STATISTICAL ANALYSIS PLAN

Version 2.0, dated 08-Dec-2021

A Randomized, Placebo-Controlled, Double-Blind Phase 3 Clinical Study to Investigate the Long-Term Safety of Fezolinetant in Women Suffering From Vasomotor Symptoms (Hot Flashes) Associated with Menopause

Skylight 4

ISN/Protocol 2693-CL-0304

IND number: 130277

EudraCT 2019-000275-16

NCT04003389

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

1 Astellas Way
Northbrook, IL 60062

Table of Contents

I. LIST OF ABBREVIATIONS AND KEY TERMS 4
1 INTRODUCTION 5
2 STUDY OBJECTIVE(S) AND DESIGN 5
2.1 Study Objective(s) 5
Primary Objective 5
Secondary Objective 5
Exploratory Objectives 5
2.2 Study Design 5
2.3 Randomization 6
3 SAMPLE SIZE 6
4 ANALYSIS SETS 7
4.1 Full Analysis Set (FAS) 7
4.2 Safety Analysis Set (SAF) 8
4.3 Endometrial Health Set (EH Set) 8
5 ENDPOINTS 8
5.1 Primary Safety Endpoint 8
5.2 Secondary Safety Endpoints 8
5.3 Exploratory Endpoints 8
5.3.1 Menopause-Specific Quality of Life (MENQoL) 9
5.3.2 EQ-5D-5L with Visual Analog Scale (VAS) 9
6 STATISTICAL METHODOLOGY 9
6.1 General Considerations 9
6.2 Study Population 10
6.2.1 Disposition of Subjects 10
6.2.2 Protocol Deviations 10
6.2.3 Demographic and Other Baseline Characteristics 10
6.2.4 Previous and Concomitant Medications 11
6.3 Study Drug Exposure and Compliance 11
6.4 Analysis of Efficacy 12
6.4.1 Analysis of Primary Efficacy Endpoint(s) 12
6.4.2 Analysis of Secondary Efficacy Endpoints 12

6.4.3 Analysis of Exploratory Endpoints 12
6.5 Analysis of Safety 13
6.5.1 Adverse Events 14
6.5.2 Clinical Laboratory Evaluation 16
6.5.3 Vital Signs 17
6.5.4 Electrocardiograms 17
6.5.5 Endometrial Health Assessment 18
6.5.6 Bone Mass 19
6.5.7 Other Safety-Related Assessments 19
6.5.8 Subgroups of Interest for safety 19
6.6 Analysis of Pharmacokinetics 20
6.6.1 Estimation of Pharmacokinetic Parameters 20
6.7 Analysis of Pharmacodynamics 20
6.8 Other Analyses 20
6.9 Interim Analysis (and Early Discontinuation of the Clinical Study) 20
6.10 Additional Conventions 20
6.10.1 Analysis Windows 20
6.10.2 Imputation Rules for Incomplete Dates 22
7 REVISION AND RATIONALE 23
7.1 List of Changes in SAP Version 2.0 23
8 REFERENCES 24
9 APPENDICES 25
9.1 Appendix 1 Additional Statistical Analysis due to COVID-19 25
9.2 Appendix 2 Author and Approver Signatures 26

I. LIST OF ABBREVIATIONS AND KEY TERMS

List of Abbreviations

Refer to the protocol for the rest of abbreviations.

Abbreviations	Description of abbreviations
AESI	Adverse Events of Special Interest
ANCOVA	Analysis of Covariance
BMI	Body Mass Index
CI	Confidence Interval
CMH	Cochran Mantel Haenszel
COVID-19	Coronavirus Disease 2019
CRF	Case Report Form
DMC	Data Monitoring Committee
FAS	Full Analysis Set
HLT	High Level Term
LSMP	Liver Safety Monitoring Panel
MMRM	Mixed Model Repeated Measures
NAFLD	Non-Alcoholic Fatty Liver Disease
NASH	Non-Alcoholic SteatoHepatitis
PD	Protocol Deviation
PRO	Patient Reported Outcome
PT	Preferred Term
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SD	Standard Deviation
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
TLF	Tables, Listings and Figures
VMS	Vasomotor Symptoms
EH	Endometrial Health

List of Key Terms

Refer to the protocol for the rest of key terms.

Terms	Definition of terms
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 final version of this SAP will be approved prior to database lock and unblinding the subject treatment assignments.

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

2 STUDY OBJECTIVE(S) AND DESIGN

2.1 Study Objective(s)

Primary Objective

- To evaluate the long-term safety and tolerability of fezolinetant in women seeking treatment for relief of Vasomotor Symptoms (VMS) associated with menopause.
- To evaluate the effect of fezolinetant on endometrial health after long-term treatment in women seeking treatment for relief of VMS associated with menopause.

Secondary Objective

- To evaluate the effect of fezolinetant on bone mineral density after long-term treatment in women seeking treatment for relief of VMS associated with menopause.

Exploratory Objectives

- To evaluate the effect of fezolinetant on subject-reported quality of life measures.
- To evaluate the pharmacokinetics of fezolinetant and its metabolite, ESN259564.

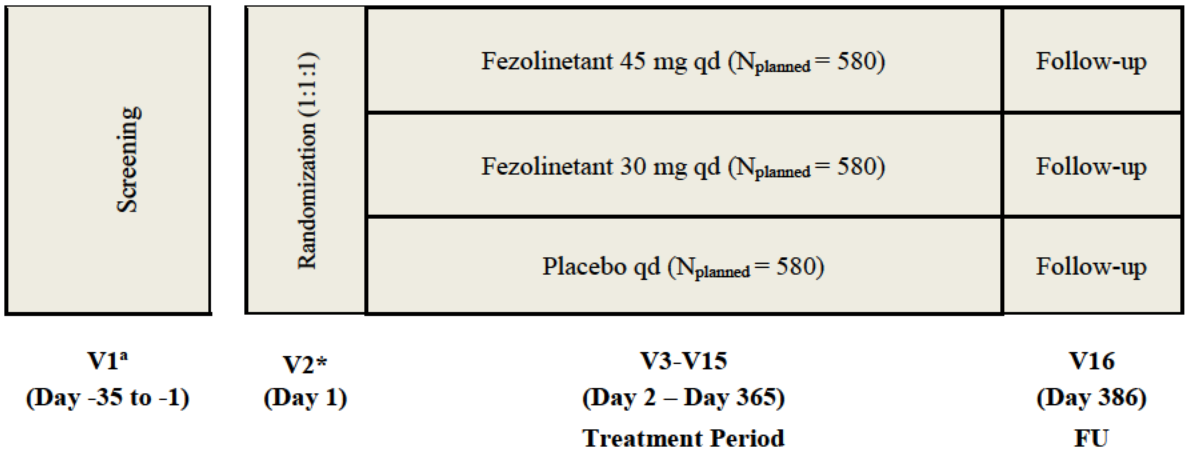
2.2 Study Design

This is a 52-week randomized, placebo-controlled, double-blind, parallel-group, multicenter clinical study to assess the safety and tolerability of fezolinetant in women seeking treatment for VMS associated with menopause.

The study visits will be performed on an outpatient basis. This study will consist of a screening period (days -35 to -1, including the screening visit [visit 1] assessments), a 52 week treatment period (day 1 [visit 2] to week 52 [visit 15]) and a follow up visit (week 55 [visit 16]) 3 weeks after the last dose of study drug. An extra 15 screening days are allowed for repeat biopsy, if necessary (days -50 to -1, including the screening visit [visit 1] assessments).

Approximately 1740 total subjects will be randomized into the study. Subjects will be randomized 1:1:1 into the following treatment groups:

- Fezolinetant 45 mg qd (approximately 580 subjects),
- Fezolinetant 30 mg qd (approximately 580 subjects),
- Placebo qd (approximately 580 subjects)



a. Screening is to be performed up to 35 days prior to randomization.
FU: follow-up; qd: once daily; V: visit.
* Refer to the schedule of assessments for visit 2b.

A Data Monitoring Committee (DMC) will oversee the safety of fezolinetant and a Liver Safety Monitoring Panel (LSMP) will monitor elevations of liver function tests for the duration of the study.

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

2.3 Randomization

Subjects will be randomized to receive fezolinetant 45 mg qd, fezolinetant 30 mg qd, or placebo in a blinded fashion such that 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. 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 specification manual for the study.

3 SAMPLE SIZE

The primary objective of this study is to assess long-term safety and tolerability. The sample size in this study is not calculated based on the statistical power for efficacy evaluation to detect treatment difference.

The total of 1740 subjects are planned to be randomized 1:1:1 to the fezolinetant 45 mg once daily group (580), fezolinetant 30 mg once daily group (580), or placebo group (580). This sample size would provide high probability to observe events of special interest that has a

fairly low background event rate of less than 1%. With the sample size, the following table illustrates the probability of observing 1 or more events, 2 or more events and 3 or more for different background event rates ranging from 0.10% to 0.90%.

	Per Treatment Arm		
Sample Size (1:1:1)	**n = 580**		
Background Event Rate	**Prob(#>=3)**	**Prob(#>=2)**	**Prob(#>=1)**
0.10%	2.11%	11.53%	44.03%
0.20%	11.18%	32.29%	68.69%
0.30%	25.33%	51.94%	82.49%
0.40%	40.94%	67.43%	90.22%
0.50%	55.46%	78.62%	94.54%
0.60%	67.63%	86.28%	96.95%
0.70%	77.16%	91.35%	98.30%
0.80%	84.27%	94.62%	99.05%
0.90%	89.38%	96.69%	99.47%

Prob(#>=1) means the probability of observing 1 or more events.

Prob(#>=2) means the probability of observing 2 or more events.

Prob(#>=3) means the probability of observing 3 or more events.

In addition, with an assumed background rate of 0.26% such as for endometrial hyperplasia, this sample size would provide the final number of evaluable subjectsto demonstrate that the point estimate for the rate of the event is less than or equal to 1% and the upper bound of the one-sided 95% confidence interval is less than or equal to 4% (as recommended in the FDA guidance on Estrogen and Estrogen/Progestin Drug Products to Treat Vasomotor Symptoms) with greater than 95% probability assuming up to 60% (including baseline, ED and subject refusal of endometrial biopsy at EOT) of subjects may not have evaluable biopsy data.

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 analysis set and the full analysis set (for the PRO exploratory endpoints) will be made prior to database lock for the primary report and prior to unblinding.

4.1 Full Analysis Set (FAS)

The full analysis set (FAS) will consist of all subjects who are randomized and receive at least one dose of study drug. This will be the primary analysis set for the PRO exploratory endpoints. 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 (SAF)

The safety analysis set (SAF) consists of all randomized subjects who took at least 1 dose of study drug. A subject erroneously receiving a treatment that is different from the 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 Endometrial Health Set (EH Set)

The endometrial health set (EH Set) will be defined as a subset of SAF who had evaluable biopsy at baseline and met the criteria specified in Section 6.5.5. This analysis set will be used to evaluate the incidence rate of endometrial hyperplasia and incidence rate of endometrial cancer and incidence rate of disordered proliferative endometrium.

5 ENDPOINTS

5.1 Primary Safety Endpoint

The primary variable will require the evaluation of the effect of fezolinetant on the following:

- Frequency and severity of AEs.
- Percentage of subjects with endometrial hyperplasia.
- Percentage of subjects with endometrial cancer.

5.2 Secondary Safety Endpoints

The secondary endpoints are:

- Change from baseline in endometrial thickness at 12 months.
- Percentage of subjects with disordered proliferative endometrium.
- Change from baseline in bone mass density (BMD) and trabecular bone score (TBS) at hip and spine at 12 months.
- Vital signs: sitting systolic and diastolic blood pressure and pulse rate.
- Laboratory tests: hematology, biochemistry and urinalysis.
- C-SSRS.
- ECG parameters.

5.3 Exploratory Endpoints

The exploratory endpoints are:

- Mean change on the Menopause-Specific Quality of Life (MENQOL) Total Score from baseline to specified time points (week 4, 12, 24 and 52).
- Mean change on the MENQOL Domain Scores from baseline to specified time points (week 4, 12, 24 and 52).
- Mean change on the Euro-Qol 5D-5L (EQ-5D-5L) Visual Analog Scale (VAS) from baseline to specified time points (week 4, 12, 24 and 52).

- Euro-Qol 5D-5L (EQ-5D-5L) domain scores at specified time points (week 4, 12, 24 and 52).
- Change from baseline to specified time points in serum concentrations of sex hormones and sex hormone-binding globulin (SHBG) (week 4, 12, 24, 52 and 55).
- Plasma concentration of fezolinetant and the fezolinetant metabolite ESN259564 at specified timepoints (week 4, 12, 24 and 52).

5.3.1 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.

5.3.2 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 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.

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, if applicable, will be conducted using 2-sided tests at the 5% significance level.

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.

6.2 Study Population

In general, data such as patient disposition, demographics and baseline characteristics will be summarized for SAF population.

6.2.1 Disposition of Subjects

Disposition of subjects will be summarized for the SAF by treatment group and overall. Number of subjects who complete or prematurely discontinue from the treatment or study (ie, follow up period) will be summarized by treatment group and overall. For the discontinuation, 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 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 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

The 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. 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. Hormone replacement therapy history will be summarized descriptively by treatment groupand total group for SAF.

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 patients took after the first dose of study medication and through 30 days from last dose of study drug. 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, fezolinetant total 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;
- > 84 to ≤ 168 days;
- > 168 to ≤ 252 days;
- > 252 to ≤ 365 days;
- > 365 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;
- ≥ 168 days;
- ≥ 252 days;
- ≥ 365 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 treatment period), not to the planned treatment period.

Percent overall compliance will be summarized in two ways for the SAF by treatment group, fezolinetant total and overall:

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

To evaluate impact of the COVID-19 to study drug intake, additional analyses 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

6.4.1 Analysis of Primary Efficacy Endpoint(s)

Not applicable.

6.4.1.1 Primary Analysis for Primary Efficacy Endpoint(s)

Not applicable.

6.4.2 Analysis of Secondary Efficacy Endpoints

Not applicable.

6.4.3 Analysis of Exploratory Endpoints

The exploratory endpoints include the MENQOL and the EQ-5D-5L which will be assessed at baseline and weeks 4, 12, 24 and 52. The exploratory PRO endpoints will be analyzed for the FAS set if not specified. Summary statistics will be provided by treatment group. For analyses by visit, the visit windows from Section 6.10.1 will be applied.

The exploratory PRO variables include the effect of fezolinetant on the following:

- Mean change on the Menopause-Specific Quality of Life (MENQOL) Total Score from baseline to specified time points.
- Mean change on the MENQOL Domain Scores from baseline to specified time points.
- Mean change on the Euro-Qol 5D-5L (EQ-5D-5L) Visual Analog Scale (VAS) from baseline to specified time points.
- Euro-Qol 5D-5L (EQ-5D-5L) domain scores

For the treatment comparison of continuous endpoints, a mixed model for repeated measures (MMRM) will be used. The model will include treatment, visit, and strata smoking status (current vs former/never) as factors, with baseline weight and baseline measurement as covariates, as well as an interaction of treatment by visit and an interaction of baseline measurement by visit. 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. EQ-5D domain scores-will be summarized and analyzed using Cochran Mantel Haenszel (CMH) test with modified ridit scores. EQ-5D will also be adjusted for the baseline score.

6.5 Analysis of Safety

Overall long-term safety of fezolinetant and the effect of fezolinetant on endometrial health will be the primary objective of this study. The safety assessments include adverse events, laboratory assessments, vital signs, C-SSRS, ECG, endometrial health assessment and imaging (mammogram, DXA, TVU). In all analysis of safety, the treatment group includes placebo, fezolinetant 30mg, fezolinetant 45mg and fezolinetant total will be presented by treatment group for SAF, unless specified otherwise. For analyses by visit, the analysis windows from Section 6.10.1 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.

Primary Safety Variables

- Frequency and severity of AEs.
- Percentage of subjects with endometrial hyperplasia.
- Percentage of subjects with endometrial cancer.

Secondary Safety Variables

- Change from baseline in endometrial thickness at 12 months.
- Percentage of subjects with disordered proliferative endometrium.
- Change from baseline in bone mass density (BMD) and trabecular bone score (TBS) at hip and spine at 12 months.
- 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 laboratory tests: hematology, biochemistry and urinalysis.
- C-SSRS at each timepoint.
- Change from baseline to each timepoint in ECG parameters

Exploratory Safety Variables

- Change from baseline to specified time points in serum concentrations of sex hormones and sex hormone-binding globulin (SHBG) (week 4, 12, 24, 52 and 55).
- Plasma concentration of fezolinetant and the fezolinetant metabolite ESN259564 at specified timepoints (week 4, 12, 24 and 52).

In general, summary statistics will be provided for all the safety parameters. Treatment comparisons will be performed for endometrial and bone density related endpoints. For AESI, the rate, odds ratio and their corresponding 95% exact confidence intervals will be presented, which will be constructed based on Santner-Snell approach. An ANCOVA model, with treatment and strata smoking status (current vs former/never) as factors, with baseline weight and baseline as covariates, will be used for single assessment continuous endpoints. Each following section will specify the analysis method appropriately.

The change and percentage reduction in serum concentrations of sex hormones and sex hormone-binding globulin (SHBG) from baseline to each analysis visit will be summarized.

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.

6.5.1 Adverse Events

Per protocol, a treatment emergent adverse event (TEAE) is defined as the AEs that occurred on or after first dose of study drug and 21 days after the last dose of study drug (when AE collection closes).

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 causally 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.

In order to adjust for differences in subjects' durations and the potential differential dropout rates between the treatment groups, an overview table to report exposure adjusted incidence rate (per 100 subject-years) for the above AE categories will be summarized. The total subject-years of exposure is calculated as the sum over subjects of the total duration of exposure divided by 365.25:

$$\sum_i \frac{ld_i - fd_i + 1}{365.25},$$

where fd_i and ld_i denote the first dose date of study drug and the last dose date of study drug for subject i.

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.

Among TEAEs, following adverse events of special interest (AESI) are defined:

- Adverse event of endometrial hyperplasia/cancer or disordered proliferative endometrium
- Adverse event of liver test elevations
- Adverse event of uterine bleeding
- Adverse event of thrombocytopenia
- Adverse event of bone fractures
- Adverse event of abuse liability
- Adverse event of depression
- Adverse event of wakefulness
- Adverse event 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.

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

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

Number and percentage of subjects with platelets < 150x10^9/L will be separarely summarized for each treatment group.

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]).

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 investigational period (from day of first dose of study drug until last dose of 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 > 3×ULN) and Total Bilirubin > 1.5×ULN
- (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 investigational period. The number and percentage of subjects meeting the criteria during the double-blind investigational 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) and BMI and weight 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, including changes from baseline. If multiple ECG variable values are available in one visit, use the mean value.

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

The QTc interval will be summarized using frequency tables for each treatment 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 treatment visit.

6.5.5 Endometrial Health Assessment

Ultrasound results (endometrial thickness) will be summarized 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. For change from baseline in endometrial thickness, an ANCOVA model, with treatment and strata smoking status (current vs former/never) as factors, with baseline weight and baseline value as covariates, will be used for treatment comparisons.

The descriptive reasons for not evaluable biopsy will be summarized. It includes reasons of not done from eCRF page, reasons of not done from lab and pathologiest readings.

For all postbaseline biopsies, the concordance of 2 of the 3 pathologists reviews will be accepted as the final diagnosis. If there is no agreement among the three pathologists, the most severe diagnosis will then be used as the final diagnosis.

The satisfactory endometrial biopsy results will be summarized for the endometrial health set (EH set). EH set consists of all subjects who are randomized and receive at least one dose of study drug, had the postbaseline biopsy done within 30 days after the last dose of study drug, and

- had an acceptable biopsy at Baseline (at least one endometrial biopsy with satisfactory tissue and no read of hyperplasia, disordered proliferative pattern or malignant); and
- had an satisfactory endometrial biopsy result after or on Day 326 or had a postbaseline final diagnosis of hyperplasia, disordered proliferative or malignant prior to Day 326.

Baseline biopsy is the last non-missing value on or prior to first dose of study drug. When more than one biopsy results are present at postbaseline, if no final diagnosis of hyperplasia, disordered proliferative pattern or malignant, the last non-missing satisfactory result within the window will be counted; otherwise, the more severe result will be counted.

For percentage of subjects with final diagnosis of endometrial hyperplasia, percentage of subjects with endometrial cancer, and percentage of subjects with disordered proliferative endometrium, the exact (Clopper-Pearson) upper one sided 95% confidence interval will be provided.

6.5.6 Bone Mass

Dual-energy X-ray absorptiometry (BMD, TBS) data will be summarized 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. The records marked as poor image quality will be excluded from the analysis. Change from baseline in bone mass density (BMD) and trabecular bone score (TBS) at hip and spine at 12 months will be derived for separate locations. An ANCOVA model, with treatment and smoking status (current vs former/never) as factors, with baseline weight and baseline value as covariates, will be used for treatment comparisons.

A sensitivity analysis will be performed by including patients with baseline but missing any post-baseline value, assuming no changes (the change score to be 0).

6.5.7 Other Safety-Related Assessments

C-SSRS

C-SSRS (including suicidal ideation and suicidal behavior) results at each visit including baseline will be summarized by treatment group.

Imaging

A mammogram at week 52/EOT/ED will be conducted if it coincides with the regularly scheduled routine screening mammogram of the patient, in accordance with local medical practice guidelines and the patient's primary care physician. Mammogram results will only be listed, if applicable.

6.5.8 Subgroups of Interest for safety

The following factors are medically important subgroups for safety analysis: Race, smoking status, BMI category, NAFLD, and NASH. The data are being analyzed by subgroup are as follows:

- Age (< 55 years, ≥ 55 years): Treatment emergent AE (TEAE) table by SOC and PT, and Categorical Liver function table
- Race 1 (White, Non-White): Treatment emergent AE (TEAE) table by SOC and PT, and Categorical Liver function table
- Race 2 (White, Black, Asian, Other): Treatment emergent AE (TEAE) table by SOC and PT, and Categorical Liver function table
- Race 3 (Black, Non-Black): Treatment emergent AE (TEAE) table by SOC and PT, and Categorical Liver function table
- Geographical region (Europe, North American): Treatment emergent AE (TEAE) table by SOC and PT, and Categorical Liver function table

- Smoking Status(current, former/never, per substance use – tabacco history CRF page): Treatment emergent AE (TEAE) table by SOC and PT, and Categorical Liver function table
- BMI category (>=18.5 kg/m^2 to <25 kg/m^2, >=25 kg/m^2 to <30 kg/m^2, >=30 kg/m^2) : Treatment emergent AE (TEAE) table by SOC and PT, and Categorical Liver function table
- Isolated non-alcoholic fatty liver disease (NAFLD) (Yes, No, per targeted medical history CRF page): Categorical liver function table
- Non-alcoholic steatohepatitis (NASH) (Yes, No, per targeted medical history CRF page): Categorical liver function table
- Diabetic Status (Yes, No): Categorical liver function table

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 [luteinizing hormone (LH), follicle-stimulating hormone (FSH), estradiol (E2), Testosterone Total/Free, Androstenedione, dehydroepiandrosterone (DHEA), Estrone] and sex hormone-binding globulin (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 will be calculated as date of visit/assessment – first dose date +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).

Except for adverse events, for the analysis window (weeks prior to the follow-up period), the last day is always 1 day after the last day of exposure to double-blind treatment.

The end of overall treatment is defined as the last day of treatment +1 and the follow-up period 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 (day)**				
		a)	**b)**	**c)**	**d)**	**e)**
Baseline	Day 1	Last non-missing value on or prior 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 99	72 to 99	58 to 127	2 to 127	
Week 16	Day 113	100 to 127	100 to 127			
Week 20	Day 141	128 to 155	128 to 155			
Week 24	Day 169	156 to 183	156 to 183	128 to 267	128 to 267	
Week 28	Day 197	184 to 211	184 to 211			
Week 32	Day 225	212 to 239	212 to 239			
Week 36	Day 253	240 to 267	240 to 267			
Week 40	Day 281	268 to 295	268 to 295			
Week 44	Day 309	296 to 323	296 to 323			
Week 48	Day 337	324 to 351	324 to 351			
Week 52	Day 365	352 to 376	352 to 376	268 to 376	268 to 376	2 to 376
Follow-up	Day 386	Last treatment + 2 days and onwards				

a) Apply to Clinical laboratory and urinalysis
b) Apply to Vital signs
c) Apply to Blood PD/PK sample, EQ-5D-5L, MENQoL
d) Apply to C-SSRS
e) Apply to 12-lead ECG

For TVU and DXA, the week 52 analysis window is day 2 to last treatment + 30 days. For endometrial biopsy, refer to Section 6.5.5 for analysis window.

As a general rule, 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 after 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.

The Follow Up visit (Week 55) will include all data collected beyond 1 day after the last dose of the study drug. If there are more than one value, then the value that is closest to Day 386 from the last dose of study drug will be selected for the analysis. The same logic will be applied as in above for more than one value.

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

SAP Sections	Description	Rationale
Relevant Sections	Added additional statistical analysis due to COVID-19	To evaluate potential impact of COVID-19
2.1, 5.1, 5.2, 5.3	Modified the primary objective and secondary objective	To be consistent with the protocol amendments (version 2.1)
2.2	Modified the sample size	To be consistent with the protocol amendments (version 2.1)
3	Modified the sample size and its rational	To be consistent with the protocol amendments (version 2.1)
5.3, 6.4.3	Added two exploratory endpoints (change of baseline in serum concentrations and plasma concentration) and its statistical analysis.	These two endpoints were missed in the previous version.
6.3	Modified the categories of drug duration and cumulative duration	To make it one year
6.4.3	Added analysis method for EQ-5D domain scores	EQ-5D domain scores are categorical endpoint. Statistical analysis method for continuous endpoints is not applicable.
6.5	Removed the significance testing for dichotomized endpoints	To be consistent with the protocol amendments (version 3.0)
6.5.1	Added exposure adjusted TEAE tables	To take the differences in subjects' duration and potential differential dropout rates between treatment groups into consideration
6.5.1	Modified list of AESIs	To be consistent with the protocol amendments (version 2.1)
6.5.2.1	Added one category	To reflect CIOMS and FDA guidance
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
6.5.5	Updated the biopsy analysis window	To make the window longer.
6.5.6	Modified the analysis to exclude the records marked as poor image quality.	To reflect the true BMD changes but rather artificial ones.
6.5.8	Added subgroup on Age (<55, ≥55), BMI cut-off 30 kg/m2, Geographical region (Europe, North American) and Diabetic Status (Yes, No)	To add medically meaningful subgroups

8 REFERENCES

- 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.
- ICH Harmonized Tripartite Guideline E 3. Structure and Content of Clinical Study Reports, November 1995. (www.ich.org; Guidelines)
- ICH Harmonized Tripartite Guideline E 9. Statistical Principles for Clinical Trials, February 1998. (www.ich.org; Guidelines)
- 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.
- 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, subjects were randomized into the study before Astellas paused recruitment in all the clinical studies on March 30, 2020. After Astellas lifted recruitment pause, the rest of the planned subjects were randomized into the study.

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
6.5.1	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.

Appendix 2: Signatures

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

PPD PhD
PPD
Date (DD Mmm YYYY)

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

PPD PhD
PPD
Date (DD Mmm YYYY)

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

PPD MD, PhD
PPD
Date (DD Mmm YYYY)