension period (FAS_F). This will be the primary analysis set for descriptive efficacy analyses. The randomized treatment for each subject will be used for summaries by treatment group based on the FAS_F, even if a subject erroneously received a different treatment.

4.2 Safety Analysis Set

The safety analysis set for the 52-week analysis will consist of all subjects who are randomized and receive at least 1 dose of fezolinetant during either the double-blind or the extension period (SAF_F), and will be used for safety analyses. A subject erroneously receiving a treatment different from their randomized treatment will be assigned to the treatment group that the subject received as first dose.

The SAF_F will be used for summaries of demographic and baseline characteristics and all safety and tolerability related variables.

4.3 Per Protocol Set

Per protocol analysis sets are not applicable beyond the double-blind treatment period.

4.4 Pharmacokinetics Analysis Set

The PKAS will be defined in a separate analysis plan. Results of the population PK analysis will not be reported in the Clinical Study Report but in a separate population PK report.

4.5 Pharmacodynamic Analysis Set

The pharmacodynamic analysis set for the extension (PDAS_F) will include the subjects from the administered population (received at least 1 dose of fezolinetant during either the double-blind or the extension period) for whom sufficient pharmacodynamic measurements

were collected during the study. The PDAS_F will be used for all analyses of pharmacodynamic data (e.g., sex hormones).

5 ENDPOINTS

The calculation of efficacy baseline for VMS frequency and severity is described in the 12-week analysis SAP, and will be calculated based on the number of moderate to severe hot flashes in the 10 days immediately prior to randomization.

For general efficacy endpoints collected at visits, the last non-missing assessment on or prior to the first dose of study treatment (investigational product or placebo) is the baseline.

For all general safety endpoints, in contrast, the baseline value will be the last non-missing value taken on or prior to first dose of fezolinetant.

Note that the baseline calculations differ between efficacy and safety analyses. The efficacy endpoints are subjective and there is expected to be a non-ignorable placebo effect, so the entire double-blind portion of the study is considered post-baseline for all subjects. For safety endpoints, any changes while on placebo during the double-blind portion of the study are expected to be within-subject variation and changes from baseline will be assessed beginning with exposure to fezolinetant.

The table below provides additional details regarding the baselines for specific endpoints. The terms R_{BL} and $WK12_{BL}$ indicate where the baseline is established, where R_{BL} is initial randomization and $WK12_{BL}$ is the start of the extension period at Week 12.

Endpoint	**Baseline**	
	fezolinetant / fezolinetant	**placebo / fezolinetant**
General efficacy	R_{BL}	R_{BL}
General safety and TVU	R_{BL}	$WK12_{BL}$
Hormones	R_{BL}	$WK12_{BL}$
Demographics	R_{BL}	R_{BL}
Medical history	R_{BL}	R_{BL}
Prior medications	R_{BL}	$WK12_{BL}$
Concomitant medications	R_{BL}	$WK12_{BL}$
Study drug exposure and compliance	R_{BL}	$WK12_{BL}$
TEAE	R_{BL}	$WK12_{BL}$
Clinical lab summary shifts and liver safety assessments	R_{BL}	$WK12_{BL}$
Protocol deviations	R_{BL}	R_{BL}
Biopsy (Listing only)	R_{BL}	R_{BL}

The 12-week double-blind placebo period will be completely described as part of the treatment comparisons within the 12-week double-blind analyses. As a result, all endpoints in the 52-week analysis are from the first dose of fezolinetant and onward (see Section 6.10 for additional details).

The ePRO daily diary, ePRO visit-based questionnaires, and CRF-based endpoints are all described in the 12-week analysis SAP.

Only analyses pertaining to endpoints that extend to time points beyond the 12-week double-blind period will be described in this SAP for the 52-week study.

The analysis visits for each endpoint and their corresponding definitions of calculation windows are in Section 6.10.1.1 (Efficacy Analysis Windows).

5.1 Primary Efficacy Endpoints

There is no primary efficacy endpoint beyond the double-blind treatment period. However, some VMS definitions from the 12-week analysis SAP are repeated below as the subjects continue to record their VMS symptoms in their electronic diary during the extension treatment period.

Subjects will record the number of vasomotor symptoms and severity of each vasomotor symptom via the ePRO daily diary.

The severity of an individual VMS is defined as follows [FDA Draft Guidance for Industry, 2003 and EMA CHMP Guideline, 2005]:

- Mild: sensation of heat without sweating
- Moderate: sensation of heat with sweating, able to continue activity
- Severe: sensation of heat with sweating, causing cessation of activity

For both post-baseline frequency and severity, a daily average per week will be derived if any information for 4 or more days was reported.

5.2 Secondary Efficacy Endpoints

5.2.1 Key Secondary Endpoint

There is no key secondary efficacy endpoint beyond the double-blind treatment period.

5.2.2 Secondary Endpoints

The secondary efficacy objectives that extend beyond the double-blind treatment period and examine the effect of fezolinetant over time are the following:

- Mean change in the frequency of moderate to severe VMS from baseline to week 24 (descriptive)
- Mean change in the Severity of moderate to severe VMS from baseline to week 24 (descriptive)

These will be described as a time point of particular interest within the exploratory VMS endpoint below.

5.3 Exploratory Efficacy Endpoints

The exploratory efficacy objectives that extend beyond the double-blind treatment period and examine the effect of fezolinetant over time are the following:

- Mean change in the frequency and severity of moderate and severe VMS from baseline to each analysis visit in the extension period and the follow-up visit
    - Earlier time points will also be presented for completeness

- Change in serum concentrations of sex hormones and sex hormone-binding globulin (SHBG) from baseline to each analysis visit
- Mean change in serum concentrations of bone specific alkaline phosphatase (BSAP), procollagen type 1 amino-terminal propeptide (P1NP) and carboxy-terminal telopeptide of type I collagen (CTX) from baseline to each analysis visit
- Plasma concentrations of fezolinetant and metabolite ES259564 at pre-specified time points
- Score on the Patient Global Impression of Change (PGI-C) in VMS at each visit
- Mean change on the PROMIS SD SF 8b total score from baseline to each visit
- Mean change on the Patient-Reported Outcomes Measurement Information System Sleep-Related Impairment – Short Form 8a (PROMIS SRI SF 8a) total score from baseline to each visit
- Score on the Patient Global Impression of Severity in Sleep Disturbance (PGI-S SD) at each visit
- Score on the Patient Global Impression of Change in Sleep Disturbance (PGI-C SD) to each visit
- Mean change on the Menopause-Specific Quality of Life (MENQOL) total score from baseline to each visit
- Mean change on the MENQOL domain scores from baseline to each visit
- Mean change on the Euro-Qol 5D-5L (EQ-5D-5L) Visual Analog Scale (VAS) from baseline to each visit
- Mean change on the Work Productivity and Activity Impairment questionnaire specific to VMS (WPAI-VMS) domain scores from baseline to each visit.

5.4 Safety Endpoints

Safety evaluations will include the following endpoints, but are not limited to:

- Frequency and severity of AEs
- TVUs and endometrial biopsy findings
- Change from baseline to each timepoint in vital signs: sitting systolic and diastolic blood pressure and pulse rate
- Change from baseline to each timepoint in ECG parameters
- Change from baseline to each timepoint in laboratory tests: hematology, biochemistry and urinalysis

6 STATISTICAL METHODOLOGY

6.1 General Considerations

As a result of this study design and the re-randomization of subjects originally assigned to placebo, there are four treatment groups which will be described in the 52-week analysis:

- Subjects originally randomized to 30 mg: 30 mg => 30 mg; per protocol, these subjects will have up to 12 months of fezolinetant exposure.

- Subjects randomized to placebo who will be re-randomized to 30 mg: placebo => 30 mg; per protocol, these subjects will have up to 9 months of fezolinetant exposure.
- Subjects originally randomized to 45 mg: 45 mg => 45 mg; per protocol, these subjects will have up to 12 months of fezolinetant exposure.
- Subjects randomized to placebo who will be re-randomized to 45 mg: placebo => 45 mg; per protocol, these subjects will have up to 9 months of fezolinetant exposure.

Continuous data will be summarized descriptively including the number of subjects (n), mean, standard deviation (SD), median, minimum and maximum. Categorical data will be summarized by frequencies and percentages. Percentages by categories will be based on the number of subjects with no missing data, i.e., the percentages for the non-missing categories will add up to 100%.

All summaries are descriptive and there are no statistical comparisons within the 52-week analyses.

All data summarization and analyses will be performed using SAS® on Unix. Specifications for table, figures, and data listing formats can be found in the Tables Listings and Figures (TLF) specifications.

As mentioned in the introduction, this analysis plan describes the analyses of the entire 52-week study for subjects who received fezolinetant, excluding the 12-week double blind analyses which is planned in a separate SAP.

Listings will include baseline data and data from first dose of fezolinetant, unless otherwise noted.

6.2 Study Population

In general, data such as patient disposition, demographics and baseline characteristics will be summarized for FAS_F population as the definition of FAS_F is identical to SAF_F if no patients were given wrong study drug. In the event when FAS_F is not identical to SAF_F then data will be summarized for both FAS_F and SAF_F as specified in the relevant sections.

6.2.1 Disposition of Subjects

Disposition of subjects will be summarized for the FAS_F by treatment group and overall. For the discontinuations, the primary reason reported by the investigator will be summarized.

Number and percentage of subjects for each analysis set will be summarized by treatment group and overall.

6.2.2 Protocol Deviations

The number and percentage of subjects with the following protocol deviation criteria will be summarized for the complete SAF population for each criterion and overall, by treatment group at the time the deviation occurs and total as well as by investigative site. Subjects deviating from a criterion more than once will be counted once for the corresponding criterion within a treatment group. The summary will include a column for PD during the

12-week placebo period. Placebo subjects deviating from a criterion while on placebo and again while on active treatment will be counted once in each treatment (e.g., placebo and fezolinetant) for the corresponding criterion, and the subject will be described in a footnote. However, such a subject would only be counted once in the Total column.

The listing will be based on the complete SAF population including all placebo data and subjects randomized to placebo who never received fezolinetant.

The unique identifiers will be as follows:

PD1 - Entered into the study even though they did not satisfy entry criteria,

PD2 - Developed withdrawal criteria during the study and was not withdrawn,

PD3 - Received wrong treatment or incorrect dose,

PD4 - Received excluded concomitant treatment.

The number of subjects who were affected by the COVID-19 pandemic will be summarized by impact item and treatment group and overall for SAF population. The impact items include treatment discontinuation, dosing change/interruption, death from COVID-19, healthcare encounter due to COVID-19, protocol deviation including missed visit windows due to COVID-19, and impacted diary data due to COVID-19.

An overview listing of subjects who were affected by the COVID-19 pandemic will be generated by the impact items.

Detailed information on how these assessments are affected by the COVID-19 pandemic can be found in relevant data listings with a flag for COVID-19.

6.2.3 Demographic and Other Baseline Characteristics

Demographics, tobacco history (smoking status), substance use (e-cigarette, alcohol, cannabis), caffeine beverage intake, height and weight, and targeted medical history will be summarized descriptively by treatment group and total group for FAS_F and SAF_F. Age group categories include <40 years, ≥40 years to <46 years, ≥46 years to <51 years, ≥51 years to <56 years, ≥56 years to <61 years, ≥61 years to <66 years, and ≥66 years. Note that current versus former or never smoking status is a stratification factor for randomization. Hormone replacement therapy history will be summarized descriptively by treatment group and total group for SAF_F.

Medical history is coded in MedDRA, and will be summarized by System Organ Class (SOC) and Preferred Term (PT) by treatment group and total group for the SAF_F.

These parameters will be summarized using the data collected at baseline.

6.2.4 Previous and Concomitant Medications

Previous and concomitant medications will be summarized by therapeutic subgroup (ATC 2nd level) and chemical subgroup (ATC 4th level) and preferred WHO name by treatment group for the FAS_F and SAF_F. Subjects taking the same medication multiple times will be counted once per medication and investigational period.

Previous medications are defined as medications that patients started prior to first administration of fezolinetant. Concomitant medications are defined as any medications that patients took after the first dose of fezolinetant and through 30 days from last dose of study drug. Medications that started prior to first administration of fezolinetant and continued while study drug was given will be counted in both previous and concomitant medications.

6.3 Study Drug Exposure and Compliance

Duration and compliance of study drug will be summarized for SAF_F by treatment group and overall.

Number and percentages of subjects with the following categories of study drug duration will be summarized: >0 to ≤ 1 day, >1 to ≤ 7 days, >7 to ≤ 14 days, >14 to ≤ 21 days, >21 to ≤ 28 days, >28 to ≤ 42 days, >42 to ≤ 56 days, >56 to ≤ 84 days, >84 to ≤ 112 days, >112 to ≤ 140 days, >140 to ≤ 168 days, >168 to ≤ 196 days, >196 to ≤ 224 days, >224 to ≤ 252 days, >252 to ≤ 280 days, >280 to ≤ 308 days, >308 to ≤ 336 days, >336 to ≤ 364 days, and >364 days.

Number and percentages of subjects with the following cumulative categories of study drug duration will be summarized: ≥ 1 day, ≥ 7 days, ≥ 14 days, ≥ 21 days, ≥ 28 days, ≥ 42 days, ≥ 56 days, ≥ 84 days, ≥ 112 days, ≥ 140 days, ≥ 168 days, ≥ 196 days, ≥ 224 days, ≥ 252 days, ≥ 280 days, ≥ 308 days, ≥ 336 days, and ≥ 364 days.

Overall treatment compliance with the dosing schedule will be examined for subjects in the SAF_F whose total study drug count and first and last days of treatment are known. Compliance will be calculated compared to the actual treatment period of dosing (first to last day of fezolinetant), not to the planned treatment period.

Percent overall compliance while on fezolinetant will be summarized in two ways for the SAF_F:

- Descriptive statistics will be presented by treatment group.
- Percent compliance will be categorized according to the following categories by treatment group:
    - Less than 50%,
    - at least 50%, less than or equal to 85%,
    - greater than 85%, less than or equal to 120%,
    - Over 120%, and
    - Unknown.

To explore the impact of the COVID-19 pandemic, additional analysis will be performed by excluding the days of missing study drug/drug interruption due to COVID-19 pandemic in the duration and compliance calculation, according to the study drug interruption CRF page.

6.4 Analysis of Efficacy

All efficacy endpoints will be descriptively summarized by the four treatment groups, as described in Section 6.1 (General Considerations), for FAS_F, unless specified otherwise.

For efficacy analyses, the visit windows from Section 6.10.1.1 and 6.10.1.2 will be applied, and they are all defined relative to the start of fezolinetant.

6.4.1 Analysis of Secondary Efficacy Endpoints

The mean change in the frequency and severity of moderate and severe VMS from baseline to week 24 will be summarized descriptively, as part of the output specified below and in Section 6.10.1.1.

The analysis will be repeated by excluding diary data that are impacted by COVID-19 pandemic, which includes the diary data recorded after the subjects were diagnosed of COVID-19, after study drug discontinuation due to COVID-19, and during dose interruption due to COVID-19.

6.4.2 Analysis of Exploratory Endpoints

The mean change in the frequency and severity of moderate and severe VMS from baseline to each analysis week up to week 12, and specified weeks thereafter (as specified in Section 6.10.1.1) will be summarized, including the secondary endpoint of Week 24.

For non-diary efficacy endpoints analysis visits are used rather than weeks, as described in Section 6.10.1.1.

PROMIS (SD and SRI total scores), MENQOL (domain and total scores), and WPAI-VMS domain scores will be summarized descriptively.

PGI-S and PGI-C (both VMS and SD) and EQ-5D domain scores will be summarized descriptively. The mean and standard deviation of the EQ-5D domain scores will be presented.

Change on the Euro-Qol 5D-5L (EQ-5D-5L) Visual Analog Scale (VAS) from baseline will be summarized descriptively.

The median change in serum concentrations of sex hormones and sex hormone-binding globulin (SHBG) from baseline to each analysis visit and the mean change in serum concentrations of BSAP, P1NP and CTX from baseline to each analysis visit will be summarized.

6.5 Analysis of Safety

All analysis of safety will be presented by the four treatment groups, as described in Section 6.1 (General Considerations), for SAF_F, unless specified otherwise. For analyses by visit, the visit windows from Section 6.10.1.2 will be applied, and they are all defined relative to the start of fezolinetant. Except for adverse events and follow-up visit data, only data up to one day after the last dose date will be considered for the analyses.

The baseline value for safety endpoint will be the last non-missing value taken on or prior to first dose of fezolinetant unless otherwise noted.

6.5.1 Adverse Events

TEAE is defined as an AE observed after starting administration of fezolinetant and up to 21 days after the last dose of study drug.

A study drug-related TEAE is defined as any TEAE with a causal relationship of YES by the investigator.

The sponsor has a list of events that they classify as "always serious" events. If an AE is reported that is considered by the sponsor to be an SAE per this classification as "always serious", the AE will be treated as an SAE in tables and listings, and flagged as 'always serious' in AE listings.

An overview table will include the following

- Number of TEAEs,
- Number and percentage of subjects with TEAEs,
- Number of drug related TEAEs,
- Number and percentage of subjects with drug related TEAEs,
- Number of serious TEAEs,
- Number and percentage of subjects with serious TEAEs,
- Number of serious drug related TEAEs,
- Number and percentage of subjects with serious drug related TEAEs,
- Number of TEAEs leading to permanent discontinuation of study drug,
- Number and percentage of subjects with TEAEs leading to permanent discontinuation of study drug, and
- Number of deaths.

The number and percentage of subjects with TEAEs in the following AE categories will be summarized for each treatment group by system organ class (SOC), high level term (HLT) and preferred term (PT):

- TEAEs
- drug related TEAEs.

The number and percentage of subjects with TEAEs in the following AE categories will be summarized by SOC and PT:

- serious TEAEs,
- drug related serious TEAEs,
- TEAEs leading to permanent discontinuation of study drug,
- drug related TEAEs leading to permanent discontinuation of study drug,
- TEAEs that equal to or exceed a threshold of 5.0% in any treatment group (threshold is based on PT).
- TEAEs excluding serious adverse events that equal to or exceed a threshold of 5.0% in any treatment group.

The number and percentage of subjects with TEAEs, as classified by PT only, will be summarized for each treatment group.

The number of TEAEs and the number and percentage of subjects with TEAEs, as classified by SOC and PT will be summarized by severity and by relationship to study drug. In the subject count, if a subject has multiple TEAEs with the same SOC or PT, but with differing severity or relationship, then the subject will be counted once with the worst severity and highest degree of relationship. If severity or relationship is missing for all episodes of the event, the subject will be counted under missing severity or relationship. In the adverse event count, the adverse events will be presented in each category they were classified to. Drug related TEAEs will be presented in a similar way by severity only.

Following adverse events of special interest are defined:

- Adverse event of uterine bleeding
- Endometrial hyperplasia/cancer or disordered proliferative endometrium
- Adverse event of thrombocytopenia
- Adverse event of liver test elevations
- Adverse event of bone fractures
- Adverse events of abuse liability
- Adverse events of depression
- Adverse events of wakefulness
- Adverse events of effect on memory

Search terms defining the AESIs will be pre-specified and documented.

The number and percentage of subjects with AESIs, as classified by PT will be separately summarized for each treatment group and listings of these adverse events will be provided.

The number and percentage of subjects with TEAEs associated with COVID-19, as classified by SOC and PT, will be summarized for each treatment group and overall.

6.5.2 Clinical Laboratory Evaluation

Quantitative values evaluated by the central laboratory including hematology, biochemistry, and urinalysis will be summarized using mean, standard deviation, minimum, maximum and median by treatment group at each analysis visit. Additionally, a within-subject change will be calculated as the post-baseline measurement minus the baseline measurement and summarized in the same way. Each laboratory result will be classified as low (L), normal (N), or high (H) at each visit according to the laboratory supplied reference ranges.

The number and percentage of subjects below and above reference range will be summarized for each treatment group at each visit.

Number and percentage of subjects with platelets < 150000/uL (150,000/μL (150×10^9/L) will also be summarized.

Frequency tabulations of qualitative clinical laboratory variables (urinalysis) will be presented for each treatment group at each visit.

For hematology and biochemistry shift tables will be presented for each treatment group:

- Summary shifts of reference range changes from baseline to worst finding during the treatment period (shift from normal or high to low, shift from normal or low to high, categorized increase [shift from low to normal or from normal to high], categorized no change [value stays in the same reference range], categorized decrease [shift from high to normal or from normal to low]).

For hormone related parameters, if the value is below the low limit of quantification, the value will be imputed as half of the LLOQ. In addition, the percentage of subjects who have BLOQ values at baseline who then have measurable post-baseline values will be summarized.

6.5.2.1 Liver Safety Assessment

The liver safety assessments will be summarized by the categories below based on the measurements from Alkaline Phosphatase (ALP), Alanine Transaminase (ALT), total bilirubin, Aspartate Transaminase (AST) and their combination. These parameters will be based on measurements from a central laboratory.

The subject's highest value during the entire active-dose study period (from day of first dose of fezolinetant until last dose of extension treatment period + 1 day) will be used.

- ALT > 3xULN, > 5xULN, > 8xULN, > 10xULN, > 20xULN
- AST > 3xULN, > 5xULN, > 8xULN, > 10xULN, > 20xULN
- ALT or AST > 3xULN, > 5xULN, > 8xULN, > 10xULN, > 20xULN
- ALP > 1.5xULN
- Total Bilirubin > 2xULN
- (ALT or AST > 3xULN) and Total Bilirubin > 2xULN
- (ALT or AST > 3xULN) and ALP < 2xULN and Total Bilirubin > 2xULN

The last 2 criteria where 2 or more parameters are evaluated will be with the measurements on the same day or up to 1 day apart. The denominator for each criterion will be the number of subjects who have at least one value during the active-dose study period. The number and percentage of subjects meeting the criteria during the active-dose study period will be summarized by treatment group.

Due to the COVID-19 pandemic, subjects might not have been able to visit the clinical site and have clinical laboratory tests taken at the clinical site and evaluated by the central laboratory, in this case clinical laboratory tests were performed taken by a local laboratory. The liver biochemistry safety assessments will be summarized including central and locally collected laboratory data due to the COVID-19 pandemic. This analysis is considered the primary analysis for the liver safety assessment.

As a secondary analysis for liver biochemistry, the analysis will be repeated with the data evaluated by the central laboratory but excluding the data that are impacted by COVID-19 pandemic, which include the data recorded after the subjects were diagnosed of COVID-19, after study drug discontinuation due to COVID-19, during dose interruption due to COVID-19, and visits that are out of protocol windows or data collected via home visits due to COVID-19 recorded in the eCRF.

6.5.3 Vital Signs

Vital signs (systolic blood pressure, diastolic blood pressure, and pulse rate) will be summarized using mean, standard deviation, minimum, maximum and median by treatment group and analysis visit. Additionally, a within-subject change will be calculated per analysis visit as the post-baseline measurement minus the baseline measurement and summarized by treatment group and visit.

6.5.4 Electrocardiograms

ECG variables will be summarized using mean, standard deviation, minimum, maximum and median for each treatment group at each analysis visit, including changes from baseline.

Number and percent of subjects with normal and abnormal results as assessed by the central reader for the 12 lead ECG will be tabulated by treatment group at each analysis visit.

The QTc interval will be summarized using frequency tables for each analysis visit for values of clinical importance using the range criteria below.

	QTc Interval Criteria Value (msec)
Normal	≤ 450
Borderline	> 450
Prolonged	> 480
Clinically significant	> 500

Number and percent of subjects with 12 lead ECG abnormalities as well as number and percent of subjects whose 12 lead ECG reading changed from normal at baseline to abnormal will be tabulated by treatment group at each analysis visit.

6.5.5 Other Safety-Related Assessments

C-SSRS results will be summarized by treatment group and analysis visit based on SAF_F.

Ultrasound results (endometrial thickness) and bone marker results will be summarized for the SAF_F population using mean, standard deviation, minimum, maximum and median by treatment group and visit. The within-subject change will be calculated per visit as the post-baseline measurement minus the baseline measurement and summarized by treatment group and visit. The listing will be based on the complete SAF population, including all placebo data and subjects randomized to placebo who never received fezolinetant, because these data were not included in the double-blind 12-week analysis.

Baseline biopsy will be defined as the non-missing value on or prior to first dose of study drug. Biopsy results will be listed based on the complete SAF population, including all placebo data and subjects randomized to placebo who never received fezolinetant, because biopsy data was not included in the double-blind 12-week analysis. Post-baseline biopsy results for subjects originally randomized to placebo will still be listed.

6.6 Analysis of Pharmacokinetics

The statistical methods for PK data will be described in a separate analysis plan. Results of the population PK analysis will not be reported in the Clinical Study Report but in a separate population PK report.

Plasma concentration of fezolinetant and ES259564 will be listed.

6.6.1 Estimation of Pharmacokinetic Parameters

All details of the population PK analysis will be described in a separate analysis plan.

6.7 Analysis of Pharmacodynamics

Serum concentrations of sex hormones (LH, FSH, E2, Testosterone Total/Free, Androstenedione, DHEA, Estrone) and SHBG will be summarized using descriptive statistics (number of subjects, mean, SD, median, minimum and maximum) for each analysis and for change from the baseline to each postdose visit, if the data is available.

6.8 Other Analyses

Not applicable.

6.9 Interim Analysis (and Early Discontinuation of the Clinical Study)

There will be no interim analysis.

6.10 Additional Conventions

In case of data issues or findings pertaining to the 12-week period after the 12-week lock and before the final lock, an errata form will be issued to document those data discrepancies.

6.10.1 Analysis Windows

Study Day for analysis windows during the double-blind treatment phase will be calculated as date of visit/assessment – first dose date +1.

Study Day during the extension phase will be calculated as date of visit/assessment – first dose date of the extension phase +1.

Study Day for the follow-up (FUD) phase will be calculated as date of visit/assessment – last dose date of study treatment (Date of EOT).

Fezolinetant Study Day (FD) will be calculated as date of visit/assessment – first dose date of fezolinetant +1. Note that this is the same as Study Day for subjects randomized to fezolinetant in the double-blind treatment phase. Post-baseline data prior to first dose of fezolinetant Day 1 is excluded from the analyses described in this SAP unless otherwise noted.

6.10.1.1 Efficacy Analysis Windows

For the analysis windows, the last day is always the last day of exposure to overall study treatment. For VMS diary data the last day of exposure includes events reported until

8:00AM the following day. For example, the Week 1 average will include diary data from 8AM on Day 1 through 7:59AM on Day 8.

For both VMS and non-VMS efficacy, the end of overall treatment is defined as the last day of treatment (e.g., discontinued study or entered follow-up period) and the follow-up period begins at last overall treatment + 1 day.

The study week determination for the vasomotor symptoms data is based on the following:

Analysis study week	**Analysis window**
Baseline	Non-missing days from D-10 up to and including D-1
Week 1	FD1 to FD7
Week 2	FD8 to FD14
Week 3	FD15 to FD21
Week 4	FD22 to FD28
Week 5	FD29 to FD35
Week 6	FD36 to FD42
Week 7	FD43 to FD49
Week 8	FD50 to FD56
Week 9	FD57 to FD63
Week 10	FD64 to FD70
Week 11	FD71 to FD77
Week 12	FD78 to FD84
Week 16	FD110 to FD116
Week 20	FD138 to FD144
Week 24	FD166 to FD172
Week 28	FD194 to FD200
Week 32	FD222 to FD228
Week 36	FD250 to FD256
Week 40	FD278 to FD284
Week 44	FD306 to FD312
Week 48	FD334 to FD340
Week 52	FD362 to FD368
Week 53 (Follow-up)	FUD1 to FUD7
Week 54 (Follow-up)	FUD8 to FUD14
Week 55 (Follow-up)	FUD15 to FUD21

Calculations for weeks up through Week 12 end on the last day of the study week. Calculations for weeks beyond the double-blind treatment period are centered about protocol-specified visits to match visit-based efficacy assessments. For example, the Week 12 analysis window ends on FD84 (7*12=84), but the Week 16 visit is centered around

protocol specified FD113 (FD110 to FD116). Calculations in the follow-up period end on the last day of the study week, based on study days in the follow-up period.

Weeks 44, 48 and 52 will be missing for subjects randomized to placebo during the double-blind treatment period because they are only on active treatment for 40 weeks.

A weekly average will be calculated if any information, including zero VMS events, was reported for 4 or more days.

All other efficacy assessments, including those from unscheduled visits and regardless of visit label, will be allocated to analysis visits based on the table below:

Target day	**Analysis window**	**Analysis visit**
D1	last non-missing value on or prior to day 1	Baseline
D29	FD22 to FD36	Week 4
D85	FD72 to FD 92	Week 12
D113	FD106 to FD120	Week 16
D141	FD134 to FD148	Week 20
D169	FD162 to FD176	Week 24
D197	FD190 to FD204	Week 28
D225	FD218 to FD232	Week 32
D253	FD246 to FD 260	Week 36
D281	FD274 to FD288	Week 40
D309	FD302 to FD316	Week 44
D337	FD330 to FD344	Week 48
D365	FD358 to FD372	Week 52

The VMS efficacy windows are smaller because they are calculated from daily data and are not dependent on the scheduling of visits. The non-VMS efficacy endpoint are not collected during the follow-up period.

If more than one observation exists within the analysis window, the observation closest to the scheduled visit day will be selected for that visit. If there are two observations that have the same distance from the scheduled day, the value that is before the scheduled day will be selected in the analysis. If more than one observation is made on the same day, an average value if continuous, or the worst value if categorical, will be included in the analysis.

6.10.1.2 Safety Analysis Windows

Except for adverse events, for the analysis window, the last day is always 1 day after the last day of exposure to overall study treatment.

The end of overall treatment is defined as the last day of treatment (e.g., discontinued study or entered follow-up period) +1 and the follow-up period for analysis windows begins at last overall treatment + 2 days and onward (Follow-up Study Day 2).

The data summary by visits will be done following the analysis windows specified in the table below:

Analysis Visits	Scheduled Day in Protocol	Analysis Windows (Fezolinetant day)			
		a)	b)	c)	d)
Baseline	Day 1	Last non-missing value on or prior to fezolinetant day 1 (inclusive)			
Week 2	Day 15	2 to 22			
Week 4	Day 29	23 to 43	2 to 43	2 to 57	
Week 8	Day 57	44 to 71	44 to 71		
Week 12*	Day 85	72 to 92	72 to 99	58 to 99	2 to 127
Week 14	D99	93 to 106			
Week 16	D113	100 to 127	100 to 127	100 to 127	
Week 20	D141	128 to 155	128 to 155	128 to 155	
Week 24	D169	156 to 183	156 to 183	156 to 183	128 to 267
Week 28	D197	184 to 211	184 to 211	184 to 211	
Week 32	D225	212 to 239	212 to 239	212 to 239	
Week 36	D253	240 to 267	240 to 267	240 to 267	
Week 40	D281	268 to 295	268 to 295	268 to 295	
Week 44	D309	296 to 323	296 to 323	296 to 323	
Week 48	D337	324 to 351	324 to 351	324 to 351	
Week 52**	D365	352 to 376	352 to 376	352 to 376	268 to 376
Week 55	D386	Last non-missing follow-up day			

a) Apply to liver biochemistry and INR testing

b) Apply to complete Clinical laboratory and urinalysis, Vital signs

c) Apply to Blood PD/PK sample, including hormones (PD only at Week 55)

d) C-SSRS

*The analysis window for the Week 12 ECG assessment (Target Day 85) is FDay 2 through FDay 99

**The analysis window for the Week 52 ECG is last non-missing day from FDay 100 to FUD1; The analysis window for the Week 52 bone marker is FDay 2 to FUD1; The analysis window for the Week 52 TVU and biopsy assessments is FDay 2 until last treatment + 30 days.

If more than one observation exists within the analysis window, the same rules will be followed as described above for efficacy analysis visit windows, except that if there are two observations that have the same distance from the scheduled day, the value that is after the scheduled day will be selected in the analysis.

6.10.2 Imputation Rules for Incomplete Dates

In case of missing partial start and stop dates for concomitant medications or targeted medical history, the following rules will be used:

If the start date is missing or partial:

- if the month is missing, use January

- if the day is missing, use the first day of the month under consideration
- if the year is missing, use year of the informed consent date
- if the entire date is missing, use informed consent date

If the stop date is missing or partial:

- if the month is missing, use December
- if the day is missing, use the last day of the month under consideration
- if the year or the entire date is missing, set the stop date to December 31st, 2099

If the imputed start date is after the stop date, then the imputed start date will be 1 day prior to the stop date.

For AEs, a missing or incomplete onset date will be imputed according to the following conventions.

If an onset date is missing, the imputed onset date will be the date of first dose of study drug.

If only the year is known for the AE onset date, the imputed onset date will be the latest of the following non-missing dates:

- Date of first dose of study drug
- January 1 of the year of AE onset date

If only the month and year is known for the onset date, set the surrogate onset date to the first day of that month and then apply the following rules.

- If the month and year of the onset date is prior to the month and year of the first dose of study drug, then the surrogate onset date will be the imputed onset date.
- If the month and year of the onset date is on or after the month and year of the first dose of study drug, then the imputed onset date will be the **latest** of the following non-missing dates:
    - Date of first dose of study drug
    - Surrogate onset date

      If the imputed onset date is after the adverse event end date, the imputed onset date will be the same as the adverse event end date.

7 REVISION AND RATIONALE

7.1 List of Changes in SAP Version 2.0

The changes from the Version 1 to Version 2 that may impact the statistical analyses are listed in the table below:

SAP Sections	Description	Rationale
Relevant sections (see Section 9.1 Appendix 1)	Added additional statistical analyses due to the COVID-19 pandemic	To evaluate potential impact of the COVID-19 pandemic
6.5.1	Modified the list of AESIs	To be consistent with the protocol amendments (version 3.0)
6.5.4	Removed the summary of QTc change from baseline outliers	The summary of QTc change from baseline outliers provides limited additional medical relevant information
6.10.12	Upper bound of the Week 52 analysis window is modified to 376. The analysis window for the Week 52 TVU and biopsy assessments is Fezolinetant Study Day 2 (FD2) until last treatment + 30 days.	To be consistent with other study (0304) in the program.

8 REFERENCES

Avis NE, Coeytaux RR, Isom S, Prevette K, Morgan T. Acupuncture in menopause (AIM) study: a pragmatic, randomized controlled trial. Menopause. 2016;23:626-37.

Busner J, Targum SD. The Clinical Global Impressions Scale: applying a research tool in clinical practice. Psychiatry. 2007;4:28-37.

James R. Carpenter , James H. Roger & Michael G. Kenward (2013) Analysis of Longitudinal Trials with Protocol Deviation: A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation, Journal of Biopharmaceutical Statistics, 23:6, 1352-1371, DOI: 10.1080/10543406.2013.834911

EMA CHMP Guideline On Clinical Investigation Of Medicinal Products For Hormone Replacement Therapy Of Oestrogen Deficiency Symptoms In Postmenopausal Women [EMEA/CHMP/021/97 Rev. 1]. October 2005. Available at: https://www.ema.europa.eu/en/clinical-investigation-medicinal-products-hormone-replacement-therapy-oestrogen-deficiency-symptoms.

EMA/CHMP (2020a), Guidance to Sponsors on How to Manage Clinical Trials During the COVID-19 Pandemic, https://www.ema.europa.eu/en/documents/press-release/guidancesponsors-how-manage-clinical-trials-during-covid-19-pandemic_en.pdf

EMA/CHMP (2020b), Points to Consider on Implications of Coronavirus Disease (COVID-19) on Methodological Aspects of Ongoing Clinical Trials, EMA/158330/2020. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/points-considerimplications-coronavirus-disease-covid-19-methodological-aspects-ongoing-clinical_en.pdf

EuroQol Research Foundation. EQ-5D-5L: About. Version 18 Apr 2017. Available at https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/. Accessed 11 Oct 2018.

FDA Draft Guidance, Estrogen and Estrogen/Progestin Drug Products to Treat Vasomotor Symptoms and Vulvar and Vaginal Atrophy Symptoms – Recommendations for Clinical Evaluation. January 2003. Available at: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm071643.pdf. Accessed 14 Jan 2019.

FDA (2020), Guidance on Conduct of Clinical Trials of Medical Products During COVID-19 Public Health Emergency, Guidance for Industry, Investigators, and Institutional Review Boards. Updated on May 14, 2020. Available at: https://www.fda.gov/regulatoryinformation/search-fda-guidance-documents/fda-guidance-conduct-clinical-trials-medicalproducts-during-covid-19-pandemic

Hilditch JR, Lewis J, Peter A, van Maris B, Ross A, Franssen E et al. A menopause-specific quality of life questionnaire: development and psychometric properties. Maturitas. 1996;24:161-75.

ICH Harmonized Tripartite Guideline E 3. Structure and Content of Clinical Study Reports, November 1995. (www.ich.org; Guidelines; "Efficacy" Topics)

ICH Harmonized Tripartite Guideline E 9. Statistical Principles for Clinical Trials, February 1998. (www.ich.org; Guidelines; "Efficacy" Topics)

Khanna D, Maranian P, Rothrock N, Cella D, Gershon R, Khanna PP, et al. Feasibility and construct validity of PROMIS and "legacy" instruments in an academic scleroderma clinic. Value Health. 2012;15:128-34.

Lewis JE, Hilditch JR, Wong CJ. Further psychometric property development of the Menopause-Specific Quality of Life questionnaire and development of a modified version, MENQOL-Intervention questionnaire. Maturitas. 2005;50:209-21.

O'Kelly M & Ratitch B. Clinical Trials with Missing Data: A Guide for Practitioners. Statistics in Practice (2014):189 – 190.

PROMIS SD: Patient-Reported Outcomes Measurment Information System. Sleep Disturbance. 08 Jul 2015. Available at: https://www.assessmentcenter.net/documents/PROMIS%20Sleep%20Disturbance%20Scoring%20Manual.pdf. Accessed 18 Oct 2018.

PROMIS SRI: Patient-Reported Outcomes Measurment Information System. Sleep-Related Impairment. 08 Jul 2015. Available at: https://www.assessmentcenter.net/documents/PROMIS%20Sleep-Related%20Impairment%20Scoring%20Manual.pdf. Accessed 18 Oct 2018.

Reilly Associates. Work Productivity and Activity Impairment (WPAI) General Information. Updated 04 Feb 2013. Available at http://www.reillyassociates.net/WPAI_General.html. Accessed 11 Oct 2018.

Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmaco Economics. 1993;4:353-65.

van Reenen M, Janssen B. EQ-5D-5L User Guide: basic information on how to use the EQ-5D-5L instrument. Version 2.1. April 2015. Available at https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf. Accessed 11 Oct 2018.

9 APPENDICES

9.1 Appendix 1: Additional Statistical Analysis due to COVID-19

The novel coronavirus (SARS-CoV-2) is a new strain of coronavirus that had not previously been identified in humans. On January 30, 2020, the World Health Organization (WHO) declared the outbreak a public health emergency of international concern. On March 11, 2020, WHO characterized COVID-19 as a pandemic. Regulatory agencies have rapidly published guidance for clinical trial sponsors to address COVID-19 issues (FDA 2020, EMA 2020a, 2020b).

The COVID-19 pandemic has a global impact on the conduct of clinical trials of medical products. Challenges may arise, for example, from quarantines, site closures, travel limitations, interruptions to the supply chain for the investigational product, or other considerations if site personnel or subjects become infected with COVID-19. While the top priority is to protect the safety of the subjects, the other two objectives are to maintain the scientific integrity of the study and to ensure compliance with good clinical practice (GCP).

For this study, all planned subjects were randomized into the study before Astellas paused recruitment in all their clinical studies on March 30, 2020. In order to capture the relevant impact of the COVID-19 pandemic, dedicated CRF pages, the risk benefit plans, and alternative measures were implemented in the study as described in the study protocol. Since the study data may be impacted by the COVID-19 pandemic, additional statistical analyses are planned for this study.

The additional statistical analyses due to the COVID-19 pandemic are planned in the following sections of this SAP.

Section	Topic
6.2.2	Study population
6.3	Study Drug Exposure and Compliance
6.4.1	Analysis of Secondary Efficacy Endpoints
6.5.1	Adverse Events
6.5.2.1	Clinical Laboratory Evaluation/Liver Safety Assessment

9.2 Appendix 2: Author and Approver Signatures

E-signatures are attached at the end of document

Signatures

Prepared by:	*E-signatures are attached at end of document* PPD	Date:	Date (DD Mmm YYYY)
Approved by:	*E-signatures are attached at end of document* PPD	Date:	Date (DD Mmm YYYY)
Approved by:	*E-signatures are attached at end of document* PPD	Date:	Date (DD Mmm YYYY)

STATISTICAL ANALYSIS PLAN

(for the 12-week Double-Blind Treatment Period)

Version 3.0, dated 30 September 2020

A Phase 3, Randomized, Placebo-controlled, 12-week Double-blind Study, followed by a Non-Controlled Extension Treatment Period, to Assess the Efficacy and Safety of Fezolinetant in Women Suffering from Moderate to Severe Vasomotor Symptoms (Hot Flashes) Associated with Menopause

Skylight 2

ISN: 2693-CL-0302

IND number: 130277

EudraCT 018-003529-27

NCT04003142

Sponsor: **Astellas Pharma Global Development, Inc. (APGD)**

1 Astellas Way

Northbrook, IL 60062

Table of Contents

I. LIST OF ABBREVIATIONS AND KEY TERMS

List of Abbreviations

Abbreviations	Description of abbreviations
ANCOVA	Analysis of Covariance
ASCM	Analysis Set Classification Meeting
ATE	Active Treatment Extension
CI	Confidence Intervals
CMH	Cochran Mantel Haenszel
COVID-19	Coronavirus Disease 2019
FAS	Full Analysis Set
GCP	Good Clinical Practice
ISE	Integrated Summary of Efficacy
MAR	Missing at Random
MMRM	Mixed Model Repeated Measures
PAP	Psychometric Analysis Plan
PKAS	Pharmacokinetics Analysis Set
PPS	Per-Protocol Analysis Set
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SD	Standard Deviation
TLF	Tables, Listings and Figures
VMS	Vasomotor Symptoms

List of Key Terms

Terms	Definition of terms
Double-blind Treatment Period	The 12 weeks of treatment after the first administration of study drug and prior to first administration of study drug in the active treatment extension period. Only this period contains a placebo treatment.
Active Treatment Extension Period	Period of time after the first administration of study drug in the active treatment extension period after completing 12 weeks of placebo-controlled double-blind treatment until the last dose of the study drug (at Week 52 or early discontinuation).
Follow-up Period	Period of time after the last dose of study drug or last assessment of the protocol. Additional safety data and follow-up observations for sustained adverse events are conducted during this period.

1 INTRODUCTION

This Statistical Analysis Plan (SAP) contains technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes procedures for executing the statistical analysis to fulfil the objectives of the study.

The analyses described in the SAP will be conducted after all subjects complete the double-blind treatment period at 12 weeks. Only data pertaining to the double-blind treatment period (all data for subjects early discontinued during the double-blind period, or data from 12-week double-blind period prior to first dose of study drug in the active treatment extension period for subjects entered the extension period) will be used. A separate SAP will describe the analysis of the 52-week study data period, including the active treatment extension period.

This SAP will be approved prior to the database lock for the 12-week double-blind period.

Changes from the planned analyses in the finalized SAP that impact the statistical analyses will be documented and justified in the Clinical Study Report (CSR).

2 STUDY OBJECTIVE(S) AND DESIGN

2.1 Study Objective(s)

Primary objective:

- To evaluate the efficacy of fezolinetant versus placebo on the frequency and severity of moderate to severe vasomotor symptoms (VMS).
  - The estimand of the primary objective will use a hypothetical strategy and compare patients as though they had continued on the assigned treatment.

Key secondary objective:

- To evaluate the efficacy of fezolinetant versus placebo on patient-reported sleep disturbance.

Secondary objectives:

- To evaluate the effect of fezolinetant versus placebo on the frequency and severity of moderate to severe VMS at weekly time points.
- To evaluate the safety and tolerability of fezolinetant.

Exploratory objectives:

- To evaluate population pharmacokinetics (popPK) of fezolinetant and metabolite ES259564.
- To evaluate the effect of fezolinetant on pharmacodynamic markers.
- To evaluate the efficacy of fezolinetant versus placebo on the frequency and severity of mild, moderate and severe VMS.
- To evaluate the short-term and sustained effects of fezolinetant versus placebo on patient-reported sleep disturbance.

- To evaluate the effect of fezolinetant versus placebo on the following patient-reported outcomes (PROs): global assessments of VMS and sleep disturbance, overall sleep-wake function, quality of life and work productivity.

2.2 Study Design

This is a randomized, 12-week double-blind, placebo-controlled, parallel group, multicenter clinical study to assess the efficacy and safety of fezolinetant in women suffering from moderate to severe VMS associated with menopause. Approximately 450 subjects will be enrolled into this study, 150 subjects per treatment arm. Duration of treatment is 52 weeks. After completing 12 weeks of treatment, subjects on placebo will be reassigned to 30 mg or 45 mg of fezolinetant in the active treatment extension period through end of study. Subjects who were already randomized on an active arm will continue on their assigned dose for the remaining 40 weeks of treatment. Following the completion (or early discontinuation [ED]) of the treatment period (week 52), subjects will complete an end of treatment (EOT; or ED) visit and final safety follow-up visit 3 weeks after the last dose of study drug is administered (week 55).

Given the design, a 12-week treatment analysis will occur to assess efficacy and safety during the double-blind treatment phase. This will occur after all subjects have completed 12 weeks of treatment. Efficacy and safety data will be analyzed, excluding the endpoints measured after 12 weeks. Since all primary and secondary analyses only based on data through Week 12, no alpha adjustment is required as the information fraction at the 12-week analysis is 100%.

Details of the schedule of clinical assessments are available in the protocol.

2.3 Randomization

The first 12 weeks of the study are double blind. Subjects will be randomized to receive fezolinetant or placebo in a blinded fashion such that neither the investigator, sponsor's study management team, clinical staff, nor the subject will know which agent is being administered. The randomization number will be assigned based on information obtained from the Interactive Response Technology (IRT).

Subjects will be randomized in a 1:1:1 ratio to a treatment arm according to the randomization schedules and stratified by smoking status (current smoker vs former/never) through IRT. Subjects who complete the 12-week treatment period and have received placebo will be randomized to either fezolinetant 30 mg or 45 mg for the extension period. The assigned treatment in the extension period will remain blinded to the site personnel and participating subjects. The site personnel will dispense the treatment according to the IRT system's assignment. Specific procedures for randomization through the IRT are contained in the IRT specifications manual.

3 SAMPLE SIZE

A total of 450 subjects are planned to be randomized; 150 subjects in each treatment arm.

The phase 2b dose-ranging study (ESN_HF_205) included 7 active doses. In this study, the observed least-squares mean difference between fezolinetant and placebo in change from baseline to week 12 (and week 4) in mean daily frequency of moderate to severe VMS ranged from -1.8 to -3.0.

For a pairwise comparison using a 2-sample t-test at a 2-sided 2.5% alpha (adjusting for 2 active doses within study -0302), 120 subjects would provide 79% or more power to detect the following effect sizes assuming a SD of 5:

Assumed treatment difference in mean daily frequency	Power for pairwise test
-2.0	79%
-2.3	90%
-2.5	94%

For change from baseline to week 12 (and week 4) in mean severity of moderate to severe VMS, the observed mean treatment differences in the phase 2b study ranged from -0.2 and -1.0.

For a pairwise comparison using a 2-sample t-test at a 2-sided 2.5% alpha (adjusting for 2 active doses within study -0302), 120 subjects would provide the at least 79% power to detect the following effect sizes assuming a SD of 1:

Assumed treatment difference in mean severity	Power for pairwise test
-0.40	79%
-0.46	90%
-0.50	94%

NOTE: The combined power for testing all 4 co-primary endpoints will be lower than the power for each considered individually; however, the co-primary endpoints are correlated (especially the same endpoint at different time points) and this will serve to limit the potential reduction of power.

Assuming approximately 20% of subjects discontinue prematurely prior to 12 weeks of treatment in study -0302, the number of subjects will be increased from 120 to 150 subjects per arm. This should enable the analysis of the primary efficacy endpoints with sufficient power.

This sample size would also provide over 95% power to detect a difference of 4.3 from placebo on the key secondary endpoint of the PROMIS sleep disturbance questionnaire, using a 2-sample t-test at a 2-sided 2.5% alpha (adjusting for 2 active doses) assuming a SD of 7 [Avis et al, 2016].

4 ANALYSIS SETS

In accordance with International Conference on Harmonization (ICH) recommendations in guidelines E3 and E9, the following analysis sets will be used for the analyses.

The determination of whether subjects are included or excluded from the safety and efficacy analysis sets will be made prior to database unblinding for the 12-week double-blind period.

4.1 Full Analysis Set

The full analysis set (FAS) will consist of all subjects who are randomized and receive at least 1 dose of study drug. This will be the primary analysis set for efficacy analyses. The randomized treatment for each subject will be used for summaries by treatment group based on the FAS, even if a subject erroneously received a different treatment.

4.2 Safety Analysis Set

The safety analysis set (SAF) consists of all randomized subjects who took at least 1 dose of study drug, and will be used for safety analyses. A subject erroneously receiving a treatment different from their randomized treatment will be assigned to the treatment group that the patient received as first dose.

The SAF will be used for summaries of demographic and baseline characteristics and all safety and tolerability related variables.

4.3 Per Protocol Set

The per protocol sets (PPS) will consist of the subset of the FAS who do not meet criteria for exclusion from PPS listed in Section 4.3.1 below.

Final judgments on exclusion of subjects from Per Protocol Analysis Set - Week 4 (PPS4) and Per Protocol Analysis Set - Week 12 (PPS12) will be made prior to randomization unblinding.

The PPS will be a secondary analysis set for efficacy analyses. Select demographic and baseline characteristics may also be summarized for the PPS.

4.3.1 Reasons for Exclusion From PPS4

The following reasons may lead to subject's exclusion from PPS4:

- No measurement of the primary efficacy endpoint available at Week 4.
- <85% interactive diary compliance during the 4 week treatment period.
- Treatment compliance less than or equal to 85% between randomization and Week 4.

4.3.2 Reasons for Exclusion From PPS12

The following reasons may lead to subject's exclusion from PPS12:

- No measurement of the primary efficacy endpoint available at Week 12.
- <85% interactive diary compliance during the 12 week treatment period.
- Treatment compliance less than or equal to 85% between randomization and Week 12.

4.4 Pharmacokinetics Analysis Set (PKAS)

The PKAS will be defined in a separate analysis plan. Results of the population PK analysis will not be reported in the Clinical Study Report but in a separate population PK report.

4.5 Pharmacodynamic Analysis Set (PDAS)

The pharmacodynamic analysis set (PDAS) will include the subjects from the administered population for whom sufficient pharmacodynamic measurements were collected. The PDAS will be used for all analyses of pharmacodynamic data.

5 ENDPOINTS

For VMS symptoms, baseline frequency is calculated based on the number of moderate to severe hot flashes in the 10 days immediately prior to randomization. Subjects are required to have reported VMS symptoms on at least 7 of these 10 days to meet the eligibility criterion, and the average is based on all the non-missing days. The value is calculated within the ePRO device and used to assess the VMS frequency eligibility criterion, as well as for the analysis of VMS efficacy baseline. Baseline severity is calculated based on these same days, but is not provided by the ePRO device because it is not used for eligibility.

For efficacy endpoints collected at visits, the last non-missing assessment on or prior to the first dose of study treatment (investigational product or placebo) is the baseline.

For completeness, all protocol-specified endpoints are listed in the following subsections. However, analyses pertaining to endpoints at time points beyond the 12-week double-blind period will be described in the SAP for the active treatment extension period.

5.1 Primary Efficacy Endpoints

The primary efficacy objective requires the evaluation of 4 co-primary endpoints:

- Mean change in the frequency of moderate to severe VMS from baseline to week 4
- Mean change in the frequency of moderate to severe VMS from baseline to week 12
- Mean change in the severity of moderate to severe VMS from baseline to week 4
- Mean change in the severity of moderate to severe VMS from baseline to week 12

Subjects will record the number of vasomotor symptoms and severity of each vasomotor symptom via the ePRO daily diary.

The 24 hour vasomotor symptoms time frame runs from 08:00 AM to 07:59 AM (next day).

Frequency of moderate to severe vasomotor symptoms is the number of moderate to severe vasomotor symptoms per 24h.

The severity of an individual vasomotor symptom (VMS) is defined as follows [FDA Draft Guidance for Industry, 2003 and EMA CHMP Guideline, 2005]:

- Mild: sensation of heat without sweating
- Moderate: sensation of heat with sweating, able to continue activity
- Severe: sensation of heat with sweating, causing cessation of activity

Severity of post-baseline co-primary severity endpoints will be calculated using a weighted average defined as shown below. Note that the calculation for severity at baseline uses a similar formula, but does not include mild VMS events in the numerator or denominator. The calculation of baseline and post-baseline severity is based on specific direction from the FDA during the Type B end-of-phase 2 meeting of April 17, 2019. This calculation allows for the capture all hot flashes on treatment, including mild.

([number of mild hot flashes/day × 1] + [number of moderate hot flashes/day × 2] + [number of severe hot flashes/day × 3]) / total number of daily mild/moderate/severe hot flashes

At baseline, severity is zero for any individual days on which subjects have zero moderate or severe symptoms. Severity for post-baseline individual days is zero for subjects who have zero mild, moderate or severe vasomotor symptoms.

At baseline, as mentioned at the start of Section 5, frequency and severity are based on the non-missing values in the 10 days immediately prior to randomization. After baseline, daily frequency and severity for specific weeks (e.g., week 4, week 12) will be calculated as the average over non-missing days over 7 days period. Days without any reporting of VMS events will be considered missing for that day. For both post-baseline frequency and severity, a daily average per week will be derived if any information for 4 or more days was reported.

All 4 co-primary endpoints must be statistically significant for a given dose to be considered successful. Details regarding determination of statistical significance are provided in Section 6.4.

5.2 Secondary Efficacy Endpoints

5.2.1 Key Secondary Endpoint

The key secondary efficacy objective examines the effect of fezolinetant versus placebo on the following:

- Mean change in the Patient-Reported Outcomes Measurement Information System Sleep Disturbance – Short Form 8b (PROMIS SD SF 8b) total score from baseline to week 12

Alpha will only be propagated to the key secondary endpoint if all co-primary endpoints are statistically significant at both dose levels. If all the co-primary endpoints are statistically significant between fezolinetant and placebo at both doses, the statistical testing will be performed on the key secondary efficacy endpoint between fezolinetant and placebo, using Hochberg with alpha = 0.05, as part of the family-wise (overall) error rate.

5.2.1.1 Patient-Reported Outcomes Measurement Information System Sleep Disturbance – Short Form 8b (PROMIS SD SF 8b)

The PROMIS is a National Institutes of Health Roadmap initiative designed to improve PRO measures using state-of-the-science methods. The PROMIS SD SF 8b [PROMIS Sleep-Disturbance, 2015; Khanna, 2012] assesses self-reported sleep disturbance over the past 7 days and includes perceptions of restless sleep; satisfaction with sleep; refreshing

sleep; difficulties sleeping, getting to sleep or staying asleep; amount of sleep; and sleep quality. Because it assesses the patient's experience of sleep disturbance, the measure does not focus on specific sleep-disorder symptoms or ask patients to report objective measures of sleep (e.g., total amount of sleep, time to fall asleep and amount of wakefulness during sleep). Responses to each of the 8 items range from 1 to 5, and the range of possible summed raw scores is 8 to 40. Higher scores on the PROMIS SD SF 8b indicate more of the concept measured (disturbed sleep).

5.2.2 Secondary Endpoints

The secondary efficacy objectives examine the effect of fezolinetant versus placebo on the following:

- Mean change in the frequency of moderate and severe VMS from baseline to each week up to week 12
- Mean change in the severity of moderate and severe VMS from baseline to each week up to week 12
- Mean percent reduction in the frequency of moderate and severe VMS from baseline to each week up to week 12
- Percent reduction ≥ 50% and at 100% in the frequency of moderate and severe VMS from baseline to each week up to week 12
- Mean change in the frequency of moderate to severe VMS from baseline to week 24 (descriptive) [Analysis to be described in the 52-Week SAP]
- Mean change in the Severity of moderate to severe VMS from baseline to week 24 (descriptive) [Analysis to be described in the 52-Week SAP]
- Score on the Patient Global Impression of Change (PGI-C) in VMS at each visit

The severity of moderate to severe VMS for all secondary endpoints in this Section will be calculated as described in Section 5.1, where baseline severity include moderate and severe VMS events, while post-baseline severity includes mild, moderate and severe VMS events in the calculation.

5.2.2.1 VMS responder

Two separate definitions of the frequency of moderate and severe VMS responder will be used to assess each subject. The derivations at each week are as follows, using Week 12 as an example:

- Subject has ≥ 50% reduction from baseline to week 12.
    - $(VMS_{week12} - VMS_{BL})/ VMS_{BL} \leq -0.5$
      where VMS_{week12} and VMS_{BL} are the frequency of moderate and severe VMS at week 12 and baseline, respectively;
- Subject has 100% reduction from baseline to week 12 if frequency at week 12 = 0 VMS events.

Subjects with missing data will be considered as a non-responder.

Percent reduction ≥ 75% ≥ 90% in the frequency of moderate and severe VMS from baseline to each week, and the frequency of mild, moderate and severe VMS responder will be defined in a similar way (see exploratory endpoints).

5.3 Exploratory Efficacy Endpoints

- Mean change in the frequency of mild, moderate and severe VMS from baseline to each week up to week 12
- Mean change in the severity of mild, moderate and severe VMS from baseline to each week up to week 12
- Mean percent reduction in the frequency of mild, moderate and severe VMS from baseline to each week up to week 12
- Percent reduction ≥ 75% and ≥ 90% in the frequency of moderate and severe VMS from baseline to each week up to week 12 (VMS responder, which will be calculated in a similar way as stated in Section 5.2.2.1)
- Percent reduction ≥ 50%, ≥ 75%, ≥ 90%, and at 100% in the frequency of mild, moderate and severe VMS from baseline to each week up to week 12 (VMS responder, which will be calculated in a similar way as stated in Section 5.2.2.1)
- Mean change in the frequency and severity of moderate and severe VMS from baseline to each visit in the active treatment extension period and the follow-up visit [Analysis to be described in the 52-Week SAP]
- Change in serum concentrations of sex hormones and sex hormone-binding globulin (SHBG) from baseline to each visit
- Mean change in serum concentrations of BSAP, P1NP and CTX from baseline to each visit [Analysis to be described in the 52-Week SAP]
- Plasma concentrations of fezolinetant and metabolite ES259564 at pre-specified time points [Analysis to be described in the 52-Week SAP]
- Mean change on the PROMIS SD SF 8b total score from baseline to each visit
- Mean change on the Patient-Reported Outcomes Measurement Information System Sleep-Related Impairment – Short Form 8a (PROMIS SRI SF 8a) total score from baseline to each visit
- Score on Patient Global Impression of Severity in Sleep Disturbance (PGI-S SD) at each visit
- Score on Patient Global Impression of Change in Sleep Disturbance (PGI-C SD) to each visit
- Mean change on the Menopause-Specific Quality of Life (MENQOL) total score from baseline to each visit
- Mean change on the MENQOL domain scores from baseline to each week up to week 12
- Mean change on the Euro-Qol 5D-5L (EQ-5D-5L) Visual Analog Scale (VAS) from baseline to each visit
- Mean change on the Work Productivity and Activity Impairment questionnaire specific to VMS (WPAI-VMS) domain scores from baseline to each visit.

- Mean change in the severity of moderate and severe VMS from baseline to each week up to week 12 (see specification in Section 5.3.1)
- Mean and mean change from baseline in the daily frequency of moderate and severe VMS for the first week (see specification in Section 5.3.1)

Note that many of the above protocol-specified exploratory endpoints are scheduled beyond the 12-week double-blind treatment period and, as mentioned earlier, will be described in the SAP for the active treatment extension period.

5.3.1 VMS events

Exploratory endpoints that include mild post-baseline VMS events (frequency, severity, percent reduction, and responder analyses) also consider mild VMS at baseline.

The severity of moderate to severe VMS in this Section will be calculated as described in Section 5.1, where both baseline and post-baseline severity include only moderate and severe VMS events in the calculation.

For the daily frequency of moderate and severe VMS endpoint, the baseline frequency and post-baseline daily frequency will be calculated as described in Section 5.1. In addition, the co-primary endpoints will also be re-derived excluding data affected by COVID-19.

5.3.2 Hormones

Hormones being collected include estradiol (E2), follicle stimulating hormone (FSH), luteinizing hormone (LH), sex hormone-binding globulin (SHBG), total and free testosterone, androstenedione, dehydroepiandrosterone (DHEA), and estrone.

5.3.3 Bone Markers

Bone markers being collected include Bone Specific Alkaline Phosphatase (BSAP), carboxy-terminal telopeptide of Type I Collagen (CTX), Procollagen 1 N-Terminal Propeptide (P1NP). However, they only collected at the randomization and Week 52/end of treatment visit, so analyses will be described in the active treatment extension period SAP.

5.3.4 Patient Global Impression Scales (PGI-C VMS, PGI-S SD, PGI-C SD)

The PGI is comprised of 2 companion 1-item PRO measures analogous to the Clinical Global Impression (CGI) scales [Busner J & Targum SD, 2007]. These measures provide brief, stand-alone global assessments prior to and after initiating a study medication. The Patient Global Impression evaluates the following: (1) patient-perceived severity of a condition (PGI-S) and (2) patient-perceived change from the initiation of treatment (PGI-C). In this study, PGI scales will be used to evaluate meaningful within-person changes over time in VMS (PGI-C) and sleep disturbance (PGI-S and PGI-C).

The PGI-C VMS asks: “Compared to the beginning of this study, how would you rate your HFs/night sweats now?” Subject ratings range from (1) much better to (7) much worse.

The PGI-C SD asks: Compared to the beginning of this study, how well are you sleeping now?” Subject ratings range from (1) much better to (7) much worse.

The PGI-S SD asks: "How would you rate the severity of any problems you currently have while sleeping at night?" Subject ratings range from (1) no problems to (4) severe problems.

5.3.5 Patient-Reported Outcomes Measurement Information System Sleep Disturbance – Short Form 8b (PROMIS SD SF 8b)

See Section 5.2.1.1. PROMIS SD SG 8b total score will be analyzed at each visit. Based on the correspondence from the FDA in the Type B meeting minutes from April 2019, it also will be analyzed at each visit after removing question #4 - I had difficulty falling asleep.

5.3.6 Patient-Reported Outcomes Measurement Information System Sleep Disturbance – Short Form 8a (PROMIS SRI SF 8a)

The PROMIS SRI SF 8a [PROMIS Sleep-Related Impairment, 2015] is an 8-item PRO measure that evaluates self-reported perceptions of alertness, sleepiness and tiredness during usual waking hours and the perceived functional impairments during wakefulness associated with sleep problems or impaired alertness. Though this measure does not directly assess cognitive, affective or performance impairment, it does measure waking alertness, sleepiness and function within the context of overall sleep-wake function. The PROMIS SRI SF 8a is a universal rather than disease-specific instrument, and has a 7-day recall period.

Responses to each of the 8 items on the PROMIS SRI SF 8a range from 1 to 5, and the range of possible summed raw scores is 8 to 40. Higher scores indicate more of the concept measured (sleep-related impairment).

5.3.7 Menopause-Specific Quality of Life (MENQOL)

The MENQOL is a 29-item PRO measure that assesses the impact of 4 domains of menopausal symptoms, as experienced over the last week: vasomotor (items 1 to 3), psychosocial (items 4 to 10), physical (items 11 to 26) and sexual (items 27 to 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 6 (extremely bothersome) scale [Lewis et al, 2005].

Each item score ranges from 1 to 8, and each domain is scored separately; each domain mean ranges from 1 to 8 [Lewis et al, 2005; Hilditch et al, 1996]. The overall questionnaire score is the mean of the domain means. Higher scores represent more bothersome menopausal symptoms.

5.3.8 Euro-QoL 5D-5L (EQ-5D-5L) with Visual Analog Scale (VAS)

The EQ-5D-5L is a 5-item standardized measure of health status that provides a simple, generic measure of health for clinical and economic appraisal [EuroQol Research Foundation, 2018; van Reenen & Janssen, 2015]. This PRO measure comprises 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems and extreme problems. The subject is asked to indicate her health state by selecting the most appropriate statement in each of the 5 dimensions. This decision results in a 1-digit number that expresses the level selected for that dimension. The digits for the 5 dimensions

can be combined into a 5-digit number that describes the patient's health state. The data from each of the 5 dimensions will be summarized.

The EQ-5D-5L VAS is a subject-reported measure that records the respondent's self-rated health on a vertical VAS where the endpoint is labeled 'Best imaginable health state' and 'Worst imaginable health state.' The scale ranges from 0 to 100, where 100 indicates the subject is in her best possible health state and 0 indicates the subject is in her worst possible health state. Subjects mark an 'X' on the scale to rate their health status that day.

5.3.9 Work Productivity and Activity Impairment questionnaire specific to VMS (WPAI-VMS)

The WPAI-VMS is a 6-item PRO measure that examines VMS-related work productivity and activity in the preceding 7 days [Reilly Associates, 2013; Reily et al, 1993]. It consists of 4 domains: absenteeism (the percentage of work time missed because of VMS in the past 7 days), presenteeism (the percentage of impairment experienced while at work in the past 7 days because of VMS), overall work productivity loss (overall work impairment measured by combining absenteeism and presenteeism to determine the total percentage of missed time) and activity impairment (the percentage of impairment in daily activities because of VMS in the past 7 days). If the subject is unemployed, only 2 of the 6 items on this questionnaire require completion.

WPAI-VMS outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity (i.e., worse outcomes) [Reilly Associates, 2013].

5.4 Safety Endpoints

Safety evaluations will include the following endpoints and other safety-related assessments:

- Frequency and severity of AEs
- TVUs and endometrial biopsy findings [Analysis to be described in the 52-Week SAP]
- Change from baseline to each timepoint in vital signs: sitting systolic and diastolic blood pressure and pulse rate
- Change from baseline to each timepoint in ECG parameters
- Change from baseline to each timepoint in laboratory tests: hematology, biochemistry and urinalysis

6 STATISTICAL METHODOLOGY

6.1 General Considerations

Continuous data will be summarized descriptively including the number of subjects (n), mean, standard deviation (SD), median, minimum and maximum. Categorical data will be summarized by frequencies and percentages. Percentages by categories will be based on the number of subjects with no missing data, i.e. the percentages for the non-missing categories will add up to 100%.

All statistical comparisons will be conducted using 2-sided tests controlling the overall error rate at the 5% significance level, as described in Section 6.4.

All data summarization and analyses will be performed using SAS® on Unix. Specifications for table, figures, and data listing formats can be found in the Tables Listings and Figures (TLF) specifications.

As mentioned in the introduction, this analysis plan describes the analyses of the 12-week double-blind treatment period.

6.2 Study Population

In general, data such as patient disposition, demographics and baseline characteristics will be summarized for FAS population as the definition of FAS is identical to SAF if no patients were given wrong study drug. In the event when FAS is not identical to SAF then data will be summarized for both FAS and SAF as specified in the relevant sections.

6.2.1 Disposition of Subjects

Disposition of subjects will be summarized for the FAS by treatment group and overall. Number of subjects who complete or prematurely discontinue from the 12-week treatment period will be summarized by treatment group and overall. For the discontinuations, the primary reason reported by the investigator will be summarized.

Number and percentage of subjects for each analysis set will be summarized by treatment group and overall. In addition, number and percentage of subjects excluded from the PPS4 and PPS12 by reason for exclusion will be summarized for FAS.

6.2.2 Protocol Deviations

The number and percentage of subjects with the following protocol deviation criteria will be summarized for SAF for each criterion and overall, by treatment group and total as well as by investigative site. Subjects deviating from a criterion more than once will be counted once for the corresponding criterion.

The unique identifiers will be as follows:

PD1 - Entered into the study even though they did not satisfy entry criteria,

PD2 - Developed withdrawal criteria during the study and was not withdrawn,

PD3 - Received wrong treatment or incorrect dose,

PD4 - Received excluded concomitant treatment.

The number of subjects who were impacted by COVID-19 will be summarized by impact item and treatment group and overall for SAF population. The impact items include treatment discontinuation, dosing change/interruption, death from COVID-19, healthcare encounter due to COVID-19, protocol deviation including missed visit windows due to COVID-19, and impacted diary data due to COVID-19.

An overview listing of subjects who were affected by COVID-19 will be generated by the impact items.

Detailed information on how these assessments are impacted by COVID-19 can be found in relevant data listings with a flag for COVID-19.

6.2.3 Demographic and Other Baseline Characteristics

Demographics, tobacco history (smoking status), substance use (e-cigarette, alcohol, cannabis), caffeine beverage intake, height and weight, and targeted medical history will be summarized descriptively by treatment group and total group for SAF. Note that current versus former or never smoking status is a stratification factor for randomization. Age group categories include ages within <40 years, >=40 years to <46 years, >=46 years to <51 years, >=51 years to <56 years, >=56 years to <61 years, >=61 years to <66 years, and >=66 years.

Medical history is coded in MedDRA, and will be summarized by System Organ Class (SOC) and Preferred Term (PT) by treatment group and total group for the SAF.

6.2.4 Previous and Concomitant Medications

Previous and concomitant medications will be summarized by therapeutic subgroup (ATC 2nd level) and chemical subgroup (ATC 4th level) and preferred WHO name by treatment group for the SAF. Subjects taking the same medication multiple times will be counted once per medication and investigational period.

Previous medications are defined as medications that patients started prior to first administration of study medication. Concomitant medications are defined as any medications that subjects took after the first dose of study medication and either up to 30 days from last dose of study drug, or first dose of study drug during the active treatment extension period, whichever comes sooner. Medications that started prior to first administration of study drug and continued while study drug was given will be counted in both previous and concomitant medications.

6.3 Study Drug Exposure and Compliance

Duration and compliance of study drug will be summarized for SAF by treatment group and overall.

Number and percentages of subjects with the following categories of study drug duration will be summarized: >0 to ≤ 1 days, >1 to ≤ 7 days, >7 to ≤ 14 days, >14 to ≤ 21 days, >21 to ≤ 28 days, >28 to ≤ 42 days, >42 to ≤ 56 days, >56 to ≤ 84 days, and >84 days.

Number and percentages of subjects with the following cumulative categories of study drug duration will be summarized: ≥ 1 day, ≥ 7 days, ≥ 14 days, ≥ 21 days, ≥ 28 days, ≥ 42 days, ≥ 56 days and ≥ 84 days.

Overall treatment compliance with the dosing schedule will be examined for subjects in the SAF whose total study drug count and first and last days of treatment are known. Compliance will be calculated compared to the actual treatment period of dosing (first to last day of the double-blind treatment period), not to the planned treatment period.

Percent overall compliance over the 12-week double-blind treatment period will be summarized in two ways for the SAF:

- Descriptive statistics will be presented by treatment group.
- Percent compliance will be categorized according to the following categories by treatment group:
    - $< 50\%$,
    - $\geq 50\%$ to $< 85\%$,
    - $\geq 85\%$ to $< 120\%$,
    - $\geq 120\%$, and
    - Unknown.

Overall treatment compliance through Week 4 (study day 28 of exposure) will be analyzed in a similar manner.

To explore the impact of the COVID-19 pandemic, additional analysis will be performed by excluding the days of missing study drug/drug interruption due to COVID-19 in the duration and compliance calculation, according to the study drug interruption CRF page.

6.4 Analysis of Efficacy

All statistical comparisons will be conducted using 2-sided tests at the $\alpha = 0.05$ significance level unless specifically stated otherwise. There are additional testing details below regarding the 4 co-primary endpoints. All null hypotheses will be of no treatment difference, all alternative hypotheses will be 2-sided and control for 2 comparisons to placebo using the Hochberg approach, unless specifically stated otherwise.

The family-wise type I error rate for the 2 active dose groups compared to placebo for the 4 co-primary efficacy endpoints will be controlled using a Hochberg approach. All 4 co-primary endpoints must be statistically significant for a given dose to be considered successful. Given that, the largest p-value in each dose group will be used to test that dose because it represents the least significant of the co-primary endpoints.

If the larger of the two maximum p-values, one from each dose, is less than 0.05, then both dose groups are considered statistically significant. If not, then if the smaller of the p-values is less than 0.025, the second compared dose is considered statistically significant. Othewise, neither dose is considered statistically significant. The trial will be considered successful if at least one dose is statistically significant.

All analysis of efficacy will be presented by treatment group for FAS, unless specified otherwise.

In general, the mechanism of missing data due to COVID-19 will be assumed to be missing at random (MAR).

6.4.1 Analysis of Primary Efficacy Endpoints

6.4.1.1 Primary Analysis for Primary Efficacy Endpoints

For each of the 4 co-primary efficacy endpoints, a mixed models repeated measures analysis of covariance (MMRM) will be used with treatment group, week (Week 1 through Week 12) and smoking status (current vs former/never) as factors, with baseline weight and baseline VMS as covariates, as well as an interaction of treatment by week and an interaction of baseline measurement by week. An unstructured covariance structure shared across treatment groups will be used to model the within-patient errors (and then Toeplitz if model does not converge). The Kenward-Roger approximation will be used to estimate denominator degrees of freedom and adjust standard errors. This analysis will use a restricted maximum likelihood–based repeated-measures approach. The treatment difference will be estimated at all study weeks (subtract placebo change from baseline from each fezolinetant response). MMRM will use all available on-treatment data to inform mean treatment effect estimates without requiring explicit imputation for missing data due to intercurrent events (i.e., for discontinued subjects). This approach is consistent with the hypothetical strategy used for the estimand, which is to compare subjects as though they had continued on the assigned treatment. Given the primary objective of the study, the analysis should describe the pharmacologic effect of the randomized treatments. The MMRM approach, and the missing at random (MAR) assumption, allows us to hypothesize that subjects had continued on assigned treatment while not being influenced by how subjects might have done without treatment or with alternative treatment(s).

Comparisons between the fezolinetant and placebo at Week 4 and Week 12 will be calculated based on least-squares mean contrasts using a 2-tailed 95% confidence interval (CI) and p-values. The LS Means will be estimated using weights proportional to the percentage of women who were randomized as current smokers. Details regarding determination of statistical significance can be found in Section 6.4.

The hypothesis for each pairwise comparison is given as follows:

H0: The change from baseline at week 4 (or 12) for fezolinetant and placebo are the same

H1: The change from baseline at week 4 (or 12) for fezolinetant and placebo are not the same

6.4.1.2 Sensitivity Analysis for Primary Efficacy Endpoints

The following sensitivity analyses will be performed for the primary efficacy endpoints to assess the robustness of the primary analysis.

MMRM model based on PPS populations

Supportive analyses will be carried out for the co-primary efficacy endpoints based on the PPS populations. The data will be analyzed using the same MMRM model as the primary analyses, but the comparison of interest is a particular week within each population (Week 4 when using PPS4, and Week 12 when using PPS12).

Simplified MMRM model

The analysis for each of the co-primary endpoints also will be conducted using a simplified MMRM model with treatment group and week as factors, with baseline measurement as a covariate as well as an interaction of treatment by week and baseline by week.

Discontinuation-reason based multiple imputation

The MMRM analysis that will be used to perform the primary analysis of the primary endpoint assumes that missingness is at random. That is, the model assumes that the trajectory of VMS events over time for subjects who prematurely discontinue is similar to the trajectory for those observed in their own treatment arm which is valid so long as that assumption is reasonable.

Discontinuation-reason based multiple imputation (MI) will be used to examine the sensitivity of the primary analysis results to departures from that underlying assumption and will assess a situation where data for subjects who discontinue early follow a pattern which is missing not at random. Specifically, MI will be used for imputation of missing data, using "Jump to Reference" algorithm (where placebo is the reference group) [Carpenter et al. 2013] for subjects who discontinue active treatment due to AEs (AE reported as primary study drug treatment status in the End of Treatment eCRF). The analysis will be implemented using the general three-step process (imputation phase, followed by analysis phase, followed by pooling phase) described in O'Kelly and Ratitch, 2014.

1. The imputation phase will implement MI via sequential modelling for the "Jump to Reference" algorithm. In particular, when implementing the "Jump to Reference" algorithm via sequential modelling, this means that only baseline values are used in the imputation of missing values for subjects in the fezolinetant 30 mg or 45 mg groups who discontinue due to AEs (this variation of the "Jump to Reference" algorithm is also referred to as the "Unconditional Reference" approach). The imputation phase will generate M imputed datasets, where M=30.
2. The analysis phase will perform the primary MMRM analysis model of the primary efficacy endpoint as described in Section 6.4.1.1 for each of the M imputed datasets.
3. Rubin's rules will be used to generate an overall set of pooled results which combines the analysis results from the M imputed datasets.

The overall set of pooled results will present:

- Least squares (LS) mean estimates, SE and 2-sided 90% confidence interval (CI) for mean change from baseline to Week 12 within a treatment group (fezolinetant 30 mg or 45 mg, placebo),
- The difference in LS means for fezolinetant 30 mg or 45 mg versus placebo, SE for the difference and 2-sided 90% CI for the difference at Week 4 and Week 12,
- Two-sided p-value for fezolinetant 30 mg or 45 mg versus placebo at Week 4 and Week 12, obtained using the difference in the LS means.
- Descriptive summary statistics at Week 4 and Week 12 will be derived using the mean of the M imputed datasets.

COVID-19 Impact

The primary analysis will be repeated by excluding diary data that are impacted by COVID-19, which includes the diary data recorded after the subjects were diagnosed of COVID-19, after study drug discontinuation due to COVID-19, and during dose interruption due to COVID-19.

6.4.1.3 Subgroup Analysis for Primary Efficacy Endpoint

The co-primary efficacy endpoint analyses will be repeated for the following subgroups using the simplified MMRM model. However, p-values for treatment difference will not be provided. Descriptive statistics and least square mean (SE) will be provided for each subgroup.

The subgroups will be defined below.

- Smoking status (current versus former/never), per randomization strata.
- Race (African-American vs Non-African-American, Asian vs Non-Asian, White vs Non-White)
- BMI (>=18.5 kg/m^2 to <25 kg/m^2, >=25 kg/m^2 to <30 kg/m^2, >=30 kg/m^2)
- Age (<55, >=55)

For subjects who are recorded more than one race, the subjects will be excluded from the race subgroups, except:

- If no African-American is recorded, then the subject is included in Non-African-American subgroup
- If no Asian is recorded, then the subject is included in Non-Asian subgroup
- If no White is recorded, then the subject is included in Non-White subgroup.

6.4.2 Analysis of Key Secondary Efficacy Endpoint

The PROMIS SD SF 8b scale will be analyzed using MMRM, similar to the primary analysis of the co-primary endpoints, with spatial power as the backup covariance structure.

If all the co-primary endpoints are statistically significant between fezolinetant and placebo at both doses, then the 5% alpha from the analysis of the co-primary endpoints will be passed to the key secondary endpoint as part of the family-wise error rate.

As a sensitivity analysis, the analysis will be repeated by excluding data that are impacted by COVID-19, which includes the data recorded after the subjects were diagnosed of COVID-19, after study drug discontinuation due to COVID-19, during dose interruption due to COVID-19, and visits that are out of protocol window due to COVID-19 recorded in the eCRF.

6.4.3 Analysis of Secondary Efficacy Endpoints

The mean change in the frequency and severity of moderate and severe VMS from baseline to each week up to week 12 comes from the same model as the co-primary analyses, but examines the results at all weeks.

For percent reduction at weeks 4 and 12, an MMRM model as described in Section 6.4.1.1 will be used.

For each of the secondary responder endpoints, logistic regression will be used for the analysis for each week and endpoint. The analyses will include the treatment group and smoking status (current vs. former/never) as factors and baseline measurement (mean frequency of vasomotor symptoms) as a covariate.

For these responder endpoints, the number of subjects who achieve the reduction will be summarized by study week with descriptive statistics for each treatment arm. A Kaplan-Meier estimate will be used to estimate the time to response. Smooth average (SA) will be used for the time to response. SA value for day X-1, day X and Day X+1 would be assigned to day X. The time to response will be the first day with at least 50% (or 100%, or the exploratory endpoints 75%, 90%) response for both original numbers and SA number. For the non-responders, it will be censored at the end of the last day of the double-blind treatment (e.g., 8am the following day).

Comparisons between the fezolinetant and placebo will be calculated based on the odds ratio using a 2-tailed 95% confidence interval (CI) and p-values.

PGI-C VMS score will be analyzed using Cochran Mantel Haenszel (CMH) test with modified ridit scores.

For the changes in the frequency and severity of moderate and severe VMS from baseline to each week up to week 12 that are derived from diary data, the analysis will be repeated by excluding diary data that are impacted by COVID-19, which includes the diary data recorded after the subjects were diagnosed of COVID-19, after study drug discontinuation due to COVID-19, and during dose interruption due to COVID-19.

For the PGI-C VMS scores that are collected by visit, the analysis will be repeated by excluding data that are impacted by COVID-19, which includes the data recorded after the subjects were diagnosed of COVID-19, after study drug discontinuation due to COVID-19, during dose interruption due to COVID-19, and visits that are out of protocol window due to COVID-19 recorded in the eCRF.

6.4.4 Analysis of Exploratory Endpoints

Assessment of the mean change in the frequency and severity of mild, moderate and severe VMS from baseline to each week up to week 12 will be analyzed using the same MMRM model as the primary analyses.

For non-VMS efficacy endpoints Visits are used rather than weeks, as described in Section 6.10.1.1.

Percent reduction, PROMIS (SD and SRI total scores), MENQOL (domain and total scores), and WPAI-VMS domain scores will be assessed using an MMRM model as described in [Section 6.4.1.1 Co-Primary Analysis].

VMS responders will be analyzed as described for the secondary responder endpoints.

PGI-S and PGI-C (SD) and EQ-5D domain scores-will be summarized and analyzed using Cochran Mantel Haenszel (CMH) test with modified ridit scores. PGI-S and EQ-5D will also be adjusted for the baseline score. The mean and standard deviation will also be provided for the EQ-5D domain scores.

The severity calculation excludes mild events at both baseline and post-baseline, using the same model as the primary analyses.

The mean change in the severity of moderate and severe VMS from baseline to each week up to week 12 (both baseline and post-baseline severity include only moderate and severe VMS events in the calculation) will be analyzed using the same MMRM model as the primary analyses.

The median change in serum concentrations of sex hormones and sex hormone-binding globulin (SHBG) from baseline to weeks 4 and 12 will be summarized.

The daily frequency of moderate and severe VMS for the first week will be summarized and graphically displayed by each day.

Anexploratory analysis will be conducted using the thresholds for clinical meaningful within subject change from an anchor based analysis.

In order to provide further context for interpretation of change in the frequency of hot flashes, the anchor based analysis will be conducted to empirically define the threshold for meaningful within-subject reduction. A separate Psychometric Analysis Plan (PAP) [RTI Health Solutions] will provide the detailed description of the psychometric analytic methods to be used in determining the clinically meaningful within-subject reductions at week 4 and week 12. The thresholds will be estimated according to the prespecified PAP.

A subject will be classified as a responder if her change from baseline to Week 4 is equal or larger than the threshold at Week 4. Similarly, a subject will be classified as a responder if her change from baseline to Week 12 is equal or larger than the threshold at Week 12.

A logistic regression analysis will be conducted to compare each dose of fezolinetant to placebo. The logistic regression model will include the treatment group and smoking status (current vs. former/never) as factors, and baseline mean frequency of vasomotor symptoms as a covariate. Missing data will be imputed by considering the following approaches: (1) last-observation-carried-forward (primary), (2) missing as a non-responder, and (3) including observed cases only. Odds ratio (fezolinetant over placebo), its 2-sided 95% CI and nominal p-value will be presented.

This analysis will be conducted using the thresholds that are calculated from the pooled studies (ASP2693-CL-0301 and ASP2693-CL-0302) in the Integrated Summary of Efficacy

(ISE). An additional analysis will be conducted using the thresholds that are calculated from each study to further characterize the study results.

6.5 Analysis of Safety

All analysis of safety will be presented by treatment group for SAF, unless specified otherwise. For analyses by visit, the visit windows from Section 6.10.1.2 will be applied. Except for adverse events and follow-up visit data, only data up to one day after the last dose date will be considered for the analyses.

The baseline value for safety endpoint will be the last non-missing value taken on or prior to first dose of study drug unless otherwise noted.

6.5.1 Adverse Events

Per protocol, a TEAE is defined as an AE observed after starting administration of the study drug and up to 21 days after the last dose of the study drug.

A TEAE during the DB period, is defined as an AE observed after starting administration of the study drug and up to 21 days after the last dose of the double-blind study drug, and before the first dose of treatment for extension phase for subjects who entered extension treatment period.

A study drug-related TEAE is defined as any TEAE with a causal relationship of YES by the investigator.

The sponsor has a list of events that they classify as “always serious” events. If an AE is reported that is considered by the sponsor to be an SAE per this classification as “always serious”, the AE will be treated as an SAE in tables and listings, and flagged as ‘always serious’ in AE listings.

An overview table will include the following

- Number of TEAEs,
- Number and percentage of subjects with TEAEs,
- Number of drug related TEAEs,
- Number and percentage of subjects with drug related TEAEs,
- Number of serious TEAEs,
- Number and percentage of subjects with serious TEAEs,
- Number of serious drug related TEAEs,
- Number and percentage of subjects with serious drug related TEAEs,
- Number of TEAEs leading to permanent discontinuation of study drug,
- Number and percentage of subjects with TEAEs leading to permanent discontinuation of study drug, and
- Number of deaths.

The number and percentage of subjects with TEAEs in the following AE categories will be summarized by system organ class (SOC), high level term (HLT) and preferred term (PT):

- TEAEs,
- Drug-related TEAEs.

The number and percentage of subjects with TEAEs in the following AE categories will be summarized by SOC and PT:

- Serious TEAEs,
- Drug-related serious TEAEs,
- TEAEs leading to permanent discontinuation of study drug,
- Drug-related TEAEs leading to permanent discontinuation of study drug,
- TEAEs that equal to or exceed a threshold of 5% in any treatment group (threshold is based on PT),
- TEAEs excluding serious adverse events that equal to or exceed a threshold of 5.0% in any treatment group.

The number and percentage of subjects with TEAEs, as classified by PT only, will be summarized for each treatment group.

The number of TEAEs and the number and percentage of subjects with TEAEs, as classified by SOC and PT will be summarized by severity and by relationship to study drug. In the subject count, if a subject has multiple TEAEs with the same SOC or PT, but with differing severity or relationship, then the subject will be counted once with the worst severity and highest degree of relationship. If severity or relationship is missing for all episodes of the event, the subject will be counted under missing severity or relationship. In the adverse event count, the adverse events will be presented in each category they were classified to. Drug related TEAEs will be presented in a similar way by severity only.

Following adverse events of special interest are defined:

- Adverse event of uterine bleeding
- Endometrial hyperplasia, cancer, or disordered proliferative endometrium
- Adverse event of thrombocytopenia
- Adverse event of liver test elevations
- Adverse event of bone fractures
- Adverse events of abuse liability
- Adverse events of depression
- Adverse events of wakefulness
- Adverse events of effect on memory

Search terms defining the AESIs will be pre-specified and separately documented.

The number and percentage of subjects with AESIs, as classified by PT will be separately summarized for each treatment group and listings of these adverse events will be provided.

The number and percentage of subjects with TEAEs associated with COVID-19, as classified by SOC and PT, will be summarized for each treatment group and overall.

6.5.2 Clinical Laboratory Evaluation

Quantitative values evaluated by the central laboratory including hematology, biochemistry, and urinalysis will be summarized using mean, standard deviation, minimum, maximum and median by treatment group at each analysis visit. Additionally, a within-subject change will be calculated as the post-baseline measurement minus the baseline measurement and summarized in the same way. Each laboratory result will be classified as low (L), normal (N), or high (H) at each visit according to the laboratory supplied reference ranges.

The number and percentage of subjects below and above reference range will be summarized for each treatment group at each visit.

Number and percentage of subjects with platelets < 150000/uL (150 10^9/L) will also be summarized.

Frequency tabulations of qualitative clinical laboratory variables (urinalysis) will be presented for each treatment group at each visit.

For hematology and biochemistry shift tables will be presented for each treatment group:

- Summary shifts of reference range changes from baseline to worst finding during the treatment period (shift from normal or high to low, shift from normal or low to high, categorized increase [shift from low to normal or from normal to high], categorized no change [value stays in the same reference range], categorized decrease [shift from high to normal or from normal to low]).

For hormone related parameters, if the value is below the low limit of quantification, the value will be imputed as half of the LLOQ. In addition, the percentage of subjects who have BLOQ values at baseline who then have measurable post-baseline values will be summarized.

6.5.2.1 Liver Safety Assessment

The liver safety assessments will be summarized by the categories below based on the measurements from Alkaline Phosphatase (ALP), Alanine Transaminase (ALT), total bilirubin, Aspartate Transaminase (AST) and their combination. These parameters will be based on measurements from a central laboratory.

The subject's highest value during the double-blind investigational period (from day of first dose of study drug until first dose of double-blind treatment period) will be used.

- ALT > 3xULN, > 5xULN, > 8xULN, > 10xULN, >20xULN
- AST > 3xULN, > 5xULN, > 8xULN, > 10xULN, >20xULN
- ALT or AST > 3xULN, > 5xULN, > 8xULN, > 10xULN, >20xULN
- ALP > 1.5xULN
- Total Bilirubin >2xULN
- (ALT or AST > 3xULN) and Total Bilirubin > 2xULN
- (ALT or AST > 3xULN) and ALP< 2xULN and Total Bilirubin > 2xULN

The last 2 criteria where 2 or more parameters are evaluated will be with the measurements on the same day or up to 1 day apart. The denominator for each criterion will be the number of subjects who have at least one value during the double-blind period. The number and percentage of subjects meeting the criteria during the double-blind period will be summarized by treatment group.

Due to COVID-19, patients might not be able to visit the clinical site and have clinical laboratory tests taken at the clinical site and evaluated by the central laboratory, in this case clinical laboratory tests can be taken by a local laboratory. The liver biochemistry safety assessments will be summarized including central and locally collected laboratory data due to COVID-19. This analysis is considered the primary analysis for the liver safety assessments.

The secondary analysis for liver biochemistry will be repeated by the data evaluated by the central laboratory but excluding the data that are impacted by COVID-19, which include the data recorded after the subjects were diagnosed of COVID-19, after study drug discontinuation due to COVID-19, during dose interruption due to COVID-19, and visits that are out of protocol window or data collected via home visits due to COVID-19 recorded in the eCRF.

6.5.3 Vital Signs

Vital signs (systolic blood pressure, diastolic blood pressure, and pulse rate) will be summarized using mean, standard deviation, minimum, maximum and median by treatment group and visit. Additionally, a within-subject change will be calculated per visit as the post-baseline measurement minus the baseline measurement and summarized by treatment group and visit.

6.5.4 Electrocardiograms

ECG variables will be summarized using mean, standard deviation, minimum, maximum and median for each treatment group at each treatment visit and time point, including changes from baseline.

Number and percent of subjects with normal and abnormal results as assessed by the central reader for the 12 lead ECG will be tabulated by treatment group at each treatment visit and time point.

The QTc interval will be summarized using frequency tables for each treatment visit and time point for values of clinical importance using the range criteria below.

	QTc Interval Criteria Value (msec)
Normal	≤ 450
Borderline	> 450
Prolonged	> 480
Clinically significant	> 500

Number and percent of subjects with 12 lead ECG abnormalities as well as number and percent of subjects whose 12 lead ECG reading changed from normal at baseline to abnormal will be tabulated by treatment group at each treatment visit.

6.5.5 Other Safety-Related Assessments

C-SSRS results will be summarized by treatment group and visit.

6.6 Analysis of Pharmacokinetics

The statistical methods for PK data will be described in a separate analysis plan. Results of the population PK analysis will not be reported in the Clinical Study Report but in a separate population PK report.

Plasma concentration of fezolinetant and ES259564 will be listed.

6.6.1 Estimation of Pharmacokinetic Parameters

All details of the population PK analysis will be described in a separate analysis plan.

6.7 Analysis of Pharmacodynamics

Serum concentrations of sex hormones (LH, FSH, E2, Testosterone Total/Free, Androstenedione, DHEA, Estrone) and SHBG will be summarized using descriptive statistics (number of subjects, mean, SD, median, minimum and maximum) for each visit and time point and for change from the baseline to each postdose visit, if the data is available

6.8 Other Analyses

Not applicable.

6.9 Interim Analysis (and Early Discontinuation of the Clinical Study)

There will be no interim analysis.

6.10 Additional Conventions

6.10.1 Analysis Windows

Study Day for analysis windows during the double-blind treatment phase will be calculated as date of visit/assessment – first dose date +1.

Study Day for the follow-up phase will be calculated as date of visit/assessment – last dose date of double-blind treatment.

As noted in Section 1, only data pertaining to the double-blind treatment phase will be used in the 12-week treatment phase analysis:

- The active treatment extension period data (beginning from the second day of the active treatment extension period) will be excluded, assuming the Week 12 assessments are done before the first dose of active treatment extension period.
- Follow-up data that also falls into the active treatment extension period will be excluded (i.e., early discontinuations within the active treatment extension period phase).

6.10.1.1 Efficacy Analysis Windows

For the analysis windows (weeks prior to the extension period), the last day is always the last day of exposure to double-blind treatment. For VMS diary data the last day of exposure includes events reported until 8:00AM the following day. For example, the Week 1 average will include diary data from 8AM on Day 1 through 7:59AM on Day 8.

For both VMS and non-VMS efficacy, the end of overall treatment is defined as the last day of treatment (e.g., discontinued study or entered follow-up period) and the follow-up period begins at last overall treatment + 1 day.

The study week determination for the vasomotor symptoms data is based on the following:

Analysis study week	**Analysis window**
Baseline	Non-missing days from D-10 up to and including D-1
Week 1	D1 to D7
Week 2	D8 to D14
Week 3	D15 to D21
Week 4	D22 to D28
Week 5	D29 to D35
Week 6	D36 to D42
Week 7	D43 to D49
Week 8	D50 to D56
Week 9	D57 to D63
Week 10	D64 to D70
Week 11	D71 to D77
Week 12	D78 to D84
Extension period visits	Defined in extension period SAP

A weekly average will be calculated if any information, including zero VMS events, for 4 or more days was reported.

All other efficacy assessments, including those from unscheduled visits and regardless of visit label, will be allocated to analysis visits based on the table below:

Target day	**Analysis window**	**Analysis visit**
D1	last non-missing value on or prior to day 1	Baseline
D29	D22 to D36	Week 4
D85	D72 to Day 92	Week 12
Beginning on Day 99	Defined in extension period SAP	Extension period visits

If more than one observation exists within the analysis window, the observation closest to the scheduled visit day will be selected for that visit. If there are two observations that have the same distance from the scheduled day, the value that is before the scheduled day will be selected in the analysis. If more than one observation is made on the same day, an average value if continuous, or the worst value if categorical, will be included in the analysis.

6.10.1.2 Safety Analysis Windows

Except for adverse events, for the analysis window (weeks prior to the follow-up period), data within 1 day after the last day of exposure to double-blind treatment will be included.

The end of overall treatment is defined as the last day of treatment (e.g., discontinued study or entered follow-up period) +1 and the follow-up period begins at last overall treatment + 2 days and onward.

The data summary by visits will be done following the analysis windows specified in the table below:

Analysis Visits	**Scheduled Day in Protocol**	**Analysis Windows (day)**		
		a)	**b)**	**c)**
Baseline	Day 1	Last non-missing value on or priot to day 1 (inclusive)		
Week 2	Day 15	2 to 22		
Week 4	Day 29	23 to 43	2 to 43	2 to 57
Week 8	Day 57	44 to 71	44 to 71	
Week 12*	Day 85	72 to 92	72 to 92	58 to 92
Extension Period Visits	Beginning on day of first dose of extension period treatment	Defined in extension period SAP		

a) Apply to liver biochemistry and INR testing

b) Apply to complete Clinical laboratory and urinalysis, Vital signs

c) Apply to Blood PD/PK sample

*The analysis window for the Week 12 ECG assessment (Target Day 85) is Day 2 through Day 92

If more than one observation exists within the analysis window, the same rules will be followed as described above for efficacy analysis visit windows, except that if there are two observations that have the same distance from the scheduled day, the value that is after the scheduled day will be selected in the analysis.

6.10.2 Imputation Rules for Incomplete Dates

In case of missing partial start and stop dates for concomitant medications or targeted medical history, the following rules will be used:

If the start date is missing or partial:

- if the month is missing, use January
- if the day is missing, use the first day of the month under consideration
- if the year is missing, use year of the informed consent date
- if the entire date is missing, use informed consent date

If the stop date is missing or partial:

- if the month is missing, use December
- if the day is missing, use the last day of the month under consideration
- if the year or the entire date is missing, set the stop date to December 31st, 2099

If the imputed start date is after the stop date, then the imputed start date will be 1 day prior to the stop date.

For AEs, a missing or incomplete onset date will be imputed according to the following conventions.

If an onset date is missing, the imputed onset date will be the date of first dose of study drug.

If only the year is known for the AE onset date, the imputed onset date will be the latest of the following non-missing dates:

- Date of first dose of study drug
- January 1 of the year of AE onset date

If only the month and year is known for the onset date, set the surrogate onset date to the first day of that month and then apply the following rules.

- If the month and year of the onset date is prior to the month and year of the first dose of study drug, then the surrogate onset date will be the imputed onset date.
- If the month and year of the onset date is on or after the month and year of the first dose of study drug, then the imputed onset date will be the **latest** of the following non-missing dates:
  - Date of first dose of study drug
  - Surrogate onset date

    If the imputed onset date is after the adverse event end date, the imputed onset date will be the same as the adverse event end date.

7 REVISION AND RATIONALE

7.1 List of Changes in SAP Version 2.0

The changes from the Version 1 to Version 2 that may impact the statistical analyses are listed in the table below:

SAP Sections	Description	Rationale
Relevant sections	Added additional statistical analysis due to COVID-19	To evaluate potential impact of COVID-19
5.2.2	Moved PGI-C to Secondary endpoints Section	To be consistent with the protocol amendments (version 3.0)
5.3	Added 75%, 90% responder endpoints	To add clinically relevant exploratory endpoints
5.3, 5.3.1	Added an endpoint in the severity of moderate and severe VMS (include only moderate and severe VMS events in the calculation)	To extend the endpoints in a sensitivity analysis to exploratory endpoints for all visits
5.3, 5.3.1, 6.4.4	Added an endpoint in daily frequency of moderate and severe VMS for the first week	To explore early onset by adding daily VMS frequency for the first week
5.3.1, 6.4.4	Removed Frequency and severity of 12-hour daytime and nighttime VMS events.	Planned not to perform analysis on the 12-hour daytime and nighttime VMS events
6.2.3, 6.2.4	Removed analysis on FAS from demographics and baseline characteristics, medical history/medication Sections	The planned analysis on SAF is sufficient to provide information
6.4.1.3	Added subgroup on Age (<55, >=55)	To add medically meaningful subgroups
6.4.1.4	Added conditional analysis based on clinical meaningfulness of the statistical change in moderate to severe hot flash frequency	Potentially in supporting the primary endpoints
6.5.1	Modified the list of AESIs	To be consistent with the protocol amendments (version 3.0)
6.5.2.1	Added a category for liver assessment, (ALT or AST > 3xULN) and ALP< 2xULN and Total Bilirubin > 1.5xULN	To be aligned with CIOMS Guidance 2020
6.5.4	Removed the summary of QTc change from baseline outliers	The summary of QTc change from baseline outliers provides limited additional medical relevant information

7.2 List of Changes in SAP Version 3.0

The changes from the Version 2 to Version 3 that may impact the statistical analyses are listed in the table below:

SAP Sections	Description	Rationale
6.4.1.4, 6.4.4	Removed Section 6.4.1.4 on conditional analysis based on clinical meaningfulness of the statistical change in moderate to severe hot flash frequency, and added the analysis to Section 6.4.4 (Analysis of Exploratory Endpoints)	To reflect FDA's feedback.
6.5.2.1	Removed the category for liver assessment, (ALT or AST > 3xULN) and ALP< 2xULN and Total Bilirubin > 1.5xULN	The analysis of this category is only planned for Integrated Summary of Safety.

8 REFERENCES

Avis NE, Coeytaux RR, Isom S, Prevette K, Morgan T. Acupuncture in menopause (AIM) study: a pragmatic, randomized controlled trial. Menopause. 2016;23:626-37.

Busner J, Targum SD. The Clinical Global Impressions Scale: applying a research tool in clinical practice. Psychiatry. 2007;4:28-37.

James R. Carpenter, James H. Roger & Michael G. Kenward (2013) Analysis of Longitudinal Trials with Protocol Deviation: A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation, Journal of Biopharmaceutical Statistics, 23:6, 1352-1371, DOI: 10.1080/10543406.2013.834911

EMA CHMP Guideline On Clinical Investigation Of Medicinal Products For Hormone Replacement Therapy Of Oestrogen Deficiency Symptoms In Postmenopausal Women [EMEA/CHMP/021/97 Rev. 1]. October 2005. Available at: https://www.ema.europa.eu/en/clinical-investigation-medicinal-products-hormone-replacement-therapy-oestrogen-deficiency-symptoms.

EMA/CHMP (2020a), Guidance to Sponsors on How to Manage Clinical Trials During the COVID-19 Pandemic, https://www.ema.europa.eu/en/documents/press-release/guidancesponsors-how-manage-clinical-trials-during-covid-19-pandemic_en.pdf

EMA/CHMP (2020b), Points to Consider on Implications of Coronavirus Disease (COVID-19) on Methodological Aspects of Ongoing Clinical Trials, EMA/158330/2020. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/points-considerimplications-coronavirus-disease-covid-19-methodological-aspects-ongoing-clinical_en.pdf

EuroQol Research Foundation. EQ-5D-5L: About. Version 18 Apr 2017. Available at https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/. Accessed 11 Oct 2018.

FDA Draft Guidance, Estrogen and Estrogen/Progestin Drug Products to Treat Vasomotor Symptoms and Vulvar and Vaginal Atrophy Symptoms – Recommendations for Clinical Evaluation. January 2003. Available at: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm071643.pdf. Accessed 14 Jan 2019.

FDA (2020), Guidance on Conduct of Clinical Trials of Medical Products During COVID-19 Public Health Emergency, Guidance for Industry, Investigators, and Institutional Review Boards. Updated on May 14, 2020. Available at: https://www.fda.gov/regulatoryinformation/search-fda-guidance-documents/fda-guidance-conduct-clinical-trials-medicalproducts-during-covid-19-pandemic

Hilditch JR, Lewis J, Peter A, van Maris B, Ross A, Franssen E et al. A menopause-specific quality of life questionnaire: development and psychometric properties. Maturitas. 1996;24:161-75.

ICH Harmonized Tripartite Guideline E 3. Structure and Content of Clinical Study Reports, November 1995. (www.ich.org; Guidelines; "Efficacy" Topics)

ICH Harmonized Tripartite Guideline E 9. Statistical Principles for Clinical Trials, February 1998. (www.ich.org; Guidelines; "Efficacy" Topics)

JoAnn V. Pinkerton, Ginger Constantine, Eunhee Hwang, Ru-fong J. Cheng Desvenlafaxine compared with placebo for treatment of menopausal vasomotor symptoms: a 12-week, multicenter, parallel-group, randomized, double-blind, placebo-controlled efficacy trial, Menopause: The Journal of The North American Menopause Society. 2012, Vol. 20, No. 1, pp. 28-37.

Khanna D, Maranian P, Rothrock N, Cella D, Gershon R, Khanna PP, et al. Feasibility and construct validity of PROMIS and "legacy" instruments in an academic scleroderma clinic. Value Health. 2012;15:128-34.

Lewis JE, Hilditch JR, Wong CJ. Further psychometric property development of the Menopause-Specific Quality of Life questionnaire and development of a modified version, MENQOL-Intervention questionnaire. Maturitas. 2005;50:209-21.

O'Kelly M & Ratitch B. Clinical Trials with Missing Data: A Guide for Practitioners. Statistics in Practice (2014):189 – 190.

PROMIS SD: Patient-Reported Outcomes Measurment Information System. Sleep Disturbance. 08 Jul 2015. Available at: https://www.assessmentcenter.net/documents/PROMIS%20Sleep%20Disturbance%20Scoring%20Manual.pdf. Accessed 18 Oct 2018.

PROMIS SRI: Patient-Reported Outcomes Measurment Information System. Sleep-Related Impairment. 08 Jul 2015. Available at: https://www.assessmentcenter.net/documents/PROMIS%20Sleep-Related%20Impairment%20Scoring%20Manual.pdf. Accessed 18 Oct 2018.

Reilly Associates. Work Productivity and Activity Impairment (WPAI) General Information. Updated 04 Feb 2013. Available at http://www.reillyassociates.net/WPAI_General.html. Accessed 11 Oct 2018.

Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmaco Economics. 1993;4:353-65.

RTI Health Solutions. Interpretation of Meaningful Within-Patient Reduction in the Frequency of Hot Flashes Associated With Menopause. Psychometric Analysis Plan.

van Reenen M, Janssen B. EQ-5D-5L User Guide: basic information on how to use the EQ-5D-5L instrument. Version 2.1. April 2015. Available at https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf. Accessed 11 Oct 2018.

9 APPENDICES

9.1 Appendix 1 Additional Statistical Analysis due to COVID-19

The novel coronavirus (SARS-CoV-2) is a new strain of coronavirus that had not previously been identified in humans. On January 30, 2020, the World Health Organization (WHO) declared the outbreak a public health emergency of international concern. On March 11, 2020, WHO characterised COVID-19 as a pandemic. Regulatory agencies have rapidly published guidance for clinical trial sponsors to address COVID-19 issues (FDA 2020, EMA 2020a, 2020b).

The COVID-19 pandemic has a global impact on the conduct of clinical trials of medical products. Challenges may arise, for example, from quarantines, site closures, travel limitations, interruptions to the supply chain for the investigational product, or other considerations if site personnel or subjects become infected with COVID-19. While the top priority is to protect the safety of the subjects, the other two objectives are to maintain the scientific integrity of the study and to ensure compliance with good clinical practice (GCP).

For this study, all planned subjects were randomized into the study before Astellas paused recruitment in all their clinical studies on March 30, 2020. In order to capture the relevant impact of COVID-19, dedicated CRF pages, the risk benefit plans, and alternative measures were implemented in the study. Since the study data may be impacted by COVID-19, additional statistical analyses are planned for this study.

The additional statistical analysis due to COVID-19 is planned in the following sections.

Section	Topic
6.2.2	Study population
6.3	Study Drug Exposure and Compliance
6.4	Analysis of Efficacy
6.4.1.2	Sensitivity Analysis for Primary Efficacy Endpoints
6.4.2	Key Secondary Endpoints
6.4.3	Secondary Endpoints
6.5.1	Adverse Events
6.5.2.1	Clinical Laboratory Evaluation

9.2 Appendix 2 Author and Approver Signatures

E-signatures are attached at the end of document.

Signatures:

Prepared by: *E-signatures are attached at end of document* Date:

PPD

Date (DD Mmm YYYY)

Approved by: *E-signatures are attached at end of document* Date:

PPD

Date (DD Mmm YYYY)

Approved by: *E-signatures are attached at end of document* Date:

PPD

Date (DD Mmm YYYY)