CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

216578Orig1s000

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

NEW DRUG APPLICATION

NDA/BLA #:	216578
Drug Name:	Fezolinetant tablet 30 mg and 45 mg
Indication(s):	Treatment of moderate to severe vasomotor symptoms (VMS) associated with menopause
Applicant:	Astellas Pharma US, Inc.
Clinical Outcome Assessment (COA) Type:	Patient-reported Outcome (PRO)
Date(s):	Submission Date: 06/22/2022 PDUFA Due Date: 02/23/2023
Review Priority:	Priority
Biometrics Division:	Division of Biometrics IV (DB IV)
COA Statistical Review Team:	Yun Tang, PhD Lili Garrard, PhD (Concurring Reviewer, DB III)
Statistical Team:	Juan C. Vivar, PhD Solomon Chefo, PhD (Acting Team Leader)
Division of Clinical Outcome Assessment (DCOA):	Jing (Julia) Ju, PharmD, PhD Selena Daniels, PharmD, PhD (Team Leader)
Medical Division:	Division of Urology, Obstetrics and Gynecology (DUOG)
Clinical Team:	Theresa Van Der Vlugt, MD Shelley Slaughter, MD, PhD (Team Leader)
Project Manager:	Samantha Bell

Keywords: NDA Review, Patient-reported Outcome (PRO), Meaningful Change

Table of Contents

LIST OF TABLES

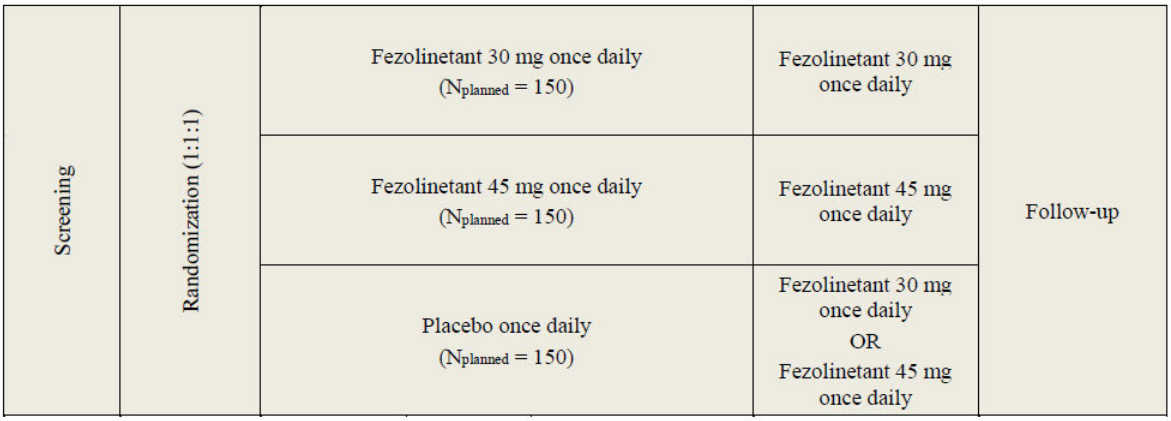
LIST OF FIGURES

LIST OF APPENDICES

1 Background

On June 22, 2022, the Applicant (Astellas) submitted an original New Drug Application (NDA 216578) to seek approval of fezolinetant for treatment of moderate to severe vasomotor symptoms (VMS) associated with menopause. Primary evidence for the efficacy of fezolinetant in the treatment of moderate to severe VMS associated with menopause came from two pivotal Phase 3 studies with identical designs, Studies 2693-CL-0301 (hereafter referred to as Study 301) and 2693-CL-0302 (hereafter referred to as Study 302). Both Studies 301 and 302 were randomized studies with a 12-week, double-blind, placebo-controlled period followed by a 40-week active treatment extension period (i.e., double-blind extension period without a placebo control), for a total of 52 weeks of treatment. Female participants (40 to 60 years of age) in both studies were randomized in a 1:1:1 ratio to receive placebo, fezolinetant 30 mg or fezolinetant 45 mg and stratified by smoking status (current smoker or former/never smoker). After completing 12 weeks of treatment, participants on placebo were re-randomized in a 1:1 ratio to receive fezolinetant 30 mg or 45 mg for another 40 weeks. Participants who completed fezolinetant 30 mg or 45 mg in the 12-week double-blind period continued to receive fezolinetant 30 mg or 45 mg during the active treatment extension period. See Figure 1 below for the study schema for both studies.

Figure 1. Study Schema for Studies 301 and 302



N: number; V: visit.

† Screening was to be performed up to 35 days prior to randomization, with a minimum of 10 days to allow for baseline data collection of VMS frequency and severity.

‡ Refer to the schedule of assessments for visits 2b and 5b.

Source: Figure 1 of Clinical Study Reports for Study 301 and Study 302

Refer to the clinical review by Dr. Theresa Van Der Vlugt for more details regarding the study design for Studies 301 and 302.

The four co-primary efficacy endpoints for Studies 301 and 302 were defined as:

- 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

The Applicant has demonstrated that participants treated with fezolinetant 30 mg and 45 mg in both studies had a statistically significant reduction from baseline to Weeks 4 and 12 in both the frequency and the severity of moderate to severe VMS relative to placebo. Specifically, the mean differences between fezolinetant treatment groups and placebo in the reduction for daily VMS frequency for both studies were observed as follows:

- Study 301: 1.87 for 30 mg and 2.07 for 45 mg at Week 4, and 2.39 for 30 mg and 2.55 for 45 mg at Week 12
- Study 302: 1.82 for 30 mg and 2.55 for 45 mg at Week 4, and 1.86 for 30 mg and 2.53 for 45 mg at Week 12

Refer to the statistical review by Dr. Juan Vivar for a detailed efficacy evaluation of fezolinetant 30 mg and 45 mg in Studies 301 and 302.

Based on the prior communications between the Applicant and FDA (see Table 1 below), the Applicant and FDA reached an agreement that to obtain an indication for the treatment of moderate to severe VMS due to menopause, fezolinetant in each phase 3 study should demonstrate (1) a statistically significant decrease from baseline in the frequency and severity of moderate to severe VMS at Weeks 4 and 12 for fezolinetant compared to placebo **and** (2) a clinically meaningful reduction in VMS frequency (i.e., the magnitude of the reduction from baseline in frequency of moderate to severe VMS at Weeks 4 and 12 exceed that of placebo at these timepoints by at least 14 moderate to severe VMS per week or 2 moderate to severe VMS per day ["14/2 Concept"]). Furthermore, the Applicant also conducted additional *exploratory* analyses to derive the clinically meaningful within-patient change thresholds (MCTs) for VMS frequency at Weeks 4 and 12. These exploratory analyses included receiver operating characteristics (ROC) method, anchor-based methods supplemented with both empirical cumulative distribution function (eCDF) curves and probability density function (PDF) curves[1], and the half-standard deviation (SD) method (i.e., half SD at baseline, which was used to define the minimally important difference). Applicant stated in the NDA submission that their prespecified primary assessment of clinically meaningful reduction in VMS frequency would be based on the "14/2 Concept". Analyses such as anchor-based analyses, ROC analyses, and the half-SD method were exploratory only and would not be used as alternative criteria for determining clinically meaningful reduction in VMS frequency. Table 1 below summarizes the

[1]Anchor-based methods supplemented with both eCDF and PDF curves are the recommended primary method to derive MCTs according to the FDA Patient-Focused Drug Development (PFDD) Guidance 4 Discussion document (https://www fda.gov/media/132505/download).

regulatory interactions pertinent to the approaches used to determine the clinically meaningful reduction in VMS frequency during the IND phase.

Table 1. Summary of Regulatory Interactions Between the Applicant and FDA Pertinent to Approaches Used to Determine the Clinically Meaningful Reduction in VMS Frequency

Regulatory Activity	Key Issues/Feedback
FDA Advice/Information Request (IR) letter dated September 20, 2019	• FDA recommended that products intended to treat moderate to severe VMS should demonstrate both a clinically meaningful and statistically significant reduction in the frequency and statistically significant reduction in the severity of VMS in the treatment groups compared with placebo with the reduction occurring within 4 weeks of initiation and maintained through 12 weeks of treatment. • FDA stated that a clinically meaningful reduction in frequency was identified as a reduction of at least two moderate to severe VMS per day or 14 moderate to severe VMS per week above placebo at Week 4 and Week 12 ("14/2 Concept"). • FDA suggested that the Applicant may propose alternative approaches to determine the clinical meaningfulness of the statistical change in VMS frequency, for example anchor-based methods using fit-for-purpose PRO instruments.
FDA Type C Meeting written response only (WRO) letter dated September 18, 2020	• FDA disagreed with the Applicant's proposal as described in the Type C meeting package dated July 21, 2020 (b) (4) • (b) (4) • FDA indicated that the Applicant may propose within-patient meaningful change analyses as the primary analysis to demonstrate clinical meaningfulness in lieu of the 14/2 Concept.
Applicant's psychometric analysis plan (PAP) version 2.0 submitted on June 24, 2021	• The PAP focused on the analyses plan used to derive the MCTs for VMS frequency. • The PAP stated that the responder analyses using the MCTs would be considered as exploratory analyses and only intended as supplementary evidence of efficacy.

Source: FDA reviewer's table

The Applicant's efficacy results have shown that fezolinetant 45 mg achieved a reduction of at least 2 moderate to severe VMS per day at both Week 4 and Week 12 in both Studies 301 and 302, whereas fezolinetant 30 mg achieved that only at Week 12 in Study 301. Using the "14/2 Concept" as the primary assessment of clinically meaningful reduction in VMS frequency, the Applicant concluded that fezolinetant 45 mg demonstrated a clinically meaningful reduction in the VMS frequency, and fezolinetant 30 mg failed to demonstrate a clinically meaningful reduction in the VMS frequency. As such, NDA 217578 requested for the labeling of the 45 mg dose and not the 35 mg dose.

This statistical review is provided to a consult received from the Clinical and DCOA teams regarding the clinical meaningfulness of fezolinetant 30 mg and 45 mg in the reduction of VMS

frequency in Studies 301 and 302 from the patients' perspective based on the anchor-based analyses (i.e., the prespecified exploratory analyses). This statistical review is based on the study protocol, statistical analysis plan, psychometric analysis plan, reports (including a final Psychometric Analysis Report), data, and Applicant's responses to FDA's Information Requests located in the Electronic Document Room at \\CDSESUB1\evsprod\NDA216578 under submission dates of June 22, 2022 and September 6, 2022.

2 Clinical Meaningfulness of VMS Frequency in Studies 301 and 302

2.1 VMS frequency

The frequency of VMS was collected using the VMS electronic diary. The baseline VMS frequency is calculated by averaging the daily counts of moderate and severe VMS events across a 10-day period prior to randomization. The average VMS frequency for Week 4 and Week 12 is calculated based on a 7-day period prior to the visit at Week 4 and Week 12, respectively. A negative change from baseline on VMS frequency indicates VMS frequency reduction. See Appendix A for a copy of VMS electronic diary.

2.2 Anchor scale

The 7-category Patient Global Impression of Change for Vasomotor Symptoms (PGIC-VMS) was the only anchor scale administered in Studies 301 and 302. Therefore, the Applicant's anchor-based analysis was based solely on PGIC-VMS. The PGIC-VMS asks the following: "Compared to the beginning of this study, how would you rate your hot flushes/night sweats now?" The response categories are as follows: Much Better, Moderately Better, A Little Better, No Change, A Little Worse, Moderately Worse, and Much Worse. See Appendix B for a copy of PGIC-VMS.

According to the Applicant's final Psychometric Analysis Report submitted as part of the NDA, the response category of "Moderately Better" on the PGIC-VMS was proposed as the target anchor response category, i.e., the anchor response category that represents clinically meaningful change to patients. Upon reviewing the information provided, the DCOA review team determined that there was a lack of qualitative evidence to support that the "Moderately Better" response category on the PGIC-VMS represents a meaningful change to participants. However, based on discussions with the DCOA team, using "Moderately Better" as the target anchor response category was considered as a conservate approach and reasonable to support further anchor-based analyses.

Refer to the DCOA review for a detailed discussion on PGIC-VMS used in the anchor-based analyses.

2.3 Methods used to derive MCTs

During the IND phase, FDA informed the Applicant that (1) anchor-based methods supplemented with eCDF and PDF curves would be the primary method to derive MCTs or a

range of MCTs, and (2) distributional-based methods such as ROC can only be supportive (see Type C Meeting WRO letters dated September 18, 2020 and July 28, 2021). In the Psychometric Analysis Plan version 3.0 submitted on March 3, 2022, the Applicant proposed to determine the MCTs using the anchor-based mean (primary) and median (supportive) change in the VMS frequency for patients who chose "Moderately Better" on PGIC-VMS; additional supportive MCT estimates included those from ROC analyses and the half-SD method. Regarding this proposal, FDA provided the following comment in the FDA Advice/IR letter dated May 24, 2022:

> *We do not agree with your proposal that "The primary threshold estimate will be determined using the descriptive mean change in the frequency of moderate/severe hot flashes for patients who had a moderate improvement on the PGIC-VMS measure (i.e., "moderately better"). The median for this group will serve as a supportive estimate.' Anchor-based methods supplemented with both empirical cumulative distribution function (eCDF) and probability density function (PDF) curves should be the primary method to derive MCTs (Refer to FDA Patient-Focused Drug Development (PFDD) Guidance 4 Discussion document: https://www.fda.gov/media/132505/download).*

However, the MCT estimates reported in Applicant's final Psychometric Analysis Report were still determined using the anchor-based mean change in the VMS frequency for participants who chose "Moderately Better" on PGIC-VMS. The median change for this group of participants were still used a supportive estimate of the threshold in addition to the supportive MCT estimates from ROC analyses and the half-SD method.

Of note, during the IND phase, FDA also recommended the Applicant the use of qualitative methods such as exit interviews to help inform the MCTs for the VMS frequency due to the limitations of the PGIC-VMS anchor scale. However, the Applicant did not follow this recommendation. Refer to the DCOA review for more details.

2.4 Anchor-based analysis

2.4.1 Anchor-based analysis population

In the NDA submission, the Applicant used blinded data from Studies 301 and 302 for their anchor-based analyses. Specifically, the primary analysis population for their anchor-based analyses consisted of full analysis set (FAS[2]) participants in Studies 301 and/or 302 who had no missing PGIC-VMS and VMS frequency at both baseline and Week 4 or Week 12 (i.e., completers population). In addition, subsamples 1, 2, 3 and 4 were used to perform sensitivity analyses. The specific sampling rules used to identify the overall and subsamples separately by study and pooled across two studies are described in Table 2 below. Depending on the analysis samples (either from individual study or pooled across the two studies), two types of MCTs for

[2] Per protocols for Studies 301 and 302, FAS is defined as all participants who were randomized and received at least one dose of study intervention.

Week 4 and Week 12 were reported in the final Psychometric Analysis Report, i.e., individual MCTs for each study separately and a common MCT that would be applied to both studies.

FDA agreed with the Applicant's primary analysis population (i.e., completers population in the pooled data from Studies 301 and 302 across treatment arms) to determine the MCT for VMS frequency (N = 909 for Week 4 and N = 796 for Week 12). FDA also agreed with the Applicant's proposal to use subsamples 3 and 4 to perform sensitivity analyses. Given the identical designs for Studies 301 and 302, FDA recommended the Applicant to derive a common MCT or a common range of MCTs, that would be applied to both studies as conveyed in the FDA Advice/IR letter dated May 24, 2022. Therefore, FDA's anchor-based analyses derived a common MCT that applied to both studies, one for Week 4 and one for Week 12.

Table 2. Description of the Analysis Samples

Sampling Rule	Study 301	Study 302	Pooled (Studies 301 and 302)
Overall (primary analysis sample to determine the MCT)	Completers population in the Study 301 FAS	Completers population in the Study 302 FAS	Completers population in the FAS
Subsample 1	50% random sample of completers population in the Study 301 FAS without replacement stratified by response of "Moderately Better" vs. the remaining response categories for PGIC-VMS	50% random sample of completers population in the Study 302 FAS without replacement stratified by response of "Moderately Better" vs. the remaining response categories for PGIC-VMS	Pooled Subsample 1 for each study
Subsample 2	Remaining 50% (those not included in Subsample 1)	Remaining 50% (those not included in Subsample 1)	Pooled Subsample 2 for each study
Subsample 3	50% random sample of completers population in the Study 301 FAS without replacement	50% random sample of completers population in the Study 302 FAS without replacement	Pooled Subsample 3 for each study
Subsample 4	Remaining 50% (those not included in Subsample 3)	Remaining 50% (those not included in Subsample 3)	Pooled Subsample 4 for each study

Source: FDA reviewer's table; adapted from Table 1 of Psychometric Analysis Report
Abbreviations: FAS, full analysis set; MCT, meaningful within-patient change threshold; PGIC-VMS, Patient Global Impression of Change (Vasomotor Symptoms)
Note: FAS is defined as all participants who were randomized and received at least one dose of study intervention. "Completers population" is defined as FAS participants in Studies 301 or 302 who had no missing PGIC-VMS and VMS frequency at both baseline and Week 4 or Week 12.

2.4.2 Anchor-based analysis results

FDA replicated the Applicant's analyses and therefore this section only presents FDA's analysis results. Table 3 below shows the distribution of VMS frequency at baseline, Weeks 4 and 12, and the change in VMS frequency from baseline to Weeks 4 or 12 by treatment arm and by study.

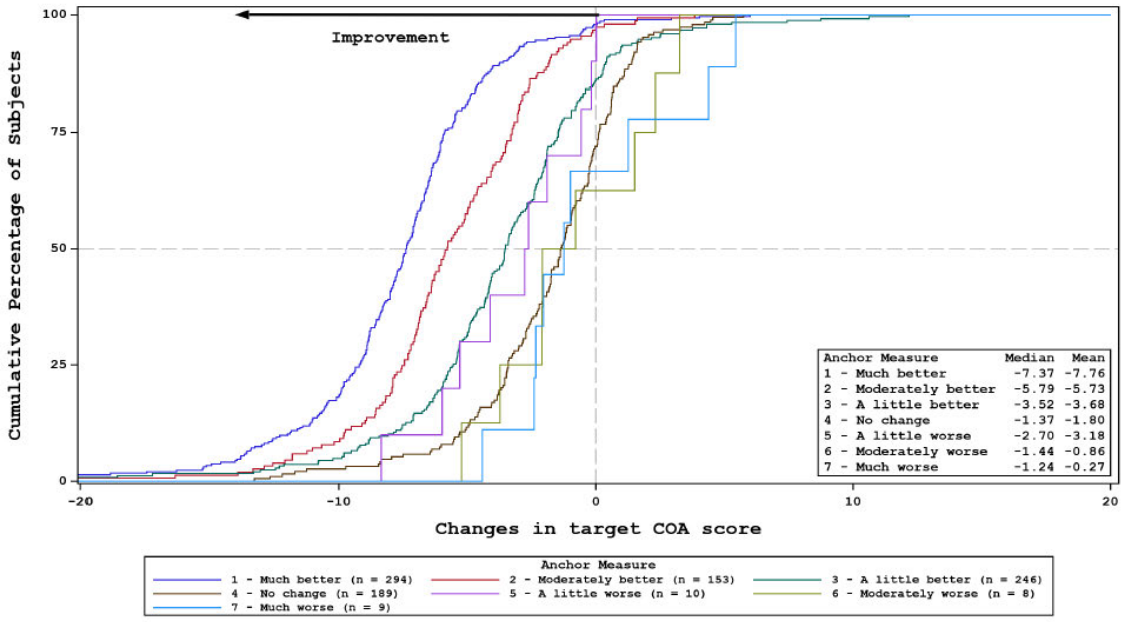
Table 3. Distribution of VMS Frequency at Baseline, Weeks 4 and 12 and Change From Baseline in VMS Frequency by Study and Treatment Arm (FAS)

	Statistic	Study 301			Study 302		
		Placebo (N = 175)	Fezolinetant 30 mg (N = 173)	Fezolinetant 45 mg (N = 174)	Placebo (N = 167)	Fezolinetant 30 mg (N = 166)	Fezolinetant 45 mg (N = 167)
Baseline	Min /Max	7.0 / 31.2	3.4 / 54.4	7.0 / 37.0	5.6 / 40.7	2.5 / 54.1	7.0 / 91.1
	Median	9.40	9.60	9.28	10.10	10.15	9.90
	[IQR]	[8.30, 11.80]	[8.19, 11.50]	[7.90, 11.80]	[8.60, 12.50]	[8.60, 12.30]	[8.60, 12.80]
	Mean (SD)	10.51 (3.79)	10.65 (4.73)	10.44 (3.92)	11.59 (5.02)	11.23 (4.88)	11.79 (8.26)
	n	175	173	174	167	166	167
	Missing	0	0	0	0	0	0
Week 4	Min /Max	0.0 / 27.2	0.0 / 14.9	0.0 / 31.6	0.0 / 48.7	0.0 / 49.1	0.0 / 68.7
	Median	7.50	5.00	4.27	7.29	4.17	4.14
	[IQR]	[4.43, 9.29]	[2.14, 7.57]	[2.29, 7.38]	[4.00, 10.43]	[1.50, 8.29]	[1.14, 7.86]
	Mean (SD)	7.25 (4.29)	5.36 (3.76)	5.20 (4.48)	8.08 (6.50)	5.79 (6.02)	5.67 (7.29)
	n	166	157	164	151	155	155
	Missing	9	16	10	16	11	12
	Change from Baseline*						
	Min /Max	-28.1 / 5.4	-52.3 / 4.2	-22.5 / 8.7	-18.6 / 12.2	-21.9 / 5.7	-28.8 / 7.4
	Median	-2.59	-5.10	-5.18	-3.44	-5.89	-6.34
	[IQR]	[-5.41, -0.57]	[-7.83, -1.99]	[-7.36, -3.01]	[-5.97, -1.24]	[-8.09, -2.89]	[-8.60, -3.16]
	Mean (SD)	-3.27 (4.18)	-5.35 (5.57)	-5.20 (4.07)	-3.64 (4.15)	-5.52 (4.23)	-6.24 (4.78)
	n	166	157	164	151	155	155
	Missing	9	16	10	16	11	12
Week 12	Min /Max	0.0 / 29.3	0.0 / 17.1	0.0 / 23.7	0.0 / 64.0	0.0 / 44.0	0.0 / 33.3
	Median	6.43	3.80	3.29	5.00	3.29	2.83
	[IQR]	[3.50, 9.17]	[1.29, 7.00]	[1.20, 6.29]	[2.41, 8.38]	[1.14, 7.00]	[0.57, 6.14]
	Mean (SD)	6.85 (4.66)	4.46 (3.72)	4.06 (3.85)	6.73 (7.58)	4.80 (5.59)	4.49 (5.39)
	n	139	131	146	140	133	145
	Missing	36	42	28	27	33	22
	Change from Baseline*						
	Min /Max	-14.9 / 8.6	-51.1 / 5.0	-34.7 / 5.6	-19.9 / 25.9	-25.7 / 6.4	-57.8 / 8.4
	Median	-3.29	-5.83	-6.70	-5.11	-6.53	-7.00
	[IQR]	[-6.29, -0.97]	[-8.77, -3.00]	[-8.51, -3.90]	[-7.36, -2.81]	[-8.40, -4.11]	[-9.20, -4.69]
	Mean (SD)	-3.67 (4.18)	-6.44 (6.15)	-6.38 (4.48)	-4.57 (5.14)	-6.43 (4.77)	-7.43 (6.47)
	n	139	131	146	140	133	145
	Missing	36	42	28	27	33	22

Source: Table 11 of Clinical Study Reports for Study 301 and Study 302; adapted and verified by the FDA reviewer
Abbreviations: FAS, fall analysis set; IQR, interquartile range; VMS, vasomotor symptoms
*A negative change indicates a reduction/improvement from baseline (i.e., a favorable outcome)

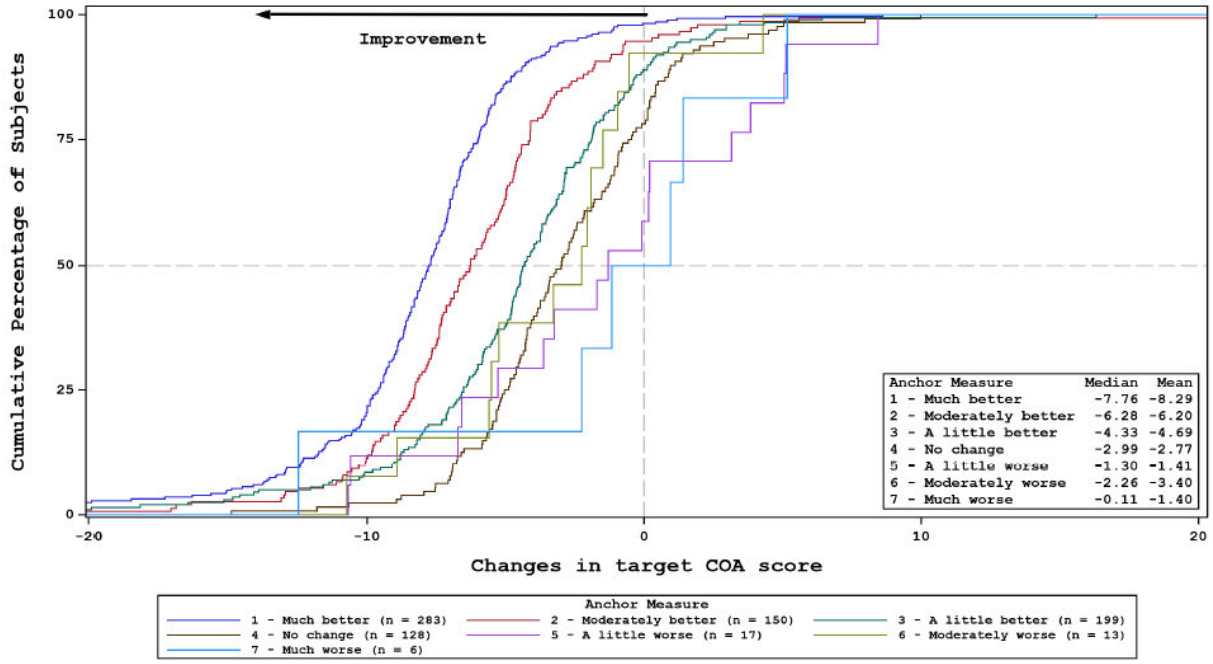
Figure 2 and Figure 3 below show the eCDF plot of change in VMS frequency at Week 4 and Week 12 from baseline by PGIC-VMS response category based on the overall sample, respectively. Based on visual inspection, both eCDF plots show a clear separation between "Moderately Better" and "A Little Better" or "No Change" categories across the entire range of scores. This provides some quantitative evidence to support FDA's decision to take a conservative approach by using "Moderately Better" as the anchor response category that represents a meaningful change to participants. The eCDF plots for subsamples 3 and 4 were similar to those for the overall sample and therefor are not presented in this review.

Figure 2. eCDF, Change in VMS Frequency from Baseline to Week 4 by PGIC-VMS Category, Pooled Overall Sample from Studies 301 and 302 (N = 909)



Source: Figure 11 of Psychometric Analysis Report; verified by the FDA reviewer
Abbreviations: eCDF, empirical cumulative distribution function; PGIC-VMS, Patient Global Impression of Change (Vasomotor Symptoms)
Note: "Overall sample" refers to participants pooled from the full analysis set (FAS) in Studies 301 and 302 who had no missing PGIC-VMS and VMS frequency at both baseline and Week 4. FAS is defined as all participants who were randomized and received at least one dose of study intervention in Studies 301 or 302.

Figure 3. eCDF, Change in VMS Frequency from Baseline to Week 12 by PGIC-VMS Category, Pooled Overall Sample from Studies 301 and 302 (N = 796)



Source: Figure 13 of Psychometric Analysis Report; verified by the FDA reviewer
Abbreviations: eCDF, empirical cumulative distribution function; PGIC-VMS, Patient Global Impression of Change (Vasomotor Symptoms)
Note: "Overall sample" refers to participants pooled from the full analysis set (FAS) in Studies 301 and 302 who had no missing PGIC-VMS and VMS frequency at both baseline and Week 12. FAS is defined as all participants who were randomized and received at least one dose of study intervention in Studies 301 or 302.

The FDA- and Applicant-proposed MCT estimates based on the PGIC-VMS anchor using the overall sample and subsamples 3 and 4 are provided in Table 4 below. FDA identified the MCT estimates using the 50th percentile from the eCDF curve of the "Moderately Better" anchor response category. In other words, the *median* change in the VMS frequency for participants who reported "Moderately Better" on PGIC-VMS was proposed by FDA as the MCTs for VMS frequency: -5.79 for Week 4 and -6.28 for Week 12. In contrast, the *mean* change in the VMS frequency for participants who reported "Moderately Better" on PGIC-VMS was proposed by the Applicant as the MCTs for VMS frequency: - 5.73 for Week 4 and -6.20 for Week 12.

Table 4. MCTs for VMS Frequency Using PGIC-VMS Derived from the Overall Sample and Subsamples 3 and 4, Pooled from Studies 301 and 302, FAS

Analysis Visit	Overall Sample (primary)	Subsample 3 (sensitivity)	Subsample 4 (sensitivity)
Week 4			
N	909	462	447
MCT by FDA	-5.79	-6.29	-5.44
MCT by Applicant	-5.73	-5.61	-5.84
Week 12			
N	796	399	397
MCT by FDA	-6.28	-6.48	-6.05
MCT by Applicant	-6.20	-5.79	-6.59

Source: FDA reviewer's analysis, End-of-Text Table 3.1 in Psychometric Analysis Report, Tables 1.1 and 1.2 "ise-fda-request-mcid-final-tables" pdf document (Applicant's IR response dated September 22, 2022),
Abbreviations: FAS, full analysis set; MCT, meaningful within-patient change threshold; PGIC-VMS, Patient Global Impression of Change (Vasomotor Symptoms)
Note: FAS is defined as all participants who were randomized and received at least one dose of study intervention in Studies 301 or 302. "Overall sample" refers to participants pooled from the FAS in Studies 301 and 302 who had no missing PGIC-VMS and VMS frequency at both baseline and Week 4 or Week 12. Subsample 3 refers to 50% of participants randomly selected from the overall sample without replacement. Subsample 4 refers to remaining participants in the overall sample (not included in subsample 3).

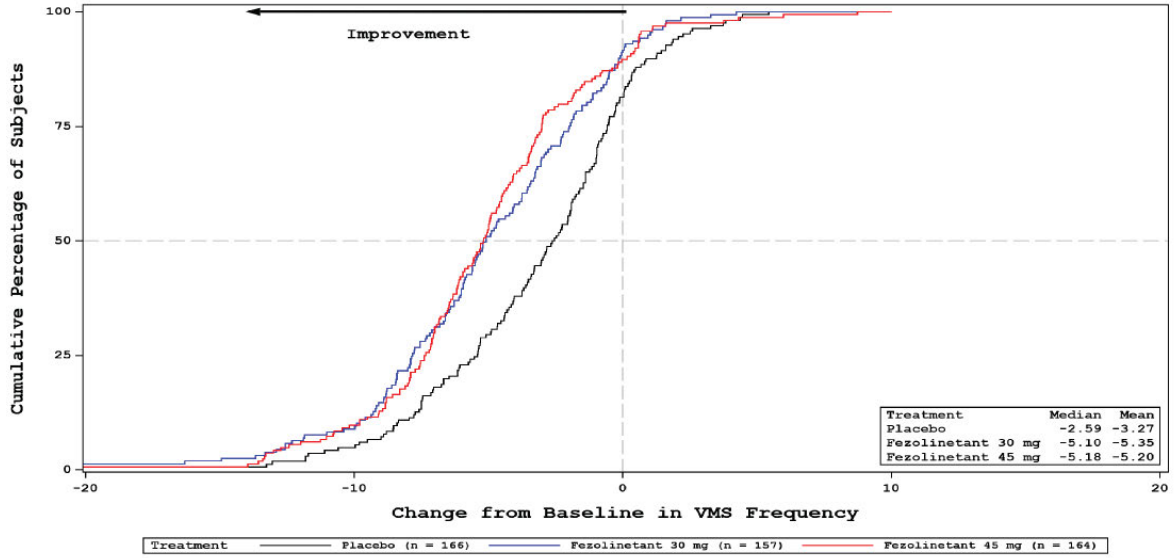
2.5 Interpretation of the clinical trial results: treatment effect on VMS frequency change

The trial results were interpreted using the identified MCTs stated in Section 2.4. The limitations of the PGIC-VMS and derivations of the thresholds need to be considered when interpreting the results. The eCDF plots of within-patient changes in VMS frequency from baseline by treatment arms for Studies 301 (see Figures 4 and 5) and 302 (see Figures 6 and 7) were examined to evaluate the clinical relevance of the observed treatment effect for VMS frequency. These figures showed a clear and consistent separation between the treatment group eCDF curves (fezolinetant 30 mg vs. placebo and fezolinetant 45 mg vs. placebo) across the entire range of score change in both studies, including both FDA- and Applicant-derived meaningful change thresholds.

To quantify the treatment effect of fezolinetant 30 mg or 45 mg compared to placebo on the proportion of participants experiencing a clinically meaningful reduction in VMS frequency, FDA examined the differences in responder rates (fezolinetant 30 mg – placebo or fezolinetant 45 mg – placebo) and corresponding 95% confidence intervals (CIs) at Weeks 4 and 12 using the FDA-derived MCTs (Tables 5 and 6). These analyses were performed on FAS by using two approaches to handle missing VMS frequency data: (1) missing as non-responder (i.e., cases with any missing data at Week 4 or Week 12 were classified as non-responders), and (2) observed cases only (i.e., cases that had VMS frequency data both at baseline and Week 4 or Week 12). As shown in Tables 5 and 6, using the FDA-derived MCTs of -5.79 for Week 4 and of -6.28 for Week 12, there were more participants on both fezolinetant 30 mg and 45 mg arms who experienced a clinically meaningful reduction in VMS frequency than on the placebo arm at both Weeks 4 and 12 for both trials, with 95% CIs for all treatment differences excluded 0.

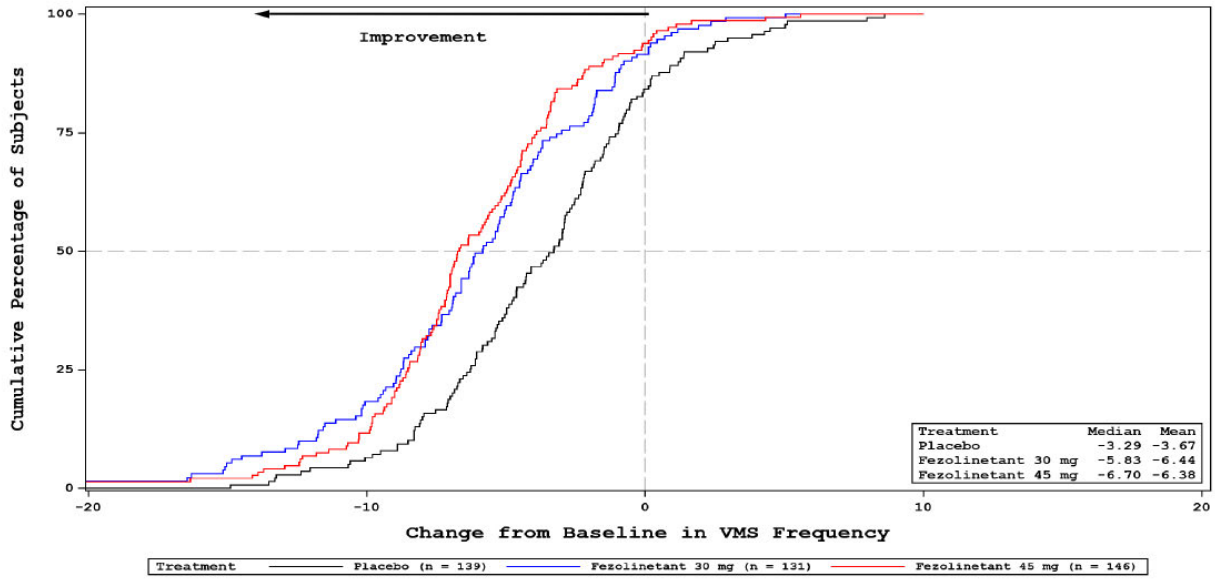
This reviewer also conducted supplementary responder analyses using the Applicant's derived-MCTs based on overall sample and subsamples 3 and 4 and FDA-derived MCTs based on subsamples 3 and 4. The results were consistent to those based on the primary FDA-derived thresholds of -5.79 for Week 4 and of -6.28 for Week 12 and are not presented in this review.

Figure 4. eCDF, Change in VMS Frequency From Baseline to Week 4 by Treatment Arms, FAS, Study 301



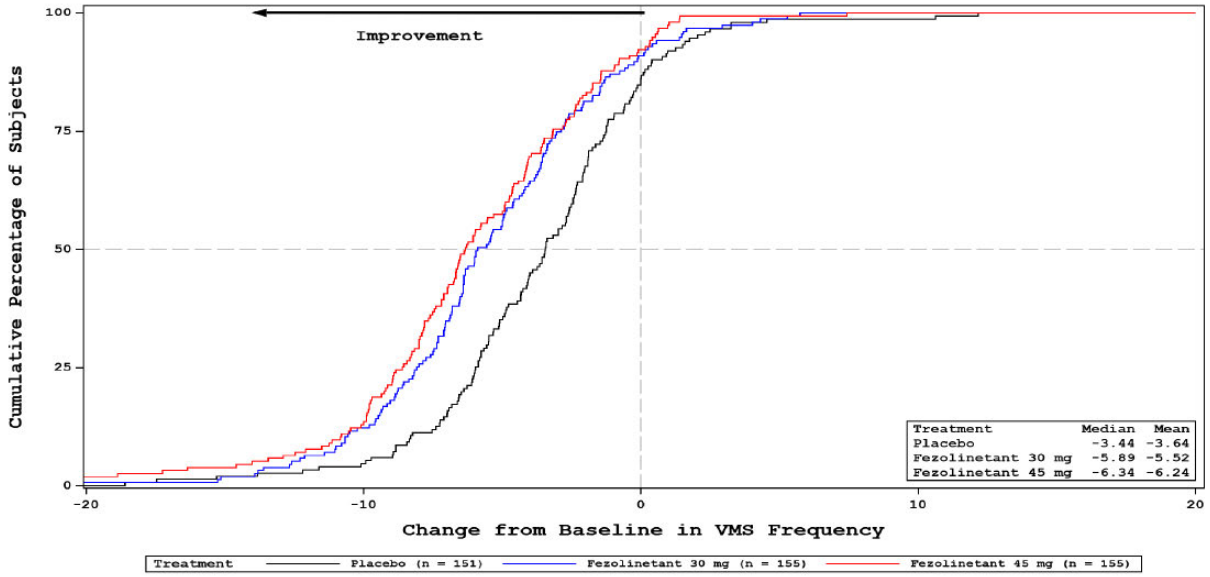
Source: Figure 8.4.6 of "ecdf-curves-vms-freq-change-scores" pdf document; verified by the FDA reviewer
Abbreviations: eCDF, empirical cumulative distribution function; FAS, full analysis set; VMS, vasomotor symptoms
Note: The figures was created based on the observed cases in the FAS (i.e., cases that had VMS frequency data at both baseline and Week 4).
FAS is defined as all participants who were randomized and received at least one dose of study intervention in Studies 301 or 302.

Figure 5. eCDF, Change in VMS Frequency From Baseline to Week 12 by Treatment Arms, FAS, Study 301



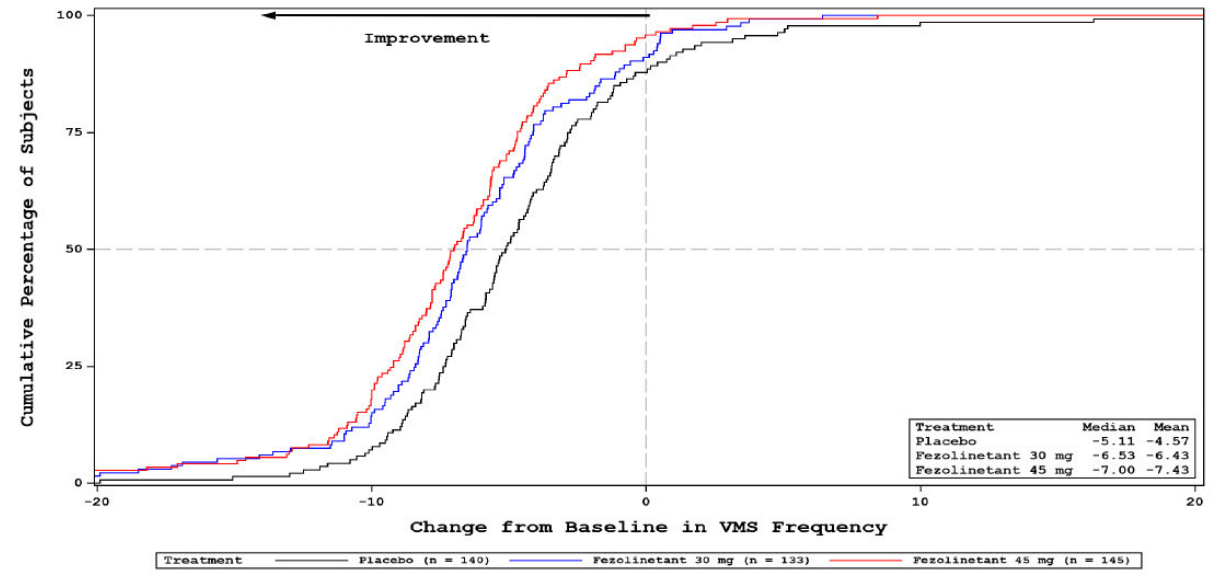
Source: Figure 8.4.7 of "ecdf-curves-vms-freq-change-scores" pdf document; verified by the FDA reviewer
Abbreviations: eCDF, empirical cumulative distribution function; FAS, full analysis set; VMS, vasomotor symptoms
Note: The figures was created based on the observed cases in the FAS (i.e., cases that had VMS frequency data at both baseline and Week 12).
FAS is defined as all participants who were randomized and received at least one dose of study intervention in Studies 301 or 302.

Figure 6. eCDF, Change in VMS Frequency From Baseline to Week 4 by Treatment Arms, FAS, Study 302



Source: Figure 8.4.8 of "ecdf-curves-vms-freq-change-scores" pdf document; verified by the FDA reviewer
Abbreviations: eCDF, empirical cumulative distribution function; FAS, full analysis set; VMS, vasomotor symptoms
Note: The figures was created based on the observed cases in the FAS (i.e., cases that had VMS frequency data at both baseline and Week 4).
FAS is defined as all participants who were randomized and received at least one dose of study intervention in Studies 301 or 302.

Figure 7. eCDF, Change in VMS Frequency From Baseline to Week 12 by Treatment Arms, FAS, Study 302



Source: Figure 8.4.9 of "ecdf-curves-vms-freq-change-scores" pdf document; verified by the FDA reviewer
Abbreviations: eCDF, empirical cumulative distribution function; FAS, full analysis set; VMS, vasomotor symptoms
Note: The figures was created based on the observed cases in the FAS (i.e., cases that had VMS frequency data at both baseline and Week 12).
FAS is defined as all participants who were randomized and received at least one dose of study intervention in Studies 301 or 302.

Table 5. Number and Percent of Female Participants With A Clinically Meaningful Reduction in VMS Frequency at Weeks 4 and 12 Based on the MCTs Derived Using the PGIC-VMS Anchor, FAS, Study 301

Data	Analysis Visit	Thresholds	Placebo (N = 175)	Fezolinetant 30 mg (N = 173)	Fezolinetant 45 mg (N = 174)	Treatment Difference (95% CI) [1] 30 mg - placebo	45 mg - placebo
Missing as Non-responder	Week 4	>= 5.79 VMS frequency reduction	38 /175 (21.7%)	67 /173 (38.7%)	72 /174 (41.4%)	17.0% (7.4%, 26.3%)	19.7% (10.0%, 28.9%)
	Week 12	>= 6.28 VMS frequency reduction	35 /175 (20.0%)	61 /173 (35.3%)	78 /174 (44.8%)	15.3% (5.9%, 24.3%)	24.8% (15.1%, 33.9%)
Observed Cases	Week 4	>= 5.79 VMS frequency reduction	38 /166 (22.9%)	67 / 157 (42.7%)	72 / 164 (43.9%)	19.8% (9.6%, 29.5%)	21.0% (11.1%, 30.6%)
	Week 12	>= 6.28 VMS frequency reduction	35 /139 (25.2%)	61 /131 (46.6%)	78 /146 (53.4%)	21.4% (10.2%, 32.4%)	28.2% (17.0%, 38.5%)

Source: FDA reviewer's analysis

Abbreviations: CI, confidence interval; CMH, Cochran-Mantel-Haenszel; FAS, full analysis set; MCT, meaningful change threshold

[1]Treatment difference was the adjusted difference based on the CMH test stratified by smoking status. The 95% CI used Newcombe method. The magnitudes of 95% CI calculated based on normal approximation were similar to their CHM-adjusted counterparts.

Note: "Missing as non-responder" refers to the situation where cases with any missing data at Week 4 or Week 12 were classified as non-responders. "Observed cases" refer to cases that had VMS frequency data both at baseline and Week 4 or Week 12.

Table 6. Number and Percent of Female Participants With A Clinically Meaningful Reduction in VMS Frequency at Weeks 4 and 12 Based on the MCTs Derived Using the PGIC-VMS Anchor, FAS, Study 302

Data	Analysis Visit	Thresholds	Placebo (N = 175) n (%)	Fezolinetant 30 mg (N = 173) n (%)	Fezolinetant 45 mg (N = 174) n (%)	Treatment Difference (95% CI) [1] 30 mg - placebo	Treatment Difference (95% CI) [1] 45 mg - placebo
Missing as Non-responder	Week 4	>= 5.79 VMS frequency reduction	41 /175 (24.6%)	78 / 173 (47.0%)	84 /174 (50.3%)	22.4% (12.2%, 32.1%)	25.7% (15.5%, 35.3%)
	Week 12	>= 6.28 VMS frequency reduction	52 /175 (31.1%)	70 /173 (42.2%)	81 / 174 (48.5%)	11.0% (0.7%, 21.1%)	17.4% (6.9%, 27.4%)
Observed Cases	Week 4	>= 5.79 VMS frequency reduction	41/151 (27.2%)	78/155 (50.3%)	84/155 (54.2%)	23.2% (12.1%, 33.1%)	27.0% (15.9%, 36.9%)
	Week 12	>= 6.28 VMS frequency reduction	52/140 (37.1%)	70/133 (52.6%)	81/145 (55.9%)	15.5% (3.5%, 26.6%)	18.7% (7.0%, 29.5%)

Source: FDA reviewer's analysis

Abbreviations: CI, confidence interval; CMH, Cochran-Mantel-Haenszel; FAS, full analysis set; MCT, meaningful change threshold

[1]Treatment difference was the adjusted difference based on the CMH test stratified by smoking status. The 95% confidence internal (CI) used Newcombe method. The magnitudes of 95% CI calculated based on normal approximation were similar to their CHM-adjusted counterparts.

Note: "Missing as non-responder" refers to the situation where cases with any missing data at Week 4 or Week 12 were classified as non-responders. "Observed cases" refer to cases that had VMS frequency data both at baseline and Week 4 or Week 12.

3 Conclusion

In the NDA submission, in addition to the "14/2 Concept" (Applicant's primary assessment of clinically meaningful reduction in VMS frequency), the Applicant conducted additional exploratory analyses using clinically meaningful within-patient change thresholds derived from anchor-based analyses to interpret the change in the frequency of moderate to severe VMS. DCOA team has concluded that the selected primary anchor PGIC-VMS has limitations which make it difficult to interpret the results of the anchor-based analyses. Acknowledging the limitations of PGIC-VMS as the only available anchor included in the studies for the VMS frequency, we observed a clear and consistent separation between the treatment group eCDF curves (fezolinetant 30 mg vs. placebo and fezolinetant 45 mg vs. placebo) across the entire range of score change, including both FDA- and Applicant-derived meaningful change thresholds as determined by PGIC-VMS at both Week 4 and Week 12 for both phase 3 trials.

5 Page(s) have been Withheld in Full as b4 (CCI/TS) immediately following this page

---

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

YUN TANG
11/22/2022 11:37:42 AM

LILI GARRARD
11/22/2022 11:47:46 AM

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDA 216578
Supplement #:	
Drug Name:	Fezolinetant tablet (30 mg and 45mg) once daily
Indication(s):	Treatment of moderate to severe vasomotor symptoms associated with menopause.
Applicant:	Astellas Pharma US Inc.
Date(s):	Submitted: 06/22/2022 PDUFA: 02/22/2023 Review Date: 11/22/2022
Review Priority:	Priority
Biometrics Division:	Division of Biometrics IV
Statistical Reviewer:	Juan C. Vivar, Ph.D.
Concurring Reviewers:	Solomon Chefo, Ph.D. (Acting Team Leader) Daphne Lin, Ph.D. Deputy Director
Medical Division:	Division of Urology, Obstetrics and Gynecology (DUOG).
Clinical Team:	Theresa Van Der Vlugt, M.D. Clinical Reviewer Shelley Slaughter, M.D. Clinical Team Leader
Project Manager:	Samantha Bell

Keywords: NDA review, mixed model repeated measures

Table of Contents

List of Tables

List of Figures

1 EXECUTIVE SUMMARY

In this New Drug Application (NDA), the Applicant seeks approval of fezolinetant for the treatment of moderate to severe vasomotor symptoms (VMS) associated with menopause. The NDA is supported by two identically designed double-blind, randomized, placebo-controlled pivotal phase 3 studies (SKYLIGHT 1 [NCT04003155] and SKYLIGHT 2 [NCT04003142]). This statistical review assessed the adequacy of the submitted information in these studies to support the efficacy of fezolinetant.

In studies SKYLIGHT 1 and SKYLIGHT 2, respectively, a total of 527 and 501 menopausal women 40 to 65 years of age who had a minimum average of 7 moderate to severe VMS (hot flashes) per day were randomized in a 1:1:1 ratio and received fezolinetant 30 mg, fezolinetant 45 mg, or placebo. Randomization was stratified by smoking status (current smokers vs. former/never smokers).

The total treatment duration of both studies was 52-week comprising 12-week double-blinded treatment period and 40-week open label extension treatment period. During the double-blinded treatment period, all subjects received their assigned treatment once daily for 12-week. After the 12-week double-blinded treatment period, subjects in the placebo arm were re-randomized in a 1:1 ratio and received either fezolinetant 30 mg or 45 mg once daily for 40 weeks. Evidence for the efficacy of fezolinetant in the treatment of VMS associated with menopause in both studies was primarily from the 12-week, placebo-controlled period.

A total of 522 menopausal women in SKYLIGHT 1 and 500 menopausal women in SKYLIGHT 2 received at least one dose of the study drug and were evaluable for efficacy. In both studies, subjects had a mean age of 54 years (range: 40 to 65 years). Most subjects in both studies were White (81%), not Hispanic or Latino (76%), and were either former smokers or had never smoked (83%). Both studies enrolled menopausal women with prior hormone therapy use (20%), with history of oophorectomy (22%), or with history of hysterectomy (32%). Most subjects in both studies completed the 12-week double-blinded treatment period (87% in SKYLIGHT 1 and 92% in SKYLIGHT 2).

In both studies, the main efficacy evaluation to support each dose of fezolinetant for the treatment of moderate to severe hot flashes due to menopause was based on the frequency and severity of moderate-to-severe hot flashes each recorded daily by the subject via an electronic diary through week 52 (See details in Section 3.2.1). In each study, the primary efficacy was assessed based on four co-primary efficacy endpoints: change in the frequency and severity of moderate to severe vasomotor symptoms per day from baseline to week 4 and from baseline to week 12. For a given dose of fezolinetant to be considered successful, all 4 co-primary endpoints must be statistically significant and clinically meaningful.

Data from both SKYLIGHT 1 and SKYLIGHT 2 studies showed a statistically significant reduction from baseline in the frequency as well as in the severity of moderate-to-severe hot flashes per day for both doses of fezolinetant compared to placebo at week 4 and week 12 (Table 1 and Table 2). For example, subjects in the combined SKYLIGHT 1 and SKYLIGHT 2 studies who had an average of 11 moderate-to-severe hot flashes per day at baseline had a mean reduction from baseline of 5 to 6 (at week 4) and 7 to 8 (at week 12) moderate-to-severe hot

flashes per day with fezolinetant as compared with a mean reduction of about 4 hot flashes per day at weeks 4 and 12 with placebo.

Although both doses of fezolinetant showed statistically significant reduction in the frequency as well as in the severity of moderate-to-severe hot flashes per day at weeks 4 and 12, fezolinetant 30 mg did not meet the clinical meaningful treatment difference of 2 or more reduction compared to placebo in the frequency of moderate-to-severe hot flashes at week 4 in SKYLIGHT 1 and SKYLIGHT 2 and at week 12 in SKYLIGHT 2. (Table 1).

Table 1. Mean Baseline and Change from Baseline to weeks 4 and 12 for Mean Frequency of Moderate to Severe Vasomotor Symptoms per 24 Hours.
(Full Analysis Set)[1]

	Placebo	**Fezolinetant**		**Difference in LS Means (95% CI) vs. Placebo**	
		30 mg	**45 mg**	**30 mg**	**45 mg**
SKYLIGHT 1					
N	175	173	174		
Baseline (SD)	10.5 (3.79)	10.7 (4.73)	10.4 (3.92)		
week 4	-3.3 (0.29)	-5.2 (0.30)	-5.4 (0.30)	-1.9 (-2.7, -1.1)	-2.1 (-2.9, -1.3)
week 12	-3.9 (0.31)	-6.3 (0.32)	-6.4 (0.31)	-2.4 (-3.3, -1.5)	-2.6 (-3.4, -1.7)
SKYLIGHT 2					
N	167	166	167		
Baseline	11.6 (5.02)	11.2 (4.88)	11.8 (8.26)		
week 4	-3.7 (0.33)	-5.5 (0.33)	-6.3 (0.33)	-1.8 (-2.7, -0.9)	-2.6 (-3.5, -1.6)
week 12	-5.0 (0.39)	-6.8 (0.39)	-7.5 (0.39)	-1.9 (-3.0, -0.8)	-2.5 (-3.6, -1.5)
POOLED					
N	342	339	341		
Baseline	11.0 (4.46)	10.9 (4.80)	11.1 (6.45)		
week 4	-3.5 (0.22)	-5.4 (0.23)	-5.8 (0.23)	-1.9 (-2.5, -1.3)	-2.3 (-2.9, -1.7)
week 12	-4.4 (0.25)	-6.6 (0.25)	-6.9 (0.25)	-2.2 (-2.8, -1.5)	-2.5 (-3.2, -1.8)

[1] All participants who were randomized and received at least 1 dose of study intervention.
Note: Summaries at weeks 4 and 12 are Least Square (LS) Means (SE) estimated from a mixed model for repeated measures analysis of covariance. SD: Standard Deviation; SE: Standard Errors.

Table 2. Mean Baseline and Change from Baseline to weeks 4 and 12 for Mean Severity of Vasomotor Symptoms per 24 Hours.
(Full Analysis Set)[1]

	Placebo	**Fezolinetant**		**Difference (95% CI) vs. Placebo**	
		30 mg	**45 mg**	**30 mg**	**45 mg**
SKYLIGHT 1					
N	175	173	174		
Baseline	2.4 (0.35)	2.4 (0.34)	2.4 (0.35)		
week 4	-0.3 (0.04)	-0.4 (0.04)	-0.5 (0.04)	-0.2 (-0.3, -0.0)	-0.2 (-0.3, -0.1)
week 12	-0.4 (0.05)	-0.6 (0.05)	-0.6 (0.05)	-0.2 (-0.4, -0.1)	-0.2 (-0.4, -0.1)
SKYLIGHT 2					
N	167	166	167		
Baseline	2.4 (0.32)	2.4 (0.33)	2.4 (0.34)		
week 4	-0.3 (0.05)	-0.5 (0.05)	-0.6 (0.05)	-0.2 (-0.3, -0.0)	-0.3 (-0.4, -0.2)
week 12	-0.5 (0.06)	-0.6 (0.06)	-0.8 (0.06)	-0.2 (-0.3, -0.0)	-0.3 (-0.5, -0.1)

	Placebo	Fezolinetant		Difference (95% CI) vs. Placebo	
		30 mg	45 mg	30 mg	45 mg
POOLED					
N	342	339	341		
Baseline	2.4 (0.34)	2.4 (0.34)	2.4 (0.35)		
week 4	-0.3 (0.03)	-0.4 (0.03)	-0.5 (0.03)	-0.2 (-0.2, -0.1)	-0.2 (-0.3, -0.25)
week 12	-0.4 (0.04)	-0.6 (0.04)	-0.7 (0.04)	-0.2 (-0.3, -0.1)	-0.2 (-0.4, -0.1)

[1] All participants who were randomized and received at least 1 dose of study intervention.
Note: Summaries at weeks 4 and 12 are Least Square (LS) Means (SE) estimated from a mixed model for repeated measures analysis of covariance. SD: Standard Deviation; SE: Standard Errors.

In summary, based on the collective efficacy evidence from the two adequate and well controlled trials of SKYLIGHT 1 and SKYLIGHT 2 studies, the reviewer concludes that the application provided substantial evidence of efficacy of fezolinetant 45 mg tablet administered once daily for the treatment of moderate to severe vasomotor symptoms associated with menopause. Although fezolinetant 30 mg was statistically superior to placebo in the mean reduction of the frequency and severity of moderate to severe VMS from baseline at week 4 and at week 12, it did not meet the clinical meaningful treatment difference criterion of 2 or more reduction compared to placebo.

2 INTRODUCTION

2.1 Overview

The Applicant, Astellas Pharma US, Inc. seeks approval of fezolinetant for the treatment of moderate to severe VMS associated with menopause. Per the Applicant, “Fezolinetant is a small-molecule, nonhormonal, selective NK3 receptor antagonist”.

The Applicant has submitted two identically designed phase 3 clinical studies (SKYLIGHT 1 and SKYLIGHT 2) to support this indication. Table 3 presents a summary of the studies addressed in this review.

Table 3. List of studies included in analysis

Study	Phase and Design	Treatment Period	Extension Treatment Period	Follow-up Period	# Subjects per Arm	Study Population
SKYLIGHT 1 (NCT04003155) SKYLIGHT 2 (NCT04003142)	Phase 3, double-blind, randomized, multicenter, placebo-controlled	12 weeks	40 weeks	3 weeks	Randomized: - 30 mg: 176 - 45 mg: 176 Placebo: 175 Randomized: - 30 mg: 166 - 45 mg: 167 Placebo: 168	Female 40-65 years old with moderate to severe VMS associated with menopause

Source: Reviewer’s summary based on protocols.

2.2 Relevant Regulatory Correspondence

Date	Application Number/ Type of Correspondence	Description
March 22, 2017	IND 130277/ Type C Meeting	Includes FDA responses about primary efficacy endpoints in proposed phase 3 clinical studies and recommendations on the calculation of mean change in severity of hot flashes.
April 17, 2019	IND 130277/ Type B End-of-phase 2 Meeting	Discussion on the design of proposed phase 3 studies. FDA stated that secondary and exploratory objectives/endpoints (b) (4) (b) (4)
September 18, 2020	IND 130277/ Type C Meeting WRO	FDA recommended that the phase 3 trials demonstrate a statistically significant decrease from baseline in the frequency and severity of VMS and reduction of at least 2 hot flashes per day.

2.3 Data sources

The study data, clinical study reports and additional information for these studies, such as the statistical analysis plans, were submitted electronically. The submitted SAS datasets for all studies were complete and well documented. These items are provided in an electronic submission located at \\CDSESUB1\evsprod\NDA216578.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The Applicant submitted both the tabulation data and analysis data (STDM and ADaM formats) for studies SKYLIGHT 1 and SKYLIGHT 2. Datasets were complete and documented. Statistical SAS programs were submitted. All statistical analyses were carried out following the pre-specified statistical analysis plan.

3.2 Evaluation of Efficacy

The efficacy assessment of fezolinetant was based on both studies SKYLIGHT 1 and SKYLIGHT 2.

3.2.1 Study Design and Endpoints

Study Design

Both SKYLIGHT 1 and SKYLIGHT 2 were randomized, 12-week double-blind, placebo-controlled, parallel group, multicenter studies designed to assess the efficacy and safety of fezolinetant in women suffering from moderate to severe VMS associated with menopause.

In both studies, women ≥ 40 years and ≤ 65 years of age seeking treatment or relief for VMS associated with menopause who had a minimum average of 7 moderate to severe hot flashes per day (or 50 to 60 per week) within the 10 days prior to randomization were enrolled at 89 sites in seven countries (United States, Canada, United Kingdom, Spain, Poland, Czech Republic and Hungary) for SKYLIGHT 1 and 88 sites in seven countries for SKYLIGHT 2 (United States, Canada, United Kingdom, Spain, Poland, Czech Republic and Latvia).

Eligible subjects who met all the studies enrollment criteria were randomized in a 1:1:1 ratio to one of the following three treatment groups. Randomization was stratified by smoking status (current smoker vs former/never smoker).

- Treatment 1: Fezolinetant 30 mg (one 30 mg tablet and one placebo tablet) once daily
- Treatment 2: Fezolinetant 45 mg (one 15 mg tablet and one 30 mg tablet) once daily
- Treatment 3: Placebo (two tablets to match) once daily

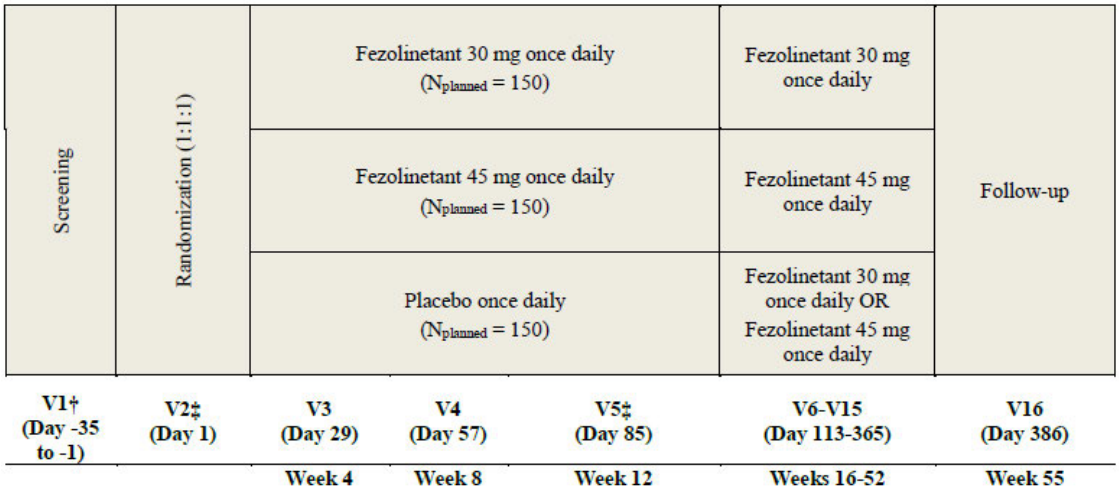
The total duration of each study (excluding a screening period of -35 to -1 days) was 55-week consisted of a 52-week treatment period and a 3-week follow-up period. During the first 12 weeks of the 55-week treatment period, subjects received the assigned treatments once daily. After completing 12 weeks of treatment, subjects in the placebo arm were re-randomized to 30 mg or 45 mg of fezolinetant in the active treatment extension period for 40 weeks of treatment through the end of study. Subjects who were in an active arm continued their assigned dose for the remaining 40-week treatment period. Following completion (or early discontinuation) of the treatment period (week 52), subjects completed an end of treatment visit and final safety follow-up visit 3 weeks after the last dose of study drug was administered (week 55). See study schema in Figure 1.

Efficacy Endpoints

The primary efficacy objective of both studies was an evaluation of superiority of each dose of fezolinetant to placebo in the following four 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

Figure 1. Study schema



Source: Figure 1 of Applicant's Clinical Study Reports.

The baseline frequency was calculated based only on the number of moderate and severe hot flashes in at least 7 of the 10 days immediately prior to randomization to meet the eligibility criterion, and the average was based on all the non-missing days. Subjects recorded daily the number (frequency) and severity of each vasomotor symptom via an electronic diary.

The severity of an individual vasomotor symptom was defined as follows:

- Mild (Mi): sensation of heat without sweating
- Moderate (Mo): sensation of heat with sweating, able to continue activity
- Severe (Se): sensation of heat with sweating, causing cessation of activity

The severity score for mild, moderate, and severe VMS was coded as 1, 2, and 3, respectively.

Severity of post-baseline co-primary severity endpoints was calculated using a weighted average over non-missing days over 7 days period defined as shown below (third column in Table 4). This calculation also includes mild VMS with an expectation that less severe VMS will be reported during study duration. Note that the calculation for severity at baseline uses a similar formula but does not include mild VMS events in the numerator or denominator and it is based on the same days as the baseline frequency. At baseline, severity was zero for any individual days on which subjects have zero moderate or severe symptoms. Severity for post-baseline individual days was zero for subjects who had zero mild, moderate, or severe vasomotor symptoms.

Table 4. Endpoint definitions

Calculation at	Frequency	Severity Score at Individual Days
Baseline	Average of the non-missing values in the 10 days before randomization	$\frac{[\#Mo/day \times 2] + [\#Se/day \times 3]}{\#(Mo + Se)/day}$
Post-baseline	Average of the non-missing values over a 7-day period	$\frac{[\#Mi/day \times 1] + [\#Mo/day \times 2] + [\#Se/day \times 3]}{\#(Mi + Mo + Se)/day}$

Source: Reviewer's summary based on protocols.
Note: For both post-baseline frequency and severity, a daily average per week was derived.

Secondary Efficacy Endpoints

Key Secondary Endpoint

The Applicant defined a key secondary efficacy endpoint in both studies as the 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.

The PROMIS SD SF 8b assessed self-reported sleep disturbance over the past seven days and included perceptions of restless sleep; satisfaction with sleep; refreshing sleep; difficulties falling sleep, getting to sleep or staying asleep; trouble sleeping; 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).

Secondary Endpoints

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

- 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

- Subject has ≥ 50% reduction from baseline to week 12 was defined as (VMS_{week12}– VMS_{BL})/ VMS_{BL} ≤ -0.5, where VMS_{week12} and VMS_{BL} were 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

- Mean change in the frequency of moderate to severe VMS from baseline to week 24
- Mean change in the Severity of moderate to severe VMS from baseline to week 24
- Score on the Patient Global Impression of Change (PGI-C) in VMS at each visit.

Reviewer's comment: (b) (4)

3.2.2 Statistical Methodologies

Analysis Sets

Full Analysis Set (FAS): Consisted of all subjects who were randomized and received at least one dose of study drug. This was the primary set for efficacy analyses. All efficacy analyses were conducted by treatment group according to the FAS.

Safety Analysis Set (SAF): Consisted of all subjects who were randomized who took at least one dose of study drug. The SAF was used for the safety analyses and the summaries of demographic and baseline characteristics.

Per Protocol Set (PPS): Consisted of the subset of subjects in the FAS who did not meet the following criteria:

- No measurement of the primary efficacy endpoint available at week 4 (week 12).
- <85% interactive diary compliance during the 4- (12-) week treatment period.
- Treatment compliance less than or equal to 85% between randomization and week 4 (week 12).

Analysis of Co-Primary Efficacy Endpoints

For each of the 4 co-primary efficacy endpoints, a mixed model repeated measures analysis of covariance (MMRM) was 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 was used to model the within-patient errors (and then Toeplitz if model did not converge). The Kenward-Roger approximation was used to estimate denominator degrees of freedom and adjust standard errors. This analysis used a restricted maximum likelihood-based repeated-measures approach.

Descriptive statistics were reported for each endpoint. Least Square (LS) estimates of mean differences from placebo for each dose and week with the associated 2-tailed 95% confidence interval (CI) were derived. The LS means were estimated using weights proportional to the percentage of women who were randomized as current smokers.

Type I Error Control (plan for multiplicity adjustment)

The Applicant proposed (b) (4)

***Reviewer's comment**: As was communicated by FDA in previous meetings with the Applicant, all the four co-primary efficacy endpoints should be tested at a significance level of 0.05.*

Handling of Missing Data

In the Applicant's primary analysis, missing data were implicitly imputed assuming a missing at random (MAR) missing data mechanism. To assess the robustness of the primary analysis results to departures from the underlying MAR assumption, the Applicant performed a discontinuation-reason based multiple imputation (MI) where data for subjects who discontinue early follow a pattern which is missing not at random. Specifically, subjects who discontinued early from the two active dose groups were multiply imputed using the placebo group. The Applicant also performed additional sensitivity analyses for the co-primary efficacy endpoints based on the per-protocol population. The data was analyzed using the same MMRM model as the primary analyses.

The key and other secondary endpoints in both studies were analyzed similarly to the analysis strategy of the co-primary efficacy endpoints.

3.2.3 Subject Disposition, Demographic and Baseline Characteristics

Subject Disposition:

Table 5 shows the summary of subject disposition and the primary reasons for study discontinuation during the 12-week double-blinded treatment period in SKYLINE 1 and 2 studies. As shown, in study SKYLIGHT 1, a total of 527 subjects were randomized. Of those, 522 took the study intervention. In SKYLIGHT 2, a total of 501 subjects were randomized. Of these, 500 took the study intervention.

Most subjects in both studies completed the 12-week double-blinded treatment period (87% in SKYLIGHT 1 and 92% in SKYLIGHT 2). The most common reason for treatment discontinuation was "withdrawal by subject" during the 12-week double-blind period (4.8% in SKYLIGHT 1 and 4.6% in SKYLIGHT 2). Due to COVID-19, 8 (1.6%) subjects in SKYLIGHT 1 and 4 (0.8%) subjects in SKYLIGHT 2 discontinued treatment from the studies.

Table 5. Summary of Subject Disposition and Reasons for Study Discontinuation

	SKYLIGHT 1			**SKYLIGHT 2**		
Category	**Placebo (N=175)**	**Fezolinetant**		**Placebo (N=167)**	**Fezolinetant**	
		30 mg (N = 174)	**45 mg (N = 173)**		**30 mg (N = 166)**	**45 mg (N = 167)**
Randomized	175	176	176	168	166	167
Safety Analysis Set †	175	174	173	167	166	167
Full Analysis Set ††	175	173	174	167	166	167
12-week Double-blind Period (Safety Analysis Set)						
Completed	152 (86.9%)	143 (82.2%)	160 (92.5%)	151 (90.4%)	152 (91.6%)	155 (92.8%)
Treatment discontinuation	23 (13.1%)	31 (17.8%)	13 (7.5%)	16 (9.6%)	14 (8.4%)	12 (7.2%)
Primary reason for study intervention discontinuation						
Adverse event	9 (5.1%)	8 (4.6%)	5 (2.9%)	1 (0.6%)	1 (0.6%)	2 (1.2%)
Death	0	0	0	0	0	0
Lost to follow-up	3 (1.7%)	4 (2.3%)	0	2 (1.2%)	1 (0.6%)	2 (1.2%)
Protocol deviation	0	2 (1.1%)	2 (1.2%)	1 (0.6%)	5 (3.0%)	0
Withdrawal by subject	9 (5.1%)	12 (6.9%)	4 (2.3%)	11 (6.6%)	6 (3.6%)	6 (3.6%)
Other	2 (1.1%)	5 (2.9%)	2 (1.2%)	1 (0.6%)	1 (0.6%)	2 (1.2%)

Source: Figure 2 and Table 3, Study Reports for SKYLIGHT 1 and SKYLIGHT 2.
† Summary was based on the actual treatment subjects received. One participant (Subject (b) (6)) randomized into Fezolinetant 45 mg treatment group was given an incorrect treatment with Fezolinetant 30 mg. This participant was considered as part of the Fezolinetant 30 mg group in the safety analyses.
†† Summary was based on the number of subject randomized to the treatment groups.
Note: At the end of the 12-week double-blinded treatment period, subjects in the placebo group were re-randomized 1:1 and received either Fezolinetant 30 mg (76 subjects each in SKYLINE 1 and 2) or Fezolinetant 45 mg (76 subjects in SKYLINE 1 and 75 subjects in SKYLINE 2) for 40-week. Subjects in the Fezolinetant 30 mg and 45 mg groups that completed the 12-week double-blinded treatment period continued to receive the assigned treatment for an additional 40-week.

Demographic and Baseline Characteristics:

Summary of demographics and baseline characteristics of each treatment group are summarized in Table 6 for both studies. As shown, all demographic and baseline characteristics were similar across treatment groups. Median age was 54 years in both studies. In SKYLIGHT 1, approximately 83% of subjects were white and 14% were black or African American. In SKYLIGHT 2, these proportions were approximately 79% and 20%, respectively. Current smokers were approximately 13% in SKYLIGHT 1 and 21% in SKYLIGHT 2. Subjects in both studies had an average BMI of 28 kg/m2 - about a third of patients had a BMI of ≥ 30 kg/m2. Both studies enrolled subjects with prior hormone therapy use (20%), with history of oophorectomy (22%), or with history of hysterectomy (32%).

Table 6. Demographic and Baseline Characteristics
(Full Analysis Set)

Parameter	Category/ Statistic	SKYLIGHT 1			SKYLIGHT 2		
			Fezolinetant			Fezolinetant	
		Placebo (n=175)	30 mg (N = 173)	45 mg (N = 174)	Placebo (n=167)	30 mg (N = 166)	45 mg (N = 167)
Age (years)	Mean (SD)	54.7 (4.8)	54.1 (4.8)	54.3 (5.1)	54.7 (4.6)	53.9 (4.9)	54.3 (5.4)
	Median	54	53	54	54	54	55
	Range	41 - 65	42 - 65	40 - 65	44 - 65	42 - 65	40 - 65
Race, n (%)	White	142 (81.1%)	147 (85.5%)	142 (81.6%)	134 (80.2%)	131 (78.9%)	132 (79.0%)
	Black or African American	28 (16.0%)	21 (12.2%)	26 (14.9%)	31 (18.6%)	35 (21.1%)	33 (19.8%)
	American Indian or Alaska Native	2 (1.1%)	0	1 (0.6%)	0	0	1 (0.6%)
	Asian	3 (1.7%)	3 (1.7%)	3 (1.7%)	1 (0.6%)	0	0
	Native Hawaiian or Other Pacific Islander	0	0	1 (0.6%)	0	0	0
	Other	0	1 (0.6%)	1 (0.6%)	1 (0.6%)	0	1 (0.6%)
	Missing	0	1	0	0	0	0
Ethnicity	Hispanic or Latino	46 (26.4%)	42 (24.3%)	48 (27.6%)	32 (19.3%)	34 (20.5%)	41 (24.6%)
	Not Hispanic or Latino	128 (73.6%)	131 (75.7%)	126 (72.4%)	134 (80.7%)	132 (79.5%)	126 (75.4%)
	Missing	1	0	0	1	0	0
BMI (kg/m^2)	Mean (SD)	28.2 (4.28)	28.1 (4.80)	28.3 (4.39)	28.2 (4.99)	27.9 (4.69)	27.9 (4.35)
	Median	28.2	27.9	27.9	27.8	27.8	27.3
	Range	18.8 - 37.7	18.0 - 37.8	18.4 - 37.9	18.6 - 38.0	18.1 - 37.6	18.0 - 37.5
BMI category n (%)	< 18.5 kg/m2	0	1 (0.6%)	1 (0.6%)	0	1 (0.6%)	1 (0.6%)
	≥ 18.5 to < 25 kg/m2	44 (25.1%)	50 (28.9%)	40 (23.1%)	53 (31.7%)	54 (32.5%)	45 (26.9%)
	≥ 25 to < 30 kg/m2	71 (40.6%)	60 (34.7%)	79 (45.7%)	62 (37.1%)	58 (34.9%)	73 (43.7%)
	≥ 30 kg/m2	60 (34.3%)	62 (35.8%)	53 (30.6%)	52 (31.1%)	53 (31.9%)	48 (28.7%)
	Missing	0	0	1	0	0	0

Parameter	Category/ Statistic	SKYLIGHT 1			SKYLIGHT 2		
			Fezolinetant			Fezolinetant	
		Placebo (n=175)	30 mg (N = 173)	45 mg (N = 174)	Placebo (n=167)	30 mg (N = 166)	45 mg (N = 167)
Weight (kg)	Mean (SD)	74.4 (12.14)	75.0 (13.86)	75.7 (12.89)	74.6 (14.68)	75.3 (14.09)	74.6 (12.45)
	Median	73.6	73.8	74.4	71.7	74.1	73
	Range	47.7 - 111.0	42.0 - 121.2	50.6 - 110.6	46.2 - 125.0	48.0 - 108.4	45.0 - 107.4
Smoking status stratification factor n (%)	Current	22 (12.6%)	21 (12.1%)	23 (13.2%)	35 (21.0%)	34 (20.5%)	34 (20.4%)
	Former/ Never	153 (87.4%)	152 (87.9%)	151 (86.8%)	132 (79.0%)	132 (79.5%)	133 (79.6%)
Prior Hormone Therapy use, n (%)	Yes	33 (19.4%)	31 (18.1%)	30 (17.8%)	31 (18.6%)	37 (22.6%)	38 (23.3%)
History of Oophorectomy, n (%)	Yes	38. (21.7%)	36 (20.8%)	38 (21.8%)	37 (22.2%)	34 (20.5%)	38 (22.8%)
History of Hysterectomy, n (%)	Yes	51 (29.1%)	60 (34.7%)	57 (32.8%)	51 (30.5%)	53 (31.9%)	56 (33.5%)

Source: Table 4 and Table 5 of Applicant's Summary of Clinical Efficacy document.

3.2.4 Results and Conclusions

Results for Co-primary Efficacy Endpoints

The four co-primary efficacy endpoints were analyzed using MMRM analysis of covariance as pre-specified in the protocol. Results are summarized in Table 7 for the frequency and Table 8 for severity of hot flashes.

Subjects enrolled in both studies had an average of 11 moderate to severe hot flashes per day at baseline and an average severity score of 2.4 unit at baseline. As shown in Table 7 and Table 8, data from both SKYLIGHT 1 and SKYLIGHT 2 studies showed a statistically significant reduction from baseline in the frequency as well as in the severity of moderate-to-severe hot flashes per day for both doses of fezolinetant compared to placebo at week 4 and week 12.

Table 7. Change in the frequency of VMS per 24 hours from baseline at week 4 and week 12

	SKYLIGHT 1			**SKYLIGHT 2**		
	Placebo	**Fezolinetant 30 mg**	**Fezolinetant 45 mg**	**Placebo**	**Fezolinetant 30 mg**	**Fezolinetant 45 mg**
Baseline (N)	175	173	174	167	166	167
Mean (SD)	10.51 (3.79)	10.65 (4.73)	10.44 (3.92)	11.59 (5.02)	11.23 (4.88)	11.79 (8.26)
Week 4 (N)	166	157	164	151	155	155
Mean (SD) change from baseline	-3.27 (4.18)	-5.35 (5.57)	-5.20 (4.07)	-3.64 (4.15)	-5.52 (4.23)	-6.24 (4.78)
LS Mean (SE) difference from placebo		-1.87 (0.42)	-2.07 (0.42)		-1.82 (0.46)	-2.55 (0.46)
95% CI		(-2.69, -1.05)	(-2.89, -1.25)		(-2.73, -0.91)	(-3.45, -1.64)
p-value		< 0.001	< 0.001		< 0.001	< 0.001
Week 12 (N)	139	131	146	140	133	145
Mean (SD) change from baseline	-3.67 (4.18)	-6.44 (6.15)	-6.38 (4.48)	-4.57 (5.14)	-6.43 (4.77)	-7.43 (6.47)
LS Mean (SE) difference from placebo		-2.39 (0.44)	-2.55 (0.43)		-1.86 (0.55)	-2.53 (0.55)
95% CI (2-sided)		(-3.25, -1.52)	(-3.40, -1.70)		(-2.94, -0.78)	(-3.60, -1.46)
p-value		< 0.001	< 0.001		< 0.001	< 0.001

Source: Table 11, Study Reports for SKYLIGHT 1 and SKYLIGHT 2.
Abbreviations: LS Means – Least Square Means; SE: Standard Error, CI – Confidence Interval; SD – Standard Deviation.
Note: The LS means, SE, CI and p-values were obtained using MMRM analysis with change from baseline as the dependent variable and treatment group, week, and smoking status (current vs former/never) as factors, with baseline measurement and baseline weight as covariates, as well as an interaction of treatment by week and an interaction of baseline measurement by week.

At week 4, subjects in fezolinetant 30 mg and 45 mg had displayed an average reduction of 5 to 6 moderate to severe hot flashes per day from baseline compared to an average reduction of 3 to 4 hot flashes for subjects in the placebo group. Similarly, at week 12, subjects in fezolinetant 30 mg and 45 mg groups had displayed an average reduction of 6 and 7 moderate to severe hot

flashes per day from baseline, respectively, compared to an average reduction of 4 to 5 hot flashes for subjects in the placebo group.

Although both doses of fezolinetant showed statistically significant reduction in the frequency of moderate-to-severe hot flashes per day at weeks 4 and 12, fezolinetant 30 mg did not meet the clinical meaningful treatment difference of 2 or more reduction at weeks 4 and 12 from baseline compared to placebo in the frequency of moderate-to-severe hot flashes (Table 7). For example, for the 45 mg arm, the LS mean differences from placebo was over 2 unit at week 4 and at week 12 in both studies. That was clinically meaningful and statistically significant. However, for the 30 mg arm, the LS mean differences from placebo was less than 2 at week 4 in SKYLIGHT 1 and SKYLIGHT 2 and at week 12 in SKYLIGHT 2.

For the co-primary efficacy endpoint of VMS severity (Table 8), the baseline severity score was about 2.4 across three arms in the two studies. As shown in Table 8, all mean changes in severity scores showed a decreasing trend with statistically significant difference between each dose of fezolinetant and placebo.

Table 8. Change in the severity of VMS per 24 hours from baseline at week 4 and week 12

	SKYLIGHT 1			**SKYLIGHT 2**		
	Placebo	**Fezolinetant 30 mg**	**Fezolinetant 45 mg**	**Placebo**	**Fezolinetant 30 mg**	**Fezolinetant 45 mg**
Baseline (N)	175	173	174	167	166	167
Mean (SD)	2.43 (0.35)	2.39 (0.34)	2.40 (0.35)	2.41 (0.32)	2.44 (0.33)	2.41 (0.34)
Week 4 (N)	166	157	164	151	155	155
Mean (SD) change from baseline	-0.28 (0.50)	-0.43 (0.56)	-0.45 (0.61)	-0.31 (0.48)	-0.47 (0.58)	-0.61 (0.63)
LS Mean (SE) difference from placebo		-0.15 (0.06)	-0.19 (0.06)		-0.15 (0.06)	-0.29 (0.06)
95% CI (2-sided)		(-0.27, -0.03)	(-0.30, -0.07)		(-0.27, -0.02)	(-0.41, -0.16)
p-value		0.012	0.002		0.021	< 0.001
Week 12 (N)	139	131	146	140	133	145
Mean (SD) change from baseline	-0.35 (0.58)	-0.57 (0.73)	-0.58 (0.75)	-0.46 (0.65)	-0.60 (0.75)	-0.74 (0.71)
LS Mean (SE) difference from placebo		-0.24 (0.08)	-0.20 (0.08)		-0.16 (0.08)	-0.29 (0.08)
95% CI (2-sided)		(-0.39, -0.09)	(-0.35, -0.06)		(-0.33, 0.00)	(-0.45, -0.13)
p-value		0.002	0.007		0.049	< 0.001

Source: Table 12, Study Reports for SKYLIGHT 1 and SKYLIGHT 2. The LS means, SE, CI and p-values were obtained using MMRM analysis with change from baseline as the dependent variable and treatment group, week, and smoking status (current vs former/never) as factors, with baseline measurement and baseline weight as covariates, as well as an interaction of treatment by week and an interaction of baseline measurement by week.

Sensitivity analyses described in section 3.2.2.4 were conducted for the four co-primary efficacy endpoints based on the Per Protocol Set. Reductions in the frequency and the severity of

moderate to severe VMS from baseline to week 4 and to week 12 relative to placebo were statistically significant, similar to those noted in the in the primary efficacy results based on the FAS population (Table 20 in the Appendix). The multiple imputation sensitivity analyses results also supported the primary analysis conclusions (Table 18 and Table 19 in the Appendix).

Analysis of Secondary Efficacy Endpoints

i) Key Secondary Endpoint

The 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 was defined in both studies as the key secondary endpoint. Table 9 shows the results for this endpoint. As shown, there was a reduction (improvement) in the PROMIS total scores for all the treatment arms in both studies but the only reduction that was statistically significant occurred in the fezolinetant 45 mg arm in SKYLIGHT 2.

Table 9. Change from Baseline in PROMIS Sleep Disturbance – Short Form 8b; 12-week Double-blind Period
(Full Analysis Set)

	SKYLIGHT 1			**SKYLIGHT 2**		
	Placebo	**Fezolinetant 30 mg**	**Fezolinetant 45 mg**	**Placebo**	**Fezolinetant 30 mg**	**Fezolinetant 45 mg**
Baseline	175	172	174	166	165	167
Mean (SD)	26.4 (6.6)	26.4 (6.6)	27.1 (7.0)	27.4 (7.0)	27.3 (6.6)	26.2 (6.6)
week 12 (N)	148	133	156	144	139	145
Mean (SD) change from baseline	-3.2 (7.3)	-3.7 (8.2)	-4.6 (7.3)	-3.6 (7.3)	-4.6 (8.1)	-4.8 (6.8)
LS Mean (SE) difference from placebo		-0.5 (0.8)	-1.1 (0.7)		-0.7 (0.7)	-2.0 (0.7)
95% CI (2-sided)		(-2.0, 1.0)	(-2.5, 0.4)		(-2.1, 0.8)	(-3.5, -0.6)
p-value		0.489	0.155		0.381	0.007

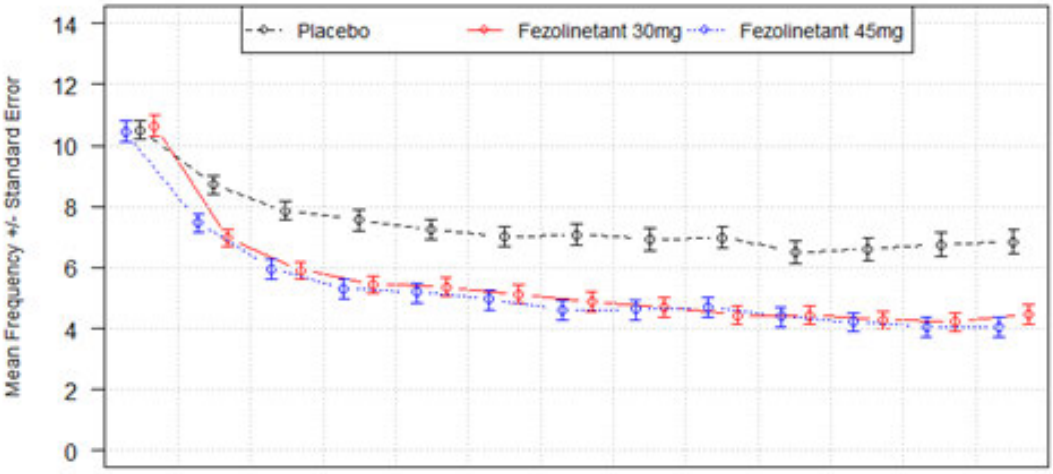
Source: Table 16, Study Reports for SKYLIGHT 1 and SKYLIGHT 2. The LS means, SE, CI and p-values were obtained using MMRM analysis with change from baseline as the dependent variable and treatment group, week, and smoking status (current vs former/never) as factors, with baseline measurement and baseline weight as covariates, as well as an interaction of treatment by week and an interaction of baseline measurement by week.

ii) Mean change in the frequency of moderate and severe VMS from baseline to each week up to week 12

Figure 2 and Figure 3 show the mean change in the frequency of moderate to severe hot flashes over time through week 12 in SKYLINE 1 and 2 studies, respectively. As shown, subjects in the fezolinetant arms (30 mg and 45 mg) had greater reductions from baseline in the mean frequency of moderate to severe hot flashes compared with the placebo arm through the 12-week double-blind period in both studies. The effect was noticed as early as week 1.

Similar profile plots for the mean change in the frequency of moderate to severe hot flashes from baseline through week 52 are presented in Appendix Figure 6 and Figure 7. As shown, placebo subjects that switched and received either dose of fezolinetant from week 12 to week 52 displayed a reduction in the frequency of moderate to severe hot flashes which further affirmed the treatment benefit of fezolinetant in the reduction of moderate to severe hot flashes.

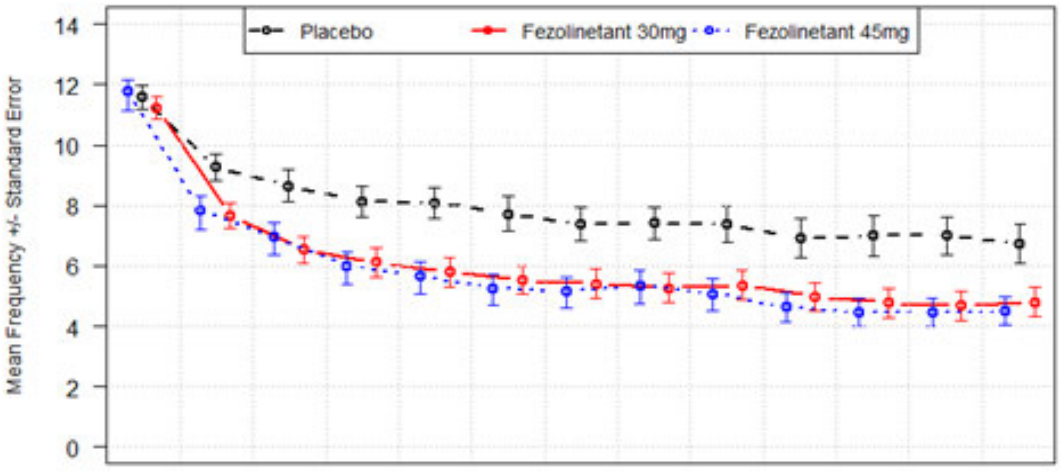
Figure 2. Mean Frequency of VMS by week (SKYLIGHT 1)



	BASE	Wk1	Wk2	Wk3	Wk4	Wk5	Wk6	Wk7	Wk8	Wk9	Wk10	Wk11	Wk12
Placebo	175	174	171	165	166	158	156	154	152	148	141	144	139
Fezolinetant 30mg	173	169	162	159	157	151	144	144	143	141	140	138	131
Fezolinetant 45mg	174	166	168	166	164	159	157	162	151	152	152	149	146

Source: Figure 3 from study report SKYLIGHT 1.

Figure 3. Mean Frequency of VMS by week (SKYLIGHT 2)



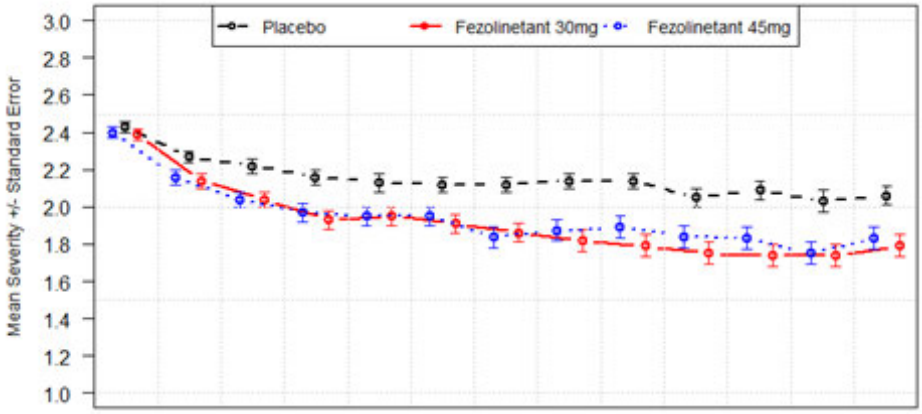
	BASE	Wk1	Wk2	Wk3	Wk4	Wk5	Wk6	Wk7	Wk8	Wk9	Wk10	Wk11	Wk12
Placebo	167	166	159	156	151	152	152	149	148	145	139	138	140
Fezolinetant 30mg	166	164	160	157	155	152	146	143	143	140	138	142	133
Fezolinetant 45mg	167	158	156	156	155	154	149	147	154	148	149	149	145

Source: Figure 3 from study report SKYLIGHT 2.

iii) <u>Mean change in the severity of moderate and severe VMS from baseline to each week up to week 12</u>

Figure 4 and Figure 5 show the mean change in the severity of VMS from baseline over time through week 12 in SKYLINE 1 and 2 studies, respectively. As shown, subjects in the fezolinetant arms (30 mg and 45 mg) had numerically greater reductions from baseline in mean severity of VMS compared with the placebo arm through the 12-week double-blind period in both studies. Similar profile plots for the mean change in the severity of hot flashes from baseline through week 52 are presented in Appendix Figure 8 and Figure 9.

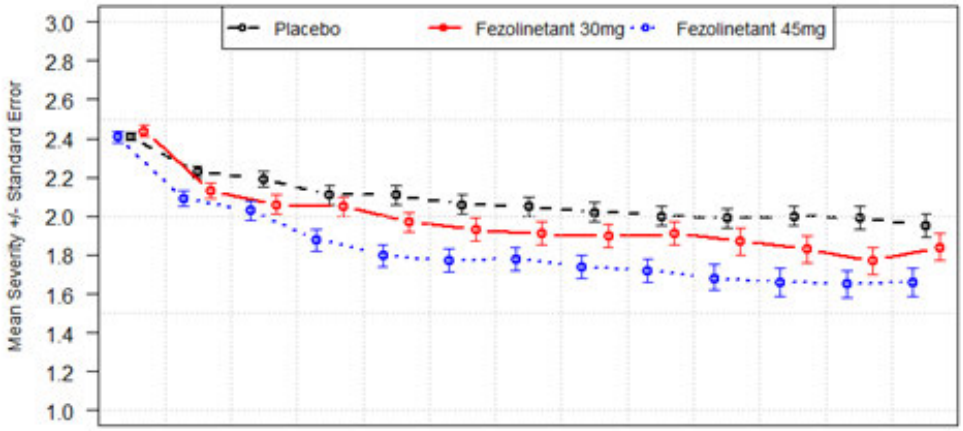
Figure 4. Mean Severity of moderate to severe VMS by week (SKYLIGHT 1)



	BASE	Wk1	Wk2	Wk3	Wk4	Wk5	Wk6	Wk7	Wk8	Wk9	Wk10	Wk11	Wk12
Placebo	175	174	171	165	166	158	156	154	152	148	141	144	139
Fezolinetant 30mg	173	169	162	159	157	151	144	144	143	141	140	138	131
Fezolinetant 45mg	174	166	168	166	164	159	157	162	151	152	152	149	146

Source: Figure 4 from study report SKYLIGHT 1.

Figure 5. Mean Severity of moderate to severe VMS by week (SKYLIGHT 2)



	BASE	Wk1	Wk2	Wk3	Wk4	Wk5	Wk6	Wk7	Wk8	Wk9	Wk10	Wk11	Wk12
Placebo	167	166	159	156	151	152	152	149	148	145	139	138	140
Fezolinetant 30mg	166	164	160	157	155	152	146	143	143	140	138	142	133
Fezolinetant 45mg	167	158	156	156	155	154	149	147	154	148	149	149	145

Source: Figure 4 from study report SKYLIGHT 2.

iv) Mean percent reduction in the frequency of moderate and severe VMS from baseline to each week up to week 12

Subjects in the treatment arms (30 mg and 45 mg) showed bigger percent reduction in the frequency of moderate to severe VMS than subjects in the placebo arm (Table 10 and Table 11). These reductions were observed after week 1 in both studies. For example, at week 4, subjects in the fezolinetant group displayed about a 50% reduction in the frequency of moderate to severe hot flashes compared a 30% reduction in the placebo group. Similarly, at week 12, subjects in the fezolinetant group displayed about a 56% - 65% reduction in the frequency of moderate to severe hot flashes compared a 35% - 45% reduction in the placebo group.

Table 10. Percent change from baseline to each week in mean frequency of moderate to severe VMS up to week 12 (SKYLIGHT 1)

Visit	Placebo	Fezolinetant		Difference in LS Means (95% CI) vs. Placebo	
		30 mg	45 mg	30 mg	45 mg
Week 1	-16.7 (26.50)	-32.3 (31.03)	-28.1 (33.60)	-15.5 (-22.0, -9.1)	-12.2 (-18.7, -5.7)
Week 2	-24.0 (30.87)	-41.9 (33.76)	-42.8 (34.08)	-17.3 (-24.3, -10.3)	-18.3 (-25.3, -11.3)
Week 3	-27.8 (33.47)	-46.9 (33.42)	-49.3 (33.62)	-18.1 (-25.3, -11.0)	-21.2 (-28.3, -14.1)
Week 4	-30.5 (35.30)	-47.8 (34.96)	-50.6 (35.44)	-16.8 (-24.2, -9.3)	-21.1 (28.5, -13.6)
Week 5	-32.4 (35.05)	-50.0 (35.48)	-53.2 (33.82)	-17.6 (-25.0, -10.1)	-21.8 (-29.2, -14.4)
Week 6	-31.8 (36.61)	-52.6 (34.39)	-56.5 (34.37)	-18.4 (-26.0, -10.8)	-23.8 (-31.4, -16.3)
Week 7	-33.9 (35.3)	-54.5 (34.31)	-55.4 (39.32)	-18.3 (-26.3, -10.3)	-21.4 (-29.4, -13.5)
Week 8	-34.0 (34.06)	-56.9 (34.88)	-54.7 (39.40)	-20.2 (-27.1, -12.2)	-21.2 (-29.0, -13.3)
Week 9	-39.1 (35.82)	-57.0 (34.80)	-57.5 (39.24)	-16.7 (-24.7, -8.6)	-18.9 (-26.8, -10.9)
Week 10	-37.4 (35.38)	-58.2 (34.77)	-59.7 (35.8)	-17.9 (-25.8, -10.0)	-19.5 (-27.3, -11.7)
Week 11	-37.1 (34.84)	-59.1 (34.88)	-60.9 (34.3)	-20.3 (-28.3, -12.4)	-22.1 (-30.0, -14.3)
Week 12	-35.0 (39.65)	-56.3 (35.87)	-61.4 (32.71)	-20.1 (-28.2, -12.0)	-24.2 (-32.2, -16.2)

Source: Table 17 and End-of-text Table 9.3.3.5 of study report SKYLIGHT 1.

Table 11. Percent change from baseline to each week in mean frequency of moderate to severe VMS up to week 12 (SKYLIGHT 2)

Visit	Placebo	Fezolinetant		Difference in LS Means (95% CI) vs. Placebo	
		30 mg	45 mg	30 mg	45 mg
Week 1	-20.9 (27.87)	-33.0 (31.4)	-35.3 (33.8)	-12.2 (-18.9, -5.4)	-15.7 (-22.4, -8.9)
Week 2	-28.4 (32.20)	-43.4 (35.35)	-43.4 (35.84)	-14.6 (-22.1, -7.1)	-16.0 (-23.5, -8.5)
Week 3	-32.7 (34.47)	-47.8 (35.69)	-51.9 (36.08)	-15.5 (23.2, -7.9)	-20.3 (-27.9, -12.6)
Week 4	-33.6 (34.14)	-51.6 (36.48)	-55.2 (36.25)	-16.3 (-24.0, -8.6)	-21.7 (-29.4, -13.9)
Week 5	-37.1 (35.36)	-53.1 (36.09)	-59.5 (34.97)	-15.6 (-23.3, -8.0)	-22.9 (-30.5, -15.2)
Week 6	-39.2 (35.18)	-53.6 (36.21)	-60.4 (-34.58)	-14.8 (-22.4, -7.2)	-22.7 (-30.3, -15.1)
Week 7	-39.5 (35.87)	-54.3 (37.55)	-60.1 (36.04)	-15.0 (22.9, -7.1)	-21.5 (29.3, -13.6)
Week 8	-40.7 (37.64)	-54.7 (36.63)	-61.4 (35.95)	-14.4 (22.3, -6.4)	-20.6 (-28.5, -12.7)
Week 9	-44.2 (37.48)	-56.6 (38.34)	-63.6 (34.65)	-12.1 (-20.1, -4.1)	-19.7 (-27.7, -11.8)
Week 10	-43.6 (37.74)	-59.0 (35.39)	-64.5 (35.38)	-14.4 (-22.3, -6.5)	-20.1 (-27.9, -12.3)
Week 11	-42.8 (38.30)	-60.0 (36.49)	-64.4 (35.81)	-15.0 (-23.0, -7.0)	-20.2 (-28.1, -12.2)
Week 12	-45.4 (39.79)	-58.6 (35.44)	-64.3 (34.92)	-13.6 (-21.6, -5.7)	-18.9 (-26.9, -11.0)

Source: Table 17 and End-of-text Table 9.3.3.5 of study report SKYLIGHT 2.

v) Percent reduction ≥ 50% and at 100% in the frequency of moderate and severe VMS from baseline to each week up to week 12

In SKYLIGHT 1, the proportion of subject who had ≥ 50% reductions in the frequency of moderate to severe VMS increased in each visit from 10.3% (placebo), 27.2% (30 mg) and 25.3% (45 mg) at week 1 to 29.7% (placebo), 44.5% (30 mg) and 56.9% (45 mg) at week 12.

Table 12. Number of responders with reductions from baseline in frequency of moderate to severe VMS by week (SKYLIGHT 1)

Visit	≥ 50% reduction			100% reduction		
		Fezolinetant			Fezolinetant	
	Placebo (N = 175)	30 mg (N =173)	45 mg (N = 174)	Placebo (N = 175)	30 mg (N = 173)	45 mg (N = 174)
Week 1	18 (10.3%)	47 (27.2%)	44 (25.3%)	0	0	0
Week 2	37 (21.1%))	64 (37.0%)	75 (43.1%)	1 (0.6%)	3 (1.7%)	4 (2.3%)
Week 3	42 (24.0%)	69 (39.9%)	89 (51.1%)	1 (0.6%)	6 (3.5%)	5 (2.9%)
Week 4	49 (28.0%)	77 (44.5%)	94 (54.0%)	5 (2.9%)	6 (3.5%)	8 (4.6%)
Week 5	47 (26.9%)	76 (43.9%)	94 (54.0%)	2 (1.1%)	10 (5.8%)	6 (3.4%)
Week 6	50 (28.6%)	78 (45.1%)	96 (55.2%)	2 (1.1%)	8 (4.6%)	10 (5.7%)
Week 7	52 (29.7%)	79 (45.7%)	98 (56.3%)	2 (1.1%)	13 (7.5%)	13 (7.5%)
Week 8	52 (29.7%)	93 (53.8%)	87 (50.0%)	2 (1.1%)	10 (5.8%)	14 (8.0%)
Week 9	56 (32.0%)	85 (49.1%)	97 (55.7%)	5 (2.9%)	15 (8.7%)	16 (9.2%)
Week 10	45 (25.7%)	84 (48.6%)	100 (57.5%)	7 (4.0%)	17 (9.8%)	18 (10.3%)
Week 11	55 (31.4%)	85 (49.1%)	100 (57.5%)	10 (5.7%)	16 (9.2%)	19 (10.9%)
Week 12	52 (29.7%)	77 (44.5%)	99 (56.9%)	6 (3.4%)	12 (6.9%)	18 (10.3%)

Source: Table 18 and End-of-Text Table 9.3.3.6.1 from study report SKYLIGHT 1.

In SKYLIGHT 2, the proportion of subjects who had ≥ 50% reductions in the frequency of moderate to severe VMS increased in each visit, from 16.8% (placebo), 27.7% (30 mg) and 34.7% (45 mg) at week 1 to 42.5% (placebo), 50.6% (30 mg) and 60.5% (45 mg) at week 12.

Table 13. Number of responders with reductions from baseline in frequency of moderate to severe VMS by week (SKYLIGHT 2)

Visit	≥ 50% reduction			100% reduction		
		Fezolinetant			Fezolinetant	
	Placebo (N = 167)	30 mg (N =166)	45 mg (N = 167)	Placebo (N = 167)	30 mg (N = 166)	45 mg (N = 167)
Week 1	28 (16.8%)	46 (27.7%)	58 (34.7%)	1 (0.6%)	1 (0.6%)	3 (1.8%)
Week 2	39 (23.4%)	71 (42.8%)	71 (42.5%)	3 (1.8%)	8 (4.8%)	4 (2.4%)
Week 3	48 (28.7%)	81 (48.8%)	89 (53.3%)	5 (3.0%)	4 (2.4%)	11 (6.6%)
Week 4	44 (26.3%)	84 (50.6%)	88 (52.7%)	3 (1.8%)	10 (6.0%)	17 (10.2%)
Week 5	54 (32.3%)	84 (50.6%)	98 (58.7%)	3 (1.8%)	12 (7.2%)	11 (6.6%)
Week 6	53 (31.7%)	81 (48.8%)	95 (56.9%)	9 (5.4%)	12 (7.2%)	17 (10.2%)
Week 7	55 (32.9%)	84 (50.6%)	92 (55.1%)	9 (5.4%)	13 (7.8%)	18 (10.8%)
Week 8	56 (33.5%)	81 (48.8%)	103 (61.7%)	10 (6.0%)	17 (10.2%)	22 (13.2%)
Week 9	64 (38.3%)	84 (50.6%)	98 (58.7%)	9 (5.4%)	14 (8.4%)	18 (10.8%)
Week 10	62 (37.1%)	85 (51.2%)	103 (61.7%)	11 (6.6%)	17 (10.2%)	25 (15.0%)
Week 11	60 (35.9%)	94 (56.6%)	105 (62.9%)	9 (5.4%)	15 (9.0%)	28 (16.8%)
Week 12	71 (42.5%)	84 (50.6%)	101 (60.5%)	9 (5.4%)	15 (9.0%)	25 (15.0%)

Source: Table 18 and End-of-Text Table 9.3.3.6.1 from study report SKYLIGHT 2.

3.2.5 Efficacy Conclusion

Fezolinetant at doses 30 mg and 45 mg demonstrated statistically significant reductions from baseline in the mean frequency of moderate to severe VMS and in the mean severity of VMS per 24 hours at week 4 and at week 12 compared to placebo. However, fezolinetant 30 mg did not meet the clinical meaningful treatment difference of 2 or more reduction compared to placebo in the frequency of VMS. Sensitivity analysis supported these findings.

3.3 **Evaluation of Safety**

Safety events were reviewed by Dr. Joo-Yeon Lee from Division of Biometrics VII and by Dr. Regina Zopf from Division of Urology, Obstetrics, and Gynecology (DUOG). Readers are referred to their respective reviews for this section.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

The analysis for the subgroups was conducted using a simplified version of the 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.

Subjects were stratified according to their smoking status. There is a small number of current smokers in SKYLIGHT 1 and SKYLIGHT 2 compared to the number of former/never smokers, across arms. This is a limitation for the interpretation of the results. Analyses for smoking status and other subgroups such as age, race and BMI were also performed for the co-primary endpoints and are shown in
Table 14 to Table 17.

Table 14. Change from Baseline in Mean Frequency of VMS from baseline at week 4 and week 12 by subgroup (SKYLIGHT 1)

	Subgroup	**Placebo**	**Fezolinetant**		**Difference in LS Means (95% CI) vs. Placebo**	
			30 mg	**45 mg**	**30 mg**	**45 mg**
Week 4						
Age	< 55 years	-3.4 (4.03)	-5.1 (4.51)	-5.3 (4.25)	-1.6 (-2.8, -0.5)	-1.8 (-2.9, -0.7)
	≥ 55 years	-3.1 (4.35)	-5.6 (6.56)	-5.1 (3.85)	-2.2 (-3.4, -1.0)	-2.5 (-3.7, -1.3)
Race	African American	-4.9 (6.14)	-5.4 (4.88)	-6.0 (3.28)	-0.8 (-3.0, 1.4)	-2.1 (-4.1, 0.0)
	Non-African American	-3.0 (3.66)	-5.3 (5.68)	-5.0 (4.20)	-2.1 (-3.0, -1.2)	-2.1 (-3.0, -1.2)
	White	-2.9 (3.7)	-5.2 (5.7)	-5.0 (4.24)	-2.0 (-2.9, -1.1)	-2.1 (-3.0, -1.2)
	Non-White	-4.8 (5.8)	-5.9 (4.69)	-6.3 (3.17)	-1.3 (-3.4, 0.7)	-2.2 (-4.0, -0.3)
BMI	≥18.5 to < 25 kg/m^2	-3.5 (3.69)	-5.0 (4.10)	-5.6 (3.86)	-1.5 (-3.0, -0.1)	-2.4 (-3.9, -0.9)
	≥25 to < 30 kg/m^2	-3.0 (3.94)	-4.7 (3.54)	-5.0 (4.44)	-1.5 (-2.8, -0.2)	-1.8 (-3.1, -0.6)
	≥ 30 kg/m^2	-3.5 (4.80)	-6.3 (7.74)	-5.2 (3.74)	-2.6 ('-4.0, -1.1)	-2.3 (-3.8, -0.7)
History of	Yes	-3.7 (0.60)	-5.1 (0.56)	-5.4 (0.57)	-1.4 (-3.0, 0.2)	-1.7 (-3.3, -0.1)

	Subgroup	Placebo	Fezolinetant		Difference in LS Means (95% CI) vs. Placebo	
			30 mg	45 mg	30 mg	45 mg
Hysterectomy	No	-3.1 (0.33)	-5.2 (0.35)	-5.4 (0.34)	-2.1 (-3.1, -1.2)	-2.3 (-3.2, -1.3)
History of Oophorectomy	Yes	-3.5 (0.72)	-4.6 (0.75)	-5.1 (0.73)	-1.0 (-3.1, 1.1)	-1.5 (-3.5, 0.5)
	No	-3.2 (0.31)	-5.3 (0.32)	-5.5 (0.32)	-2.1 (-2.9, -1.2)	-2.3 (-3.1, -1.4)
Prior Hormone Therapy Use	Yes	-3.0 (0.54)	-5.8 (0.56)	-6.7 (0.56)	-2.8 (-4.4, -1.3)	-3.7 (-5.3, -2.2)
	No	-3.4 (0.34)	-5.1 (0.34)	-5.2 (0.34)	-1.7 (-2.6, -0.7)	-1.8 (-2.7, -0.8)
Smoking Status	Current Smokers	-3.9 (4.11)	-6.4 (4.20)	-4.4 (4.33)	-2.2 (-4.7, 0.2)	-1.6 (-4.1, 0.8)
	Former/Never Smokers	-3.2 (4.20)	-5.2 (5.73)	-5.3 (4.04)	-1.9 (-2.8, -1.0)	-2.2 (-3.0, -1.3)
Week 12						
Age	< 55 years	-4.0 (4.19)	-6.2 (0.41)	-6.6 (0.39)	-2.0 (-3.1, -0.8)	-2.4 (-3.5, -1.3)
	≥ 55 years	-3.3 (4.17)	-6.8 (7.30)	-6.0 (3.95)	-2.9 (-4.2, -1.6)	-2.8 (-4.1, -1.5)
Race	African American	-3.8 (4.98)	-6.7 (5.34)	-7.0 (2.31)	-2.1 (4.3, 0.2)	-3.1 (-5.1, -1.0)
	Non-African American	-3.7 (4.05)	-6.4 (6.27)	-6.3 (4.78)	-2.5 (-3.4, -1.5)	-2.4 (-3.4, -1.5)
	White	-3.7 (4.05)	-6.4 (6.34)	-6.2 (4.85)	-2.4 (-3.4, -1.4)	-2.5 (-3.4, -1.5)
	Non-White	-3.7 (4.83)	-6.7 (5.05)	-7.18 (2.33)	-2.2 (-4.2, -0.2)	-3.2 (-5.0, -1.4)
BMI	≥18.5 kg/m^2 to < 25 kg/m^2	-3.6 (3.86)	-5.9 (4.19)	-6.2 (3.14)	-2.3 (-3.9, -0.8)	-3.0 (-4.5, -1.4)
	≥25 kg/m^2 to < 30 kg/m^2	-3.8 (4.23)	-5.9 (4.19)	-6.6 (5.41)	-1.9 (-3.4, -0.5)	-2.3 (-3.6, -1.0)
	≥ 30 kg/m^2	-3.6 (443)	-7.5 (8.54)	-6.0 (3.86)	-3.0 (-4.5, -1.5)	-2.5 (-4.1, -0.9)
History of Hysterectomy	Yes	-4.1 (0.65)	-5.9 (0.61)	-6.7 (0.62)	-1.8 (-3.6, -0.1)	-2.6 (-4.4, -0.9)
	No	-3.8 (0.34)	-6.5 (0.36)	-6.4 (0.34)	-2.7 (-3.7, -1.8)	-2.6 (-3.5, -1.6)
History of Oophorectomy	Yes	-4.2 (0.75)	-5.2 (0.78)	-6.9 (0.76)	-1.0 (-3.2, 1.1)	-2.7 (-4.9, -0.6)
	No	-3.8 (0.33)	-6.6 (0.34)	-6.4 (0.32)	-2.7 (-3.7, -1.8)	-2.5 (-3.4, -1.6)
Prior Hormone Therapy Use	Yes	-3.1 (0.69)	-6.2 (0.74)	-6.9 (0.71)	-3.1 (-5.1, -1.1)	-3.8 (-5.7, -1.8)
	No	-4.1 (0.35)	-6.3 (0.35)	-6.4 (0.34)	-2.2 (-3.2, -1.3)	-2.3 (-3.3, -1.4)
Smoking Status	Current Smokers	-3.5 (4.30)	-8.5 (5.33)	-5.8 (-3.86)	-4.1 (-6.5, -1.6)	-3.5 (-5.9, -1.1)
	Former/Never Smokers	-3.7 (4.18)	-6.2 (6.21)	-6.5 (4.57)	-2.2 (-3.1, -1.3)	-2.4 (-3.3, -1.5)

Source: Table 15 and End-of-Text Table 9.3.1.7.7, End-of-Text Table 9.3.1.7.3.1, End-of-Text Table 9.3.1.7.3.3 and End-of-Text Table 9.3.1.7.5 from Study Report of SKYLIGHT 1. The LS means, SE, and CI were obtained a MMRM analysis with change from baseline as the dependent variable and treatment group and week as factors, with baseline measurement as covariate, as well as an interaction of treatment by week and an interaction of baseline measurement by week.

Table 15. Change from Baseline in Mean Frequency of VMS from baseline at week 4 and week 12 by subgroup (SKYLIGHT 2)

	Subgroup	Placebo	Fezolinetant		Difference in LS Means (95% CI) vs. Placebo	
			30 mg	45 mg	30 mg	45 mg
Week 4						
Age	< 55 years	-3.8 (4.47)	-5.9 (4.65)	-7.0 (5.77)	-2.1 (-3.5, -0.7)	-3.1 (-4.5, -1.6)
	≥ 55 years	-3.4 (3.77)	-5.0 (3.60)	-5.5 (3.46)	-1.5 (-2.7, -0.3)	-2.0 (-3.2, -0.9)
Race	African American	-2.1 (5.55)	-5.6 (4.41)	-6.4 (5.11)	-3.5 (-6.3, -0.7)	-4.2 (-7.1, -1.4)
	Non-African American	-4.0 (3.67)	-5.5 (4.20)	-6.2 (4.72)	-1.4 (-2.4, -0.4)	-2.1 (-3.1, -1.2)
	White	-4.0 (3.69)	-5.5 (4.20)	-6.1 (4.27)	-1.4 (-2.3, -0.4)	-2.1 (-3.0, -1.1)
	Non-White	-2.2 (5.48)	-5.6 (4.41)	-6.2 (5.12)	-3.4 (-6.2, -0.7)	-4.0 (-6.7, -1.2)
BMI	≥18.5 kg/m^2 to < 25 kg/m^2	-3.8 (3.31)	-5.5 (3.85)	-6.3 (3.20)	-1.8 (-3.1, -0.4)	-2.8 (-4.2, -1.4)
	≥25 kg/m^2 to < 30 kg/m^2	-3.8 (4.240	-5.9 (4.52)	-6.2 (4.69)	-1.7 (-3.2, -0.2)	-2.0 (-3.4, -0.5)
	≥ 30 kg/m^2	-3.4 (4.83)	-5.2 (4.34)	-6.5 (5.99)	-2.1 (-4.0, -0.2)	-3.2 (-5.2, -1.3)
History of Hysterectomy	Yes	-3.6 (0.62)	-4.8 (0.60)	-6.0 (0.60)	-1.3 (-3.0, 0.4)	-2.4 (-4.1, -0.7)
	No	-3.8 (0.38)	-5.9 (0.38)	-6.3 (0.38)	-2.1 (-3.1, -1.0)	-2.5 (-3.6, -1.5)
History of Oophorectomy	Yes	-3.4 (0.92)	-4.8 (0.95)	-6.6 (0.92)	-1.4 (-4.0, 1.2)	-3.2 (-5.8, -0.7)
	No	-3.8 (0.35)	-5.7 (0.34)	-6.1 (0.35)	-1.9 (-2.9, -0.9)	-2.3 (-3.3, -1.3)
Prior Hormone Therapy Use	Yes	-3.4 (0.55)	-6.8 (0.50)	-7.5 (0.50)	-3.3 (-4.8, -1.9)	-4.1 (-5.6, -2.6)
	No	-3.8 (0.38)	-5.1 (0.39)	-5.9 (0.39)	-1.32 (-2.4, -0.3)	-2.2 (-3.2, -1.1)
Smoking Status	Current Smokers	-3.8 (4.36)	-6.7 (3.78)	-7.9 (5.61)	-2.0 (-3.9, -0.1)	-3.5 (-5.4, -1.5)
	Former/Never Smokers	-3.6 (4.11)	-5.2 (4.30)	-5.8 (4.48)	-1.7 (-2.7, -0.6)	-2.3 (-3.3, -1.3)
Week 12						
Age	< 55 years	-5.0 (0.61)	-7.4 (0.59)	-8.3 (0.63)	-2.4 (-4.1, -0.8)	-3.3 (-5.0, -1.6)
	≥ 55 years	-4.8 (4.32)	-6.1 (3.55)	-6.6 (3.97)	-1.4 (-2.6, -0.1)	-1.8 (-3.0. -0.6)
Race	African American	-2.6 (8.14)	-6.6 (6.13)	-7.4 (4.96)	-3.9 (-5.8, -1.0)	-4.5 (-7.4, -1.6)
	Non-African American	-5.0 (4.04)	-6.4 (4.41)	-7.4 (6.77)	-1.4 (-2.4, -0.4)	-2.1 (-3.1, -1.1)
	White	-5.0 (4.07)	-6.4 (4.41)	-7.2 (6.16)	-1.3 (-2.3, -0.3)	-1.9 (-2.9, -0.9)
	Non-White	-2.8 (8.03)	-6.6 (6.13)	-7.3 (4.89)	-3.8 (-6.6, -1.0)	-4.3 (-7.1, -1.4)
BMI	≥18.5 kg/m^2 to < 25 kg/m^2	-5.2 (3.05)	-6.0 (3.56)	-6.9 (2.75)	-1.0 (2.4, 0.3)	-1.7 (-3.1, -0.4)
	≥25 kg/m^2 to < 30 kg/m^2	-4.3 (5.16)	-7.1 (5.29)	-7.7 (7.93)	-2.0 (-3.9, -0.2)	-2.3 (-4.1, -0.6)
	≥ 30 kg/m^2	-4.3 (6.65)	-6.2 (5.27)	-7.7 (6.65)	-2.0 (-4.3, 0.3)	-3.5 (-5.8, 1.1)
History of Hysterectomy	Yes	-4.1 (0.87)	-6.7 (0.85)	-7.3 (0.85)	-2.6 (-5.0, -0.2)	-3.1 (-5.5, -0.7)
	No	-5.3 (0.40)	-6.9 (0.40)	-7.6 (0.40)	-1.6 (-2.7, -0.5)	-2.3 (-3.4, -1.2)

	Subgroup	Placebo	Fezolinetant		Difference in LS Means (95% CI) vs. Placebo	
			30 mg	45 mg	30 mg	45 mg
History of Oophorectomy	Yes	-3.8 (0.93)	-6.8 (0.98)	-7.9 (0.95)	-2.9 (-5.6, -0.3)	-4.1 (-6.7, -1.5)
	No	-5.3 (0.36)	-6.8 (0.35)	-7.3 (0.35)	-1.6 (-2.6, -0.5)	-2.1 (-3.0, -1.0)
Prior Hormone Therapy Use	Yes	-3.9 (0.64)	-7.2 (0.59)	-7.9 (0.58)	-3.3 (-5.0, -1.5)	-4.0 (-5.7, -2.2)
	No	-5.2 (0.46)	-6.6 (0.48)	-7.4 (0.48)	-1.4 (-2.7, -0.1)	-2.2 (-3.5, -0.9)
Smoking Status	Current Smokers	-4.1 (4.38)	-7.3 (4.94)	-8.5 (6.36)	-2.4 (-4.4, -0.3)	-3.4 (-5.5, -1.4)
	Former/Never Smokers	-4.7 (5.31)	-6.2 (4.73)	-7.2 (6.50)	-1.6 (-2.9, -0.4)	-2.3 (-3.5, -1.0)

Source: Table 15 and End-of-Text Table 9.3.1.7.7, End-of-Text Table 9.3.1.7.3.1, End-of-Text Table 9.3.1.7.3.3 and End-of-Text Table 9.3.1.7.5 from Study Report of SKYLIGHT 2. The LS means, SE, and CI were obtained a MMRM analysis with change from baseline as the dependent variable and treatment group and week as factors, with baseline measurement as covariate, as well as an interaction of treatment by week and an interaction of baseline measurement by week.

Table 16. Change from Baseline in Mean Severity of VMS from baseline at week 4 and week 12 by subgroup (SKYLIGHT 1)

	Subgroup	Placebo	Fezolinetant		Difference in LS Means (95% CI) vs. Placebo	
			30 mg	45 mg	30 mg	45 mg
Week 4						
Age	< 55 years	-0.3 (0.55)	-0.4 (0.57)	-0.4 (0.58)	-0.1 (-0.3, 0.1)	-0.1 (-0.3, 0.1)
	≥ 55 years	-0.2 (0.44)	-0.4 (0.54)	-0.5 (0.65)	-0.2 (-0.4, -0.1)	-0.3 (-0.4, -0.1)
Race	African American	-0.3 (0.33)	-0.4 (0.55)	-0.5 (0.64)	-0.2 (-0.5, 0.2)	-0.2 (-0.5, 0.1)
	Non-African American	-0.3 (0.53)	-0.4 (0.56)	-0.5 (0.60)	-0.2 (-0.3, -0.0)	-0.2 (-0.3, -0.1)
	White	-0.3 (0.50)	-0.4 (0.56)	-0.5 (0.60)	-0.1 (-0.3, -0.0)	-0.2 (-0.3, -0.1)
	Non-White	-0.3 (0.50)	-0.5 (0.55)	-0.5 (0.66)	-0.2 (-0.5, 0.2)	-0.2 (-0.5, 0.1)
BMI	≥18.5 kg/m^2 to < 25 kg/m^2	-0.4 (0.62)	-0.4 (0.54)	-0.5 (0.62)	0.0 (-0.2, 0.3)	-0.1 (-0.4, 0.2)
	≥25 kg/m^2 to < 30 kg/m^2	-0.3 (0.52)	-0.4 (0.54)	-0.4 (0.54)	-0.2 (-0.3, 0.0)	-0.1 (-0.3, 0.0)
	≥ 30 kg/m^2	-0.2 (0.33)	-0.5 (0.60)	-0.4 (0.69)	-0.3 (-0.5, -0.1)	-0.3 (-0.5, -0.1)
History of Hysterectomy	Yes	-0.3 (0.07)	-0.3 (0.07)	-0.4 (0.07)	0.0 (-0.2, 0.2)	-0.2 (-0.4, 0.0)
	No	-0.3 (0.05)	-0.5 (0.05)	-0.5 (0.05)	-0.2 (0.4, -0.1)	-0.2 (-0.3, -0.1)
History of Oophorectomy	Yes	-0.3 (0.08)	-0.2 (0.08)	-0.4 (0.08)	0.1 (-0.1, 0.3)	-0.1 (-0.3, 0.1)
	No	-0.3 (0.05)	-0.5 (0.05)	-0.5 (0.05)	-0.2 (-0.4, -0.1)	-0.2 (-0.3, -0.1)
Prior Hormone Therapy Use	Yes	-0.3 (0.11)	-0.4 (0.11)	-0.5 (0.11)	-0.1 (-0.4, 0.2)	-0.2 (-0.5, 0.1)
	No	-0.3 (0.05)	-0.4 (0.05)	-0.5 (0.05)	-0.2 (-0.3, -0.0)	-0.2 (-0.3, -0.1)
Smoking Status	Current Smokers	-0.5 (0.79)	-0.4 (0.49)	-0.7 (0.87)	0.1 (-0.3, 0.6)	-0.3 (-0.7, 0.2)
	Former/Never Smokers	-0.3 (0.44)	-0.4 (0.57)	-0.4 (0.55)	-0.2 (-0.3, -0.1)	-0.2 (-0.3, -0.1)

	Subgroup	Placebo	Fezolinetant		Difference in LS Means (95% CI) vs. Placebo	
			30 mg	45 mg	30 mg	45 mg
Week 12						
Age	< 55 years	-0.4 (0.68)	-0.6 (0.77)	-0.6 (0.75)	-0.2 (-0.4, 0.1)	-0.1 (-0.3, 0.1)
	≥ 55 years	-0.3 (0.44)	-0.5 (0.67)	-0.6 (0.75)	-0.3 (-0.5, -0.1)	-0.3 (-0.5, -0.1)
Race	African American	-0.3 (0.44)	-0.6 (0.86)	-0.6 (0.70)	-0.3 (-0.7, 0.2)	-0.3 (-0.7, 0.1)
	Non-African American	-0.4 (0.60)	-0.6 (0.71)	-0.6 (0.76)	-0.2 (-0.4, -0.1)	-0.2 (-0.4, -0.0)
	White	-0.4 (0.60)	-0.5 (0.68)	-0.6 (0.74)	-0.2 (-0.4, -0.1)	-0.2 (-0.4, -0.0)
	Non-White	-0.3 (0.41)	-0.7 (0.90)	-0.6 (0.77)	-0.3 (-0.8, 0.1)	-0.3 (-0.7, 0.1)
BMI	≥18.5 kg/m^2 to < 25 kg/m^2	-0.4 (0.69)	-0.4 (0.47)	-0.5 (0.73)	-0.0 (-0.3, 0.3)	-0.2 (-0.5, 0.1)
	≥25 kg/m^2 to < 30 kg/m^2	-0.4 (0.62)	-0.5 (0.67)	-0.6 (0.7)	-0.2 (-0.4, 0.1)	-0.1 (-0.3, 0.1)
	≥ 30 kg/m^2	-0.2 (0.38)	-0.8 (0.89)	-0.6 (0.83)	-0.5 (-0.7, -0.2)	-0.4 (-0.7, -0.1)
History of Hysterectomy	Yes	-0.4 (0.10)	-0.5 (0.09)	-0.5 (0.09)	-0.1 (-0.3, 0.2)	-0.1 (-0.4, 0.1)
	No	-0.4 (0.07)	-0.7 (0.07)	-0.6 (0.07)	-0.3 (-0.5, -0.1)	-0.2 (-0.4, -0.1)
History of Oophorectomy	Yes	-0.3 (0.08)	-0.3 (0.09)	-0.4 (0.08)	-0.0 (-0.2, 0.2)	-0.1 (-0.3, 0.1)
	No	-0.4 (0.05)	-0.7 (0.05)	-0.6 (0.05)	-0.3 (-0.5, -0.2)	-0.2 (-0.4, -0.1)
Prior Hormone Therapy Use	Yes	-0.3 (0.13)	-0.5 (0.14)	-0.7 (0.13)	-0.18 (-0.6, 0.2)	-0.3 (-0.7, 0.0)
	No	-0.4 (0.06)	-0.6 (0.06)	-0.6 (0.06)	-0.3 (-0.4, -0.1)	-0.2 (-0.4, -0.0)
Smoking Status	Current Smokers	-0.4 (0.68)	-0.5 (0.17)	-0.8 (0.16)	-0.2 (-0.6, 0.3)	-0.5 (-1.0, -0.0)
	Former/Never Smokers	-0.3 (0.56)	-0.6 (0.73)	-0.5 (0.72)	-0.2 (-0.4, -0.1)	-0.2 (-0.3, 0.0)

Source: Table 15 and End-of-Text Table 9.3.1.7.8, End-of-Text Table 9.3.1.7.4.1, End-of-Text Table 9.3.1.7.4.3 and End-of-Text Table 9.3.1.7.6 from Study Report of SKYLIGHT 1. The LS means, SE, and CI were obtained a MMRM analysis with change from baseline as the dependent variable and treatment group and week as factors, with baseline measurement as covariate, as well as an interaction of treatment by week and an interaction of baseline measurement by week.

Table 17. Change from Baseline in Mean Severity of VMS from baseline at week 4 and week 12 by subgroup (SKYLIGHT 2)

	Subgroup	Placebo	Fezolinetant		Difference in LS Means (95% CI) vs. Placebo	
			30 mg	45 mg	30 mg	45 mg
Week 4						
Age	< 55 years	-0.3 (0.50)	-0.5 (0.61)	-0.7 (0.70)	-0.1 (-0.3, 0.0)	-0.3 (-0.5, -0.1)
	≥ 55 years	-0.3 (0.46)	-0.4 (0.53)	-0.6 (0.56)	-0.2 (-0.3, 0.0)	-0.3 (-0.4, -0.1)
Race	African American	-0.29 (0.56)	-0.5 (0.76)	-0.6 (0.66)	-0.3 (-0.6, 0.1)	-0.3 (-0.6, 0.1)
	Non-African American	-0.3 (0.46)	-0.5 (0.52)	-0.6 (0.62)	-0.1 (-0.3, 0.0)	-0.3 (-0.4, -0.2)
	White	-0.3 (0.46)	-0.5 (0.52)	-0.6 (0.62)	-0.1 (-0.3, 0.0)	-0.3 (-0.4, -0.2)
	Non-White	-0.3 (0.56)	-0.5 (0.76)	-0.5 (0.67)	-0.3 (-0.6, 0.1)	-0.2 (-0.6, 0.1)

	Subgroup	Placebo	Fezolinetant		Difference in LS Means (95% CI) vs. Placebo	
			30 mg	45 mg	30 mg	45 mg
BMI	≥18.5 kg/m^2 to < 25 kg/m^2	-0.3 (0.48)	-0.6 (0.54)	-0.7 (0.61)	-0.3 (-0.5, -0.1)	-0.4 (-0.6, -0.2)
	≥25 kg/m^2 to < 30 kg/m^2	-0.3 (0.37)	-0.5 (0.64)	-0.6 (0.60)	-0.2 (-0.4, 0.1)	-0.3 (-0.5, -0.1)
	≥ 30 kg/m^2	-0.3 (0.59)	-0.4 (0.52)	-0.6 (0.69)	-0.1 (-0.3, 0.2)	-0.3 (-0.5, -0.0)
History of Hysterectomy	Yes	-0.3 (0.08)	-0.4 (0.08)	-0.6 (0.08)	-0.2 (-0.4, 0.1)	-0.4 (-0.6, -0.2)
	No	-0.4 (0.05)	-0.5 (0.05)	-0.6 (0.05)	-0.1 (-0.3, 0.0)	-0.2 (-0.4, -0.1)
History of Oophorectomy	Yes	-0.3 (0.10)	-0.4 (0.11)	-0.7 (0.11)	-0.1 (-0.4, 0.2)	-0.4 (-0.7, -0.1)
	No	-0.3 (0.05)	-0.5 (0.05)	-0.6 (0.05)	0.1 (-0.3, 0.0)	-0.3 (-0.4, -0.1)
Prior Hormone Therapy Use	Yes	-0.3 (0.11)	-0.5 (0.10)	-0.8 (0.10)	-0.2 (-0.5, 0.1)	-0.5 (-0.8, -0.2)
	No	-0.3 (0.05)	-0.5 (0.05)	-0.6 (0.05)	-0.1 (-0.3, 0.0)	-0.2 (-0.4, -0.1)
Smoking Status	Current Smokers	-0.3 (0.59)	-0.6 (0.72)	-0.7 (0.68)	-0.3 (-0.6, 0.1)	-0.4 (-0.8, -0.1)
	Former/Never Smokers	-0.3 (0.46)	-0.4 (0.53)	-0.6 (0.62)	-0.1 (-0.3, 0.0)	-0.3 (-0.4, -0.1)
Week 12						
Age	< 55 years	-0.5 (0.66)	-0.7 (0.82)	-0.8 (0.70)	-0.2 (-0.4, 0.0)	-0.3 (-0.5, -0.0)
	≥ 55 years	-0.4 (0.65)	-0.5 (0.08)	-0.7 (0.72)	-0.1 (-0.4, 0.1)	-0.3 (-0.5, -0.1)
Race	African American	-0.4 (0.72)	-0.6 (0.87)	-0.7 (0.69)	-0.3 (-0.7, 0.1)	-0.3 (-0.7, 0.1)
	Non-African American	-0.5 (0.64)	-0.6 (0.72)	-0.8 (0.71)	-0.1 (-0.3, 0.0)	-0.3 (-0.5, -0.1)
	White	-0.5 (0.64)	-0.6 (0.72)	-0.8 (0.71)	-0.1 (-0.3, 0.0)	-0.3 (-0.5, -0.1)
	Non-White	-0.4 (0.70)	-0.6 (0.87)	-0.6 (0.70)	-0.3 (-0.7, 0.2)	-0.3 (-0.7, 0.2)
BMI	≥18.5 kg/m^2 to < 25 kg/m^2	-0.4 (0.52)	-0.7 (0.72)	-0.8 (0.75)	-0.4 (-0.7, -0.1)	-0.4 (-0.7, -0.1)
	≥25 kg/m^2 to < 30 kg/m^2	-0.5 (0.60)	-0.6 (0.79)	-0.7 (0.67)	-0.1 (-0.3, 0.2)	-0.2 (-0.4, 0.1)
	≥ 30 kg/m^2	-0.5 (0.82)	-0.5 (0.76)	-0.8 (0.74)	-0.1 (-0.4, 0.3)	-0.4 (-0.7, -0.1)
History of Hysterectomy	Yes	-0.4 (0.12)	-0.6 (0.11)	-0.8 (0.11)	-0.2 (-0.5, 0.1)	-0.4 (-0.7, -0.1)
	No	-0.5 (0.07)	-0.7 (0.07)	-0.8 (0.07)	-0.2 (-0.3, 0.0)	-0.3 (-0.4, -0.1)
History of Oophorectomy	Yes	-0.4 (0.11)	-0.7 (0.11)	-0.7 (0.11)	-0.3 (-0.6, 0.0)	-0.3 (-0.6, -0.0)
	No	-0.5 (0.06)	-0.6 (0.05)	-0.8 (0.05)	-0.1 (-0.3, 0.0)	-0.3 (-0.4, -0.1)
Prior Hormone Therapy Use	Yes	-0.3 (0.13)	-0.6 (0.12)	-1.0 (0.12)	-0.3 (-0.6, 0.1)	-0.7 (-1.0, -0.3)
	No	-0.5 (0.06)	-0.6 (0.06)	-0.7 (0.07)	-0.1 (-0.3, 0.1)	-0.2 (-0.4, -0.0)
Smoking Status	Current Smokers	-0.5 (0.79)	-0.8 (0.81)	-0.8 (0.68)	-0.2 (-0.6, 0.2)	-0.3 (-0.7, 0.1)
	Former/Never Smokers	-0.4 (0.62)	-0.6 (0.73)	-0.7 (0.72)	-0.1 (-0.3, 0.0)	-0.3 (-0.5, -0.1)

Source: Table 15 and End-of-Text Table 9.3.1.7.8, End-of-Text Table 9.3.1.7.4.1, End-of-Text Table 9.3.1.7.4.3 and End-of-Text Table 9.3.1.7.6 from Study Report of SKYLIGHT 2. The LS means, SE, and CI were obtained a MMRM analysis with change from baseline as the dependent variable and treatment group and week as factors, with baseline measurement as covariate, as well as an interaction of treatment by week and an interaction of baseline measurement by week.

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

No statistical issues were found.

5.2 Collective Evidence

The data from studies SKYLIGHT 1 and SKYLIGHT 2 showed that fezolinetant at doses 30 mg and 45 mg demonstrated statistically significant reductions from baseline in the average frequency and severity of moderate to severe VMS per 24 hours at week 4 and at week 12 compared to placebo. However, fezolinetant 30 mg did not meet the clinical meaningful treatment difference of 2 or more reduction compared to placebo in the frequency of VMS.

The effect of the fezolinetant treatment (30 mg and 45 mg) on the VMS frequency was observed as early as week 1 and greater reductions compared to placebo were observed during the 12-week double-blind period.

5.3 Conclusions and Recommendations

Based on the collective efficacy evidence from the two adequate and well controlled trials of SKYLIGHT 1 and SKYLIGHT 2 studies, the reviewer concludes that the application provided substantial evidence of efficacy of fezolinetant 45 mg tablet administered once daily for the treatment of moderate to severe vasomotor symptoms associated with menopause. Although fezolinetant 30 mg was statistically superior to placebo in the mean reduction in the frequency and severity of moderate to severe VMS from baseline at week 4 and at week 12, it did not meet the clinical meaningful treatment difference criterion of 2 or more reduction compared to placebo.

5.4 Labeling Recommendation

In Section 14 of the draft label, the Applicant has proposed to include the following co-primary efficacy endpoint results from the two trials (SKYLIGHT 1 and 2).

(b) (4)

Reviewer's Remark: Overall, the Applicant's proposal to include the above co-primary efficacy endpoint results in Section 14 of the draft label appears reasonable. The reviewer has the following recommendations regarding the results presented in Table 2 and 3:

i) For the treatment difference versus placebo results in Table 2 and 3 above, we recommend that the 95% confidence intervals be presented (b) (4) for the treatment differences.

ii) (b) (4)

iii) The title of Table 3 be revised to '*Mean Baseline and Change from Baseline to weeks 4 and 12 for Mean Severity of ~~Moderate to Severe~~ VMS* (b) (4) *24 Hours in Women Treated with* (b) (4) *in* (b) (4) *1 and 2*' because the analysis of severity of VMS at the post-baseline visits also included 'mild' severity.

APPENDICES

Sensitivity Results

Table 18. Discontinuation-reason Based Multiple Imputation Sensitivity Analysis of mean frequency of VMS, change from baseline and difference from placebo; 12-week Double-blind Period

	Placebo	Fezolinetant		LS Means Difference (95% CI) vs. Placebo	
		30 mg	45 mg	30 mg	45 mg
SKYLIGHT 1					
N	175	173	174		
Baseline	10.5 (3.79)	10.7 (4.73)	10.4 (3.92)		
Change to week 4	-3.3 (4.18)	-5.3 (5.53)	-5.2 (4.06)	-1.8 (-2.5, -1.2)	-2.1 (-2.7, -1.4)
Change to week 12	-3.7 (4.18)	-6.2 (6.05)	-6.3 (4.45)	-2.2 (-3.0, -1.5)	-2.5 (-3.2, -1.8)
SKYLIGHT 2					
N	167	166	167		
Baseline	11.6 (5.02)	11.2 (4.88)	11.8 (8.26)		
Change to week 4	-3.6 (4.15)	-5.5 (4.23)	-6.2 (4.77)	-1.8 (2.6, -1.1)	-2.5 (-3.3, -1.8)
Change to week 12	-4.6 (5.14)	-6.4 (4.75)	-7.4 (6.45)	-1.9 (-2.8, -1.0)	-2.5 (-3.4, -1.6)

Source: Table 14 from study reports SKYLIGHT 1 and SKYLIGHT 2.

Table 19. Discontinuation-reason Based Multiple Imputation Sensitivity Analysis of mean severity of VMS, change from baseline and difference from placebo; 12-week Double-blind Period

	Placebo	Fezolinetant		LS Means Difference (95% CI) vs. Placebo	
		30 mg	45 mg	30 mg	45 mg
SKYLIGHT 1					
N	175	173	174		
Baseline	2.4 (0.35)	2.4 (0.34)	2.4 (0.35)		
Change to week 4	-0.3 (0.50)	-0.4 (0.55)	-0.5 (0.60)	-0.2 (-0.3, -0.1)	-0.2 (-0.3, -0.1)
Change to week 12	-0.4 (0.58)	-0.6 (0.71)	-0.6 (0.74)	-0.2 (-0.4, -0.1)	-0.2 (-0.3, -0.1)
SKYLIGHT 2					
N	167	166	167		
Baseline	2.4 (0.32)	2.4 (0.33)	2.4 (0.34)		
Change to week 4	-0.3 (0.48)	-0.5 (0.33)	-2.4 (0.34)	-0.2 (-0.3, -0.0)	-0.3 (-0.4, -0.2)
Change to week 12	-0.5 (0.65)	-0.6 (0.75)	-0.7 (0.71)	-0.2 (-0.3, -0.0)	-0.3 (-0.4, -0.2)

Source: Table 14 from study reports SKYLIGHT 1 and SKYLIGHT 2.

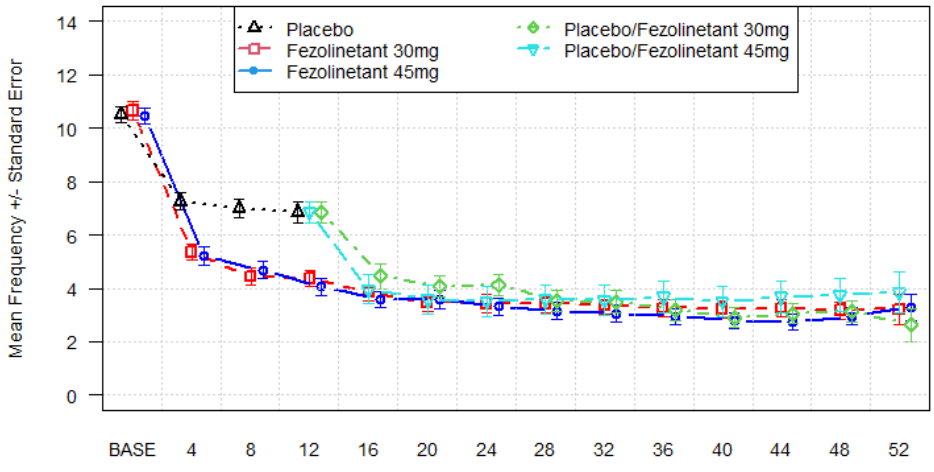
Table 20. Sensitivity Analysis of Co-Primary Endpoints, (Per Protocol Set)

Frequency	SKYLIGHT 1			SKYLIGHT 2		
	Placebo	**Fezolinetant 30 mg**	**Fezolinetant 45 mg**	**Placebo**	**Fezolinetant 30 mg**	**Fezolinetant 45 mg**
Baseline for week 4 (N)	151	141	148	145	138	136
Mean (SD)	10.46 (3.42)	10.79 (5.06)	10.58 (4.09)	11.72 (5.25)	11.53 (5.15)	12.06 (9.02)
Mean (SD) change from baseline	-2.90 (3.65)	-5.29 (5.70)	-5.18 (4.20)	-3.60 (4.20)	-5.34 (4.28)	-5.90 (4.82)
LS Mean (SE) difference from placebo		-2.18 (0.44)	-2.20 (0.43)		-1.77 (0.50)	-2.22 (0.50)
95% CI (2-sided)		(-3.03, -1.32)	(-3.05, -1.36)		(-2.76, -0.79)	(-3.21, -1.23)
MMRM p-value		< 0.001	< 0.001		< 0.001	< 0.001
Baseline for week 12 (N)	132	116	121	126	112	123
Mean (SD)	10.48 (3.55)	10.79 (5.25)	10.72 (4.17)	11.56 (4.69)	11.39 (5.32)	12.29 (9.44)
Mean (SD) change from baseline	-3.46 (4.11)	-6.03 (6.06)	-6.33 (4.77)	-4.38 (5.29)	-6.04 (4.61)	-7.37 (6.92)
LS Mean (SE) difference from placebo		-2.36 (0.50)	-2.70 (0.50)		-1.72 (0.65)	-2.72 (0.64)
95% CI (2-sided)		(-3.36, -1.37)	(-3.68, -1.72)		(-3.01, -0.44)	(-3.97, -1.46)
MMRM p-value		< 0.001	< 0.001		0.009	< 0.001
Severity	**SKYLIGHT 1**			**SKYLIGHT 2**		
	Placebo	**Fezolinetant 30 mg**	**Fezolinetant 45 mg**	**Placebo**	**Fezolinetant 30 mg**	**Fezolinetant 45 mg**
Baseline for week 4 (N)	151	141	148	145	138	136
Mean (SD)	2.41 (0.35)	2.38 (0.34)	2.41 (0.36)	2.42 (0.32)	2.44 (0.32)	2.40 (0.34)
Mean (SD) change from baseline	-0.23 (0.41)	-0.40 (0.53)	-0.43 (0.57)	-0.30 (0.45)	-0.40 (0.50)	-0.51 (0.53)
LS Mean (SE) difference from placebo		-0.18 (0.06)	-0.20 (0.06)		-0.10 (0.06)	-0.21 (0.06)
95% CI (2-sided)		(-0.29, -0.06)	(-0.31, -0.08)		(-0.22, 0.02)	(-0.33, -0.09)
MMRM p-value		0.003	< 0.001		0.09	< 0.001
Baseline for week 12 (N)	132	116	121	126	112	123
Mean (SD)	2.41 (0.36)	2.37 (0.34)	2.42 (0.37)	2.42 (0.31)	2.46 (0.33)	2.41 (0.34)
Mean (SD) change from baseline	-0.33 (0.56)	-0.52 (0.66)	-0.51 (0.68)	-0.41 (0.62)	-0.48 (0.68)	-0.69 (0.65)
LS Mean (SE) difference from placebo		-0.21 (0.08)	-0.18 (0.08)		-0.06 (0.08)	-0.28 (0.08)
95% CI (2-sided)		(-0.36, -0.06)	(-0.33, -0.03)		(-0.22, 0.11)	(-0.44, -0.12)
MMRM p-value		0.007	0.019		0.501	< 0.001

Source: Table 13, Study Reports for SKYLIGHT 1 and SKYLIGHT 2. The LS means, SE, CI and p-values were obtained a MMRM analysis with change from baseline as the dependent variable and treatment group, week, and smoking status (current vs former/never) as factors, with baseline measurement and baseline weight as covariates, as well as an interaction of treatment by week and an interaction of baseline measurement by week.

Mean Frequency and Severity by treatment. 52-week period

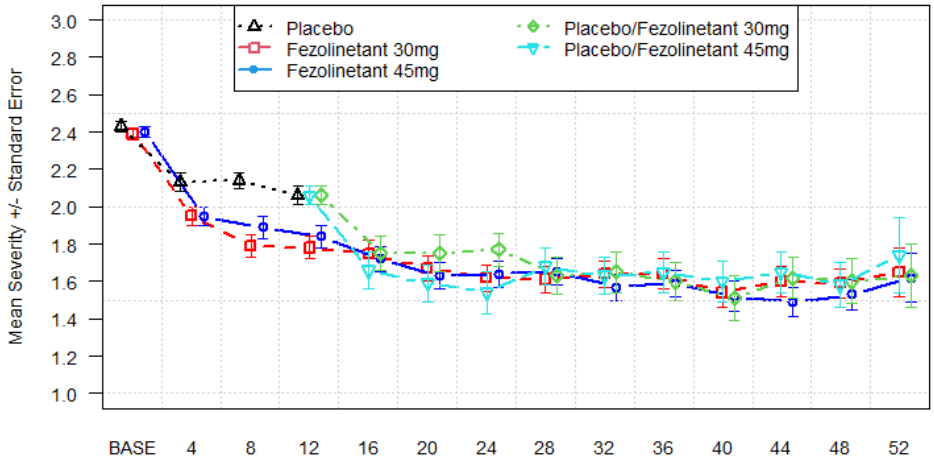
Figure 6. Mean Frequency of VMS by treatment arm until week 52 (SKYLIGHT 1)



Number of subjects	BASE	4	8	12	16	20	24	28	32	36	40	44	48	52
Placebo	175	166	152	139										
Fezolinetant 30mg	173	157	143	134	130	130	121	119	121	116	118	113	115	50
Fezolinetant 45mg	174	164	151	147	145	140	138	138	141	138	136	136	129	49
Placebo/Fezolinetant 30mg					72	69	66	65	60	60	61	55	54	21
Placebo/Fezolinetant 45mg					68	68	68	64	64	62	63	60	57	21

Source: Figure 7, clinical study report for SKYLIGHT 1.

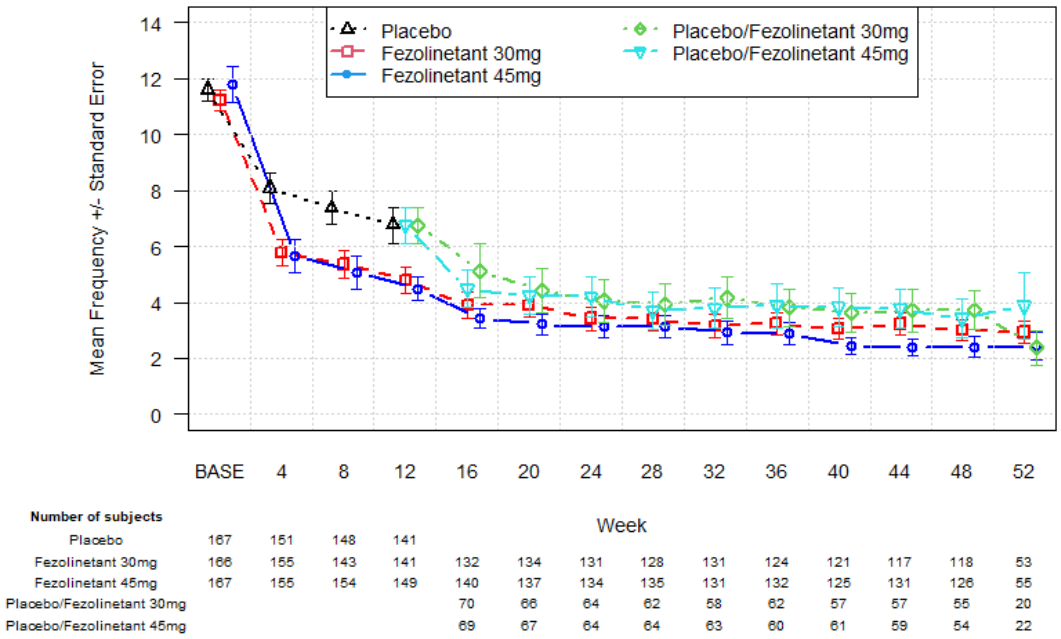
Figure 7. Mean Severity of VMS by treatment arm until week 52 (SKYLIGHT 1)



Number of subjects	BASE	4	8	12	16	20	24	28	32	36	40	44	48	52
Placebo	175	166	152	139										
Fezolinetant 30mg	173	157	143	134	130	130	121	119	121	116	118	113	115	50
Fezolinetant 45mg	174	164	151	147	145	140	138	138	141	138	136	136	129	49
Placebo/Fezolinetant 30mg					72	69	66	65	60	60	61	55	54	21
Placebo/Fezolinetant 45mg					68	68	68	64	64	62	63	60	57	21

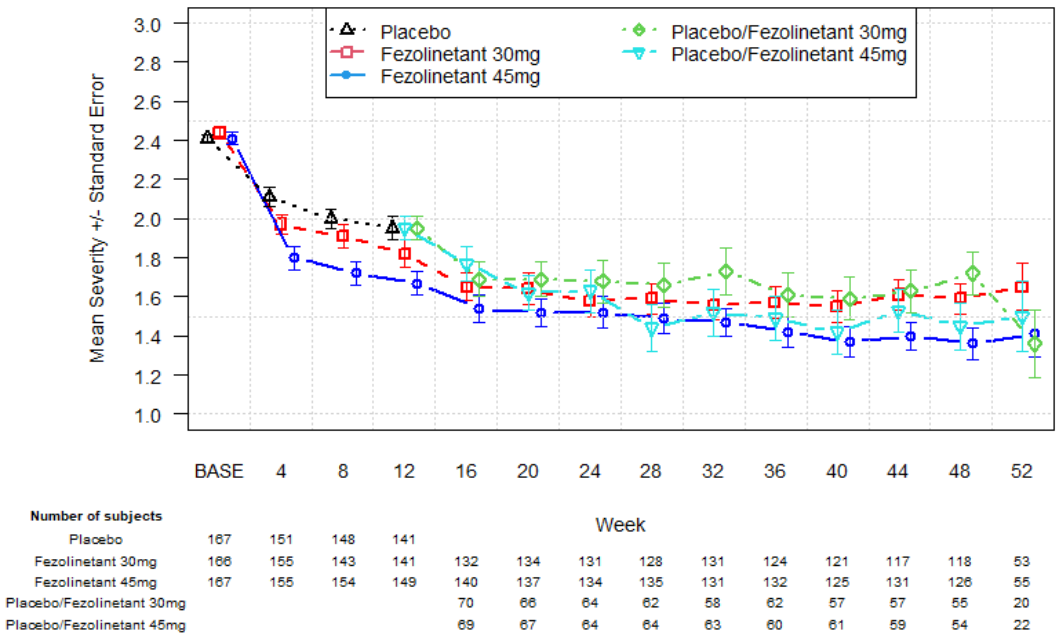
Source: Figure 8, clinical study report for SKYLIGHT 1.

Figure 8. Mean Frequency of VMS by treatment arm until week 52 (SKYLIGHT 2)



Source: Figure 7, clinical study report for SKYLIGHT 2.

Figure 9. Mean Severity of VMS by treatment arm until week 52 (SKYLIGHT 2)



Source: Figure 8, clinical study report for SKYLIGHT 2.

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

JUAN C VIVAR
11/21/2022 03:30:30 PM

SOLOMON CHEFO
11/21/2022 03:54:50 PM

TSAE YUN D LIN
11/21/2022 05:01:23 PM

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Science
Office of Biostatistics

Statistical Review and Evaluation

CARCINOGENICITY STUDIES

IND/NDA Number:	NDA 216578
Drug Name:	ESN364
Indication:	Treatment of Moderate to Severe Vasomotor Symptoms Associated with Menopause
Studies:	Carcinogenicity Studies in Rats for 104 weeks and Mice for 26 weeks
Applicant:	Study Sponsor: Ogeda SA 47 Rue Adrienne Bolland Gosselies 6041 BELGIUM Testing Facility: (b) (4)
Documents Reviewed:	Electronic submission: Submitted on August 23 2022 Electronic data: Submitted on August 23 2022
Review Priority:	Standard
Biometrics Division:	Division of Biometrics - VI
Statistical Reviewer:	Dr. Hepei Chen
Concurring Reviewer:	Dr. Karl Lin
Medical Division:	Division of Pharm-Tox for Rare Diseases, Pediatrics, Urologic and Reproductive Medicine (DPT-RPURM)
Reviewing Pharmacologist:	Dr. Miyun Tsai-Turton
Keywords:	Carcinogenicity, Dose response

Table of Contents

1. Background

In this submission the sponsor included reports of two animal carcinogenicity studies, one in rats and one in mice. These studies were to evaluate the carcinogenic potential and determine the toxicokinetics of the test article, ESN364, when administered daily via oral gavage to female Wistar Han rats for at least 104 weeks, and to male and female 001178-T (hemizygous) rasH2 mice for at least 26 weeks.

In this review the phrase "dose response relationship" refers to the linear component (trend) of the effect of treatment, and not necessarily to a strictly increasing or decreasing mortality or tumor incidence rate as dose increases.

2. Rat Study

As indicated in Table 1, one experiment in female rats was conducted with three treated groups, and one vehicle control group. Four hundred Wistar Han rats were assigned randomly in size of 100 rats per group. The dose levels for the three treated groups were 10, 30, and 100 mg/kg/day ESN364, respectively. In this review these dose groups were referred to as the low (Group 2), mid (Group 3), and high (Group 4) dose groups, respectively. The rats in the vehicle control groups were administrated with 0.5% (w/v) methylcellulose (400 cps) in reverse osmosis (RO) water, and handled for the same duration and in the same manner as the treated groups.

Table 1: Experimental Design in Female Rat Study

Group No.	No. of Toxicity Animals	Test Material	Dosage Level (mg/kg/day)
1	100	Vehicle control	0
2	100	ESN364 Low	10
3	100	ESN364 Mid	30
4	100	ESN364 High	100

Cageside observations were conducted for carcinogenicity animals once on Day 173 of the dosing phase. Detailed observations were conducted for each animal once during the predose phase and for each carcinogenicity animal prior to dosing on Day 1 and weekly (based on Day 1) throughout the dosing phase to Week 104 and on Day 728 of the dosing phase. Detailed observations were also collected for each carcinogenicity animal on days of scheduled sacrifice (all surviving animals). On each day of dosing beginning on Day 598 of the dosing phase, cageside observations were conducted for each carcinogenicity animal at approximately 3 hours postdose. A necropsy was conducted on multiple animals that were found dead or sacrificed at an unscheduled interval throughout the dosing phase. After at least 104 weeks of dosing, all surviving carcinogenicity animals/sex/group (dependent on survival), having been fasted overnight, were anesthetized with isoflurane inhalation, exsanguinated, and necropsied. Tissues indicated in the previous table (Necropsy and Macroscopic Observations section) from the animals that died or were sacrificed at scheduled or unscheduled interval were examined microscopically by the Principal Investigator for Anatomic Pathology.

2.1. Sponsor's analyses

2.1.1. Survival analysis

In the sponsor's report, tests to compare survival were performed, with a two-sided risk for increasing and decreasing mortality with dose. Tests were performed for dose response along with pairwise comparisons for each dosed group against the control group using Kaplan-Meier product-limit estimates (Kaplan and Meier, 1958), along with log-rank and Wilcoxon tests (Peto et al., 1980). These were performed using the LIFETEST procedure in SAS. The time to death or euthanasia (in weeks) was the dependent variable. The treatment group was included as the stratum. Animals with a death or euthanasia status recorded as a planned euthanasia or an accidental death were censored in the analysis. For the analysis of tumors, animals that were classified as accidental deaths were included in the incidental tables in the week indicated (in the same way as animals that died of natural causes).

Sponsor's findings:

The sponsor's analysis showed that the numbers of rats surviving to their terminal necropsy were 59 (59%), 65 (65%), 60 (60%), and 65 (65%) in the vehicle control, the low, mid, and high dose groups for female rats, respectively. There were no statistically significant differences in the incidence of mortality amongst the groups.

2.1.2. Tumor data analysis

In the sponsor's analysis, only tumors from tissues listed in the Protocol to be examined for all animals were analyzed. Statistical analysis was conducted for all such tumor types. Tests to compare tumor incidence were performed, with a one-sided risk for increasing incidence with dose. Tests were performed for dose response along with pairwise comparisons for each dosed group against the control group.

For tumors occurring in animals dying spontaneously or euthanized in a moribund condition, the Pathologist classified the context of observation as one of the following: 1. Fatal: The tumor was a factor in the demise of the animal. 2. Non-fatal: The tumor was not a factor in the demise of the animal. 3. Uncertain.

Occult or non-palpable tumors were analyzed by the International Agency for Research on Cancer (IARC) asymptotic fixed interval-based prevalence test (Peto et al., 1980). The cut off points for the interval based test were Weeks 0 to 52, 53 to 78, 79 to 92, 93 to before terminal euthanasia, and the terminal euthanasia. Fatal and non-fatal tumors were analyzed together, with a separate stratum for each. No tumors of uncertain context were noted. The test was implemented using PROC MULTTEST in the SAS system. In the case of sparse tables (10 total in a stratum), the exact form of the test was used for that stratum. Otherwise, the asymptotic version of the test was used. Animals were assigned to the terminal euthanasia stratum based on the death or euthanasia status recorded in the data and were not assigned based on their week of necropsy. Animals dying (with a cause of death of "Natural death or moribund sacrifice" or "Accidental death") after the initiation of terminal euthanasia for that group were classified as a

terminal euthanasia animal and had all tumors classified as incidental for the purposes of statistical analysis.

Observable or palpable tumors in the mammary gland or skin/subcutis were analyzed using the methods previously described for analyzing survival, using the time to death or time of detection of the tumor (in weeks) as a surrogate for the tumor onset time. The comparisons between the control and dosed groups were performed with a one-sided risk for increasing incidence with dose.

Unadjusted P-values were reported for tumors. The trend test was used for interpretation of the results; the pairwise test results were only produced to assist in determining a no observed effect level and were not utilized to determine the significance or non-significance of any tumor. Site or tumor combinations were statistically analyzed. The criteria for combination were based on Guidelines for combining neoplasms for evaluation of rodent carcinogenicity studies (McConnell et al., 1986) and as indicated by the Study Pathologist. Incidences of multiple-organ and combined neoplastic findings, such as hemangioma, fibrosarcoma, and endometrial stromal polyp, were counted by animal, not tissue type.

Adjustment for multiple testing:

In the sponsor's report, indication of a possible treatment effect were assessed on the basis of rare or common tumor type, in line with the current FDA guidelines (Food and Drug Administration Draft Guidance for Industry, 2001) and Expanded Statistical Decision Rules for Interpretations of Results of Rodent Carcinogenicity Studies of Pharmaceuticals (Lin and Rahman, 2018. The Study Pathologist determined whether a tumor type was rare or common. If the concurrent control rate in the study is >1%, the tumor was considered common, regardless of background historical rates.

Sponsor's findings:

In the sponsor's report, the Table 2 presents tumors observed which had statistically significant dose response results when evaluated at the 5% level.

Table 2. Significant Findings (p-value ⩽ 0.05) of the Pairwise Comparison Results

Tissue and Lesion	Test (Group)	Unadjusted p-value
Ovary Combined B-Adenoma, rete ovarii/B-Cystadenoma/ B-Luteoma/M-Carcinoma/M-Dysgerminoma/ M-Granulosa cell tumor, malignant/ M-Malignant granulosa/theca cell tumor	Mid v Control (Group 3 v Group 1)	0.0254
Thymus Combined B-Thymoma/M-Malignant thymoma	Trend (Groups 1,2,3,4)	0.0396
Thyroid B-Adenoma, follicular cell	Trend (Groups 1,2,3,4)	0.0075
Thyroid Combined B-Adenoma, follicular cell/ M-Carcinoma, follicular cell	Trend (Groups 1,2,3,4)	0.0277

2.2. Reviewer's analyses

To verify the sponsor's analyses and to perform additional analyses suggested by the reviewing toxicologist, this reviewer independently performed the survival and tumor data analyses using the data provided by the sponsor electronically.

2.2.1. Survival analysis

In the reviewer's analysis, the survival distributions of rats in all four groups (Groups 1, 2, 3, and 4) were estimated using the Kaplan-Meier product limit method. The dose response relationship was tested across Groups 1, 2, 3, and 4 using the likelihood ratio test, and the homogeneity of survival distributions was tested using the log-rank test. The Kaplan-Meier curves for survival rates are given in Figures 1 in the appendix for all four groups in female rats. The intercurrent mortality data of all four groups and the results of the tests for dose response relationship and homogeneity of survivals for Groups 1, 2, 3, and 4 are given in Tables 1 in the appendix for female rats, respectively.

Reviewer's findings:

The reviewer's analysis showed that the numbers of female rats surviving to their terminal necropsy were 59 (59%), 65 (65%), 60 (60%), and 65 (65%) in the vehicle control, the low, mid, and high dose groups, respectively. No statistically significant dose response relationship and pairwise comparisons in mortality was noted for female rats.

2.2.2. Tumor data analysis

The tumor data were analyzed for dose response relationships across the vehicle control group, and low, mid, and high dose groups, and pairwise comparisons of each of the three treated groups against the vehicle control group, using the Poly-k method described in the paper of Bailer and Portier (1988) and Bieler and Williams (1993).

In the ploy-k method, the adjustment for differences in mortality among treatment groups is made by modifying the number of animals at risk in the denominators in the calculations of overall tumor rates in the Cochran-Armitage test to reflect less-than-whole-animal contributions for animals that die without tumor before the end of the study (Bailer and Portier 1988). The modification is made by defining a new number of animals at risk for each treatment group. The number of animals at risk for the *i*-th treatment group R^* i is defined as $R^* i = \sum W ij$ where w *ij* is the weight for the *j*-th animal in the *i*-th treatment group, and the sum is over all animals in the group.

Bailer and Portier (1988) proposed the weight w *ij* as follows:

w*ij* = 1 to animals dying with the tumor, and

w*ij* = (t*ij* / tsacr)^3 to animals dying without the tumor,

where t*ij* is the time of death of the *j*-th animal in the *i*-th treatment group, and tsacr is the planned (or intended) time of terminal sacrifice. The above formulas imply that animals living up to the end of the planned terminal sacrifice date without developing any tumor will also be

assigned w*ij* =1 since t*ij* = tsacr. Also animals developed the tumor type being tested before the end of the study will be assigned as w*ij* = 1.

Certain treatment groups of a study or the entire study may be terminated earlier than the planned (or intended) time of terminal sacrifice due to excessive mortalities. However, based on the principle of the Intention-to-treat (ITT) analysis in randomized trials, the tsacr should not be affected by the unplanned early terminations. The tsacr should always be equal to the planned (or intended) time of terminal sacrifice. For those animals that were sacrificed later than tsacr, regardless their actual terminal sacrifice time, tsacr was used as their time of terminal sacrifice in the analysis.

One critical point for Poly-k test is the choice of the appropriate value of k, which depends on the tumor incidence pattern with the increased dose. For long term 104 week standard rat and mouse studies, a value of k=3 is suggested in the literature. Hence, this reviewer used k=3 for the analysis of this data.

Multiple testing adjustment:

For the adjustment of multiple testing, this reviewer used the methodologies suggested in the FDA guidance for statistical design and analysis of carcinogenicity studies (2001). For dose response relationship tests, the guidance suggests the use of test levels of α=0.01 for common tumors and α=0.05 for rare tumors for a submission with one two-year study in one species and one short-term study with another species, in order to keep the overall false-positive rate at the nominal level of approximately 10%. For multiple pairwise comparisons of treated group with control group, however, the guidance indicated that the corresponding multiple testing adjustment is still under development and not yet available. To be conservative, the test level of α=0.05 was used for pairwise comparisons of treated group with control group for both rare and common tumors in this study.

It should be noted that the FDA guidance for multiple testing for dose response relationship is based on a publication by Lin and Rahman (1998). In this work the authors investigated the use of this rule for Peto analysis. However, in a later work Rahman and Lin (2008) showed that this rule for multiple testing for dose response relationship is also suitable for Poly-k tests.

A rare tumor is defined as one in which the published spontaneous tumor rate is less than 1%. However, if the background information for the common or rare tumor is not available, the number of animals bearing tumors in the vehicle control group in the present study was used to determine the common or rare tumor status in the review report. That is, if the number of animals bearing tumors in the vehicle control group is 0, then this tumor is considered as the rare tumor; otherwise, if the number of animals bearing tumors in the control group is greater than or equal to 1, then this tumor is considered as the common tumor.

Reviewer's findings:

The tumor rates and the p-values of the tested tumor types are listed in Tables 2 in the appendix for female rats. The tumor types with p-values less than or equal to 0.05 for dose response

relationship and/or pairwise comparisons of treated groups and vehicle control are reported in Table 3.

Based on the criteria of adjustment for multiple testing discussed above, a statistically significant dose response relationship was noted in the incidence of adenoma follicular cell (p-value = 0.0094), without corresponding statistically significant increases in the high dose group (Group 4) when compared to the vehicle control group (Group 1), if this tumor is considered to be the common tumor. Also a statistically significant increase for the incidence of combined B-Adenoma, Rete Ovarii / B-Adenoma, Tubulostromal / B-Cystadenom / M-Carcinoma in ovary (p-value = 0.0132) was noted in the mid dose group (Group=3) when compared to the vehicle control group (Group 1), without corresponding statistically significant dose response relationship regardless the tumor classification (rare or common). No other statistically significant findings were noted in tumor data for female rats.

Table 3: Summary Table of Tumor Types with P-Values ≤ 0.05 for Dose Response Relationship and/or Pairwise Comparisons of Treated Groups and Vehicle control Group in Female Rats

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	10 mg Low (L) P - C vs. L	30 mg Mid (M) P - C vs. M	100 mg High (H) P - C vs. H
Ovary	B-Adenoma, Rete Ovarii/B-Adenoma, Tubulostromal/B-Cystadenom /M-Carcinoma	0/100 (84) 0.3606	0/100 (84) NC	6/100 (82) 0.0132 $	1/100 (84) 0.5000
	B-Luteoma/M-Granulosa Cell Tumor, Mal*/M-Malignant Granulosa/Theca*	0/100 (84) 0.2964	1/100 (84) 0.5000	1/100 (82) 0.4940	1/100 (84) 0.5000
Thymus	B-Thymoma/ M-Malignant Thymoma	5/98 (82) 0.0401 @	5/99 (84) 0.3884	8/99 (82) 0.2824	11/99 (84) 0.1025
Thyroid	B-Adenoma, Follicular Cell	3/100 (84) 0.0094 $	2/100 (84) 0.5000	4/100 (82) 0.4866	9/100 (85) 0.0685
	B-Adenoma, Follicular Cell/ M-Carcinoma, Follicular	6/100 (84) 0.0299 @	2/100 (84) 0.8615	5/100 (82) 0.4832	10/100 (85) 0.2232

[&] X/ZZ (YY): X=number of tumor bearing animals; YY=mortality weighted total number of animals; ZZ=unweighted total number of animals observed; NC = Not calculable.
$ = Statistically significant at 0.01 level in common tumor for test of dose response relationship, and at 0.05 level in rare or common tumor for test of pairwise comparisons;
@ = Not statistically significant at 0.01 level in common tumor for test of dose response relationship;

3. Mouse Study

Table 4: Experimental Design in Mouse Study

Group No.	No. of Toxicity Animals		Test Material	Dosage Level (mg/kg/day)	
	Male	Female		Male	Female
1	25	25	Vehicle control	0	0
2	25	25	ESN364 Low	50	50
3	25	25	ESN364 Mid	150	150
4	25	25	ESN364 High	450	450
5	15	15	Positive control	0	0

Two separate experiments, one in male mice and one in female mice were conducted. As indicated in Table 4, in the experiment there were three treated groups, one vehicle control

group, and one positive control group. One hundred and ten 001178-W (wild type) rasH2 mice of each sex were assigned randomly in size of 25 mice per group except for the positive control group which contains 10 mice. Three treated groups received once daily oral gavage at 50, 150, and 450 mg/kg/day of ESN364 for a minimum of 26 weeks for both male and female mice. In this review these three treated groups were referred to as the low (Group 2), mid (Group 3), and high (Group 4) dose groups, respectively. The mice in the vehicle control group (Group 1) were administrated once daily with oral gavage at 0.5% (w/v) methylcellulose (400 cps) in reverse osmosis water, and handled for the same duration and in the same manner as the treated groups. The mice in the positive control group (Group 5) were administered one intraperitoneal dose of 75 mg/kg N-methyl-N-nitrosourea (MNU) on Day 1 of the dosing phase and served as positive controls.

The same clinical examinations, laboratory investigations and pathology procedures used in the rat study were performed in the mouse study.

3.1. Sponsor's analyses

Because the mouse study was conducted by the same testing facility as the rat study, the sponsor used the same methodologies that were used for the analyses of the rat survival and tumor data.

3.1.1. Survival analysis

Sponsor’s findings:

The sponsor’s analysis showed that the numbers of mice surviving to their terminal necropsy were 24 (96%), 24 (96%), 25 (100%), and 24 (96%) in the vehicle control group, the low, mid, and high dose groups for male mice, respectively, and 23 (92%), 23 (92%), 25 (100%), and 24 (96%) for female mice respectively. No statistically significant dose response relationship and pairwise comparisons in mortality was noted for both male and female mice.

3.1.2. Tumor data analysis

Multiple testing adjustment:

The same multiple testing adjustment used in the rat study was used in the mouse study.

Sponsor’s findings:

In the sponsor’s report, for both male and female mice, there were no statistically significant differences in tumor incidence amongst the groups.

3.2. Reviewer's analyses

Similar to the rat study, this reviewer independently performed survival and tumor data analyses of mouse data to verify sponsor’s analyses. Data used in this reviewer's analyses were provided by the

sponsor electronically.

For the analysis of both the survival data and the tumor data in mice, this reviewer used similar methodologies that were used for the analyses of the rat survival and tumor data.

3.2.1. Survival analysis

Reviewer's findings:

The reviewer's analysis showed that the numbers of mice surviving to their terminal necropsy were 24 (96%), 24 (96%), 25 (100%), and 24 (96%) in the vehicle control group, the low, mid, and high dose groups for male mice, respectively, and 23 (92%), 23 (92%), 25 (100%), and 24 (96%) for female mice respectively. No statistically significant findings in mortality were noted in both male and female mice.

3.2.2. Tumor data analysis

Reviewer's findings:

The tumor rates and the p-values of the tested tumor types are listed in Tables 4A and 4B in the appendix for male and female mice, respectively. No statistically significant finding was noted in the reviewer's analysis for both male and female mice.

4. Summary

In this submission the sponsor included reports of two animal carcinogenicity studies, one in rats and one in mice. These studies were to evaluate the carcinogenic potential and determine the toxicokinetics of the test article, ESN364, when administered daily via oral gavage to female Wistar Han rats for at least 104 weeks, and to male and female 001178-T (hemizygous) rasH2 mice for at least 26 weeks.

Rat Study:

One experiment in female rats was conducted with three treated groups, and one vehicle control group. Four hundred Wistar Han rats were assigned randomly in size of 100 rats per group. The dose levels for the three treated groups were 10, 30, and 100 mg/kg/day ESN364, respectively.

The reviewer's analysis showed that the numbers of female rats surviving to their terminal necropsy were 59 (59%), 65 (65%), 60 (60%), and 65 (65%) in the vehicle control, the low, mid, and high dose groups, respectively. No statistically significant dose response relationship and pairwise comparisons in mortality was noted for female rats.

In the reviewer's analysis, a statistically significant dose response relationship was noted in the incidence of adenoma follicular cell (p-value = 0.0094), without corresponding statistically significant increases in the high dose group (Group 4) when compared to the vehicle control group, if this tumor is considered to be the common tumor. Also a statistically significant

increase for the incidence of combined B-Adenoma, Rete Ovarii / B-Adenoma, Tubulostromal / B-Cystadenom / M-Carcinoma in ovary (p-value = 0.0132) was noted in the mid dose group (Group=3) when compared to the vehicle control group (Group 1), without corresponding statistically significant dose response relationship regardless the tumor classification (rare or common). No other statistically significant findings were noted in tumor data for female rats.

Mouse Study:

Two separate experiments, one in male mice and one in female mice were conducted. In the experiment there were three treated groups, one vehicle control group, and one positive control group. One hundred and ten 001178-W (wild type) rasH2 mice of each sex, respectively, were assigned randomly in size of 25 mice per group except for the positive control group which contains 10 mice. Three treated groups received once daily oral gavage at 50, 150, and 450 mg/kg/day of ESN364 for a minimum of 26 weeks for both male and female mice.

The reviewer's analysis showed that the numbers of mice surviving to their terminal necropsy were 24 (96%), 24 (96%), 25 (100%), and 24 (96%) in the vehicle control group, the low, mid, and high dose groups for male mice, respectively, and 23 (92%), 23 (92%), 25 (100%), and 24 (96%) for female mice respectively. No statistically significant findings in mortality were noted in both male and female mice.

In the reviewer's tumor analysis, no statistically significant finding was noted for both male and female mice.

Dr. Hepei Chen.
Mathematical Statistician

Concur: Dr. Karl Lin.
Team Leader, DBVI

Cc: Archival NDA 216578
Dr. Miyun Tsai-Turton

5. Appendix

Table 1: Intercurrent Mortality Rate in Female Rats

	0 mg/kg/day Vehicle Control		10 mg/kg/day Low		30 mg/kg/day Mid		100 mg/kg/day High	
Week / Type of Death	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %	No. of Death	Cum %
0 - 52			2	2.00	2	2.00	3	3.00
53 - 78	8	8.00	4	6.00	8	10.00	8	11.00
79 - 91	14	22.00	13	19.00	15	25.00	11	22.00
92 - 104	17	39.00	12	31.00	13	38.00	13	35.00
Accidental Death	2	2.00	4	4.00	2	2.00		
Terminal sacrifice	59	59.00	65	65.00	60	60.00	65	65.00
Total	100		100		100		100	

Test	All Dose Groups	Vehicle Control vs. Low	Vehicle Control vs. Mid	Vehicle Control vs. High
Dose-Response (Likelihood Ratio)	0.8916	0.3063	0.9775	0.6191
Homogeneity (Log-Rank)	0.6956	0.3037	0.9774	0.6176

#All Cum. % Cumulative Percentage except for Terminal sacrifice;
* = Significant at 5% level; ** = Significant at 1% level.

Table 2: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Rats

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	10 mg Low (L) P - C vs. L	30 mg Mid (M) P - C vs. M	100 mg High (H) P - C vs. H
Adipose, Other	B-Lipoma	1/100 (84)	0/100 (84)	0/100 (82)	0/100 (84)
		0.7485	0.5000	0.4940	0.5000
	M-Fibrosarcoma	0/100 (84)	0/100 (84)	1/100 (82)	0/100 (84)
		0.2515	NC	0.4940	NC
Adrenal, Cortex	B-Adenoma	0/100 (84)	1/100 (84)	0/100 (82)	1/100 (84)
		0.3134	0.5000	NC	0.5000
Adrenal, Medulla	B-Pheochromocytoma	1/100 (84)	1/100 (84)	2/100 (82)	1/100 (84)
		0.4408	NC	0.4909	NC
Bone, Other	M-Carcinoma, Squamous Cell	1/100 (84)	0/100 (84)	0/100 (82)	0/100 (84)
		0.7485	0.5000	0.4940	0.5000
Brain	B-Granular Cell Tumor	1/100 (84)	0/100 (84)	1/100 (82)	1/100 (84)
		0.3904	0.5000	0.7455	NC
	M-Malignant Granular Cell T*	0/100 (84)	2/100 (85)	1/100 (82)	0/100 (84)
		0.6853	0.2515	0.4940	NC
	B-Granular Cell Tumor/ M-Malignant Granular Cell T*	1/100 (84)	2/100 (85)	2/100 (82)	1/100 (84)
		0.5541	0.5045	0.4909	NC
	B-Meningioma	0/100 (84)	0/100 (84)	0/100 (82)	1/100 (85)
		0.2537	NC	NC	0.5030
	M-Malignant Astrocytoma	0/100 (84)	0/100 (84)	1/100 (82)	0/100 (84)
		0.2515	NC	0.4940	NC
	M-Malignant Oligodendroglio*	0/100 (84)	0/100 (84)	0/100 (82)	2/100 (85)
		0.0638	NC	NC	0.2515
Cervix	B-Endometrial Stromal Tumor	0/100 (84)	1/100 (85)	0/100 (82)	0/100 (84)
		0.4955	0.5030	NC	NC
	B-Polyp, Endometrial Stromal	1/100 (84)	0/100 (84)	2/100 (82)	1/100 (84)
		0.3703	0.5000	0.4909	NC
	B-Endometrial Stromal Tumor/ B-Polyp, Endometrial Stromal	1/100 (84)	1/100 (85)	2/100 (82)	1/100 (84)
		0.4391	0.2515	0.4909	NC
	M-Carcinoma	0/100 (84)	1/100 (85)	1/100 (82)	0/100 (84)
		0.4985	0.5030	0.4940	NC
	M-Schwannoma	3/100 (85)	1/100 (85)	0/100 (82)	0/100 (84)
		0.9800	0.6897	0.8704	0.8750
Colon	B-Lipoma	1/100 (84)	0/100 (84)	0/100 (82)	0/100 (84)
		0.7485	0.5000	0.4940	0.5000

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

Table 2: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Rats

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	10 mg Low (L) P - C vs. L	30 mg Mid (M) P - C vs. M	100 mg High (H) P - C vs. H
Duodenum	M-Carcinoma	0/100 (84) 0.2515	0/100 (84) NC	0/100 (82) NC	1/100 (84) 0.5000
Heart	M-Endocardial Schwannoma	0/100 (84) 0.2515	0/100 (84) NC	0/100 (82) NC	1/100 (84) 0.5000
Hemolympho- Reticular System	M-Histiocytic Sarcoma	1/100 (84) 0.7485	0/100 (84) 0.5000	0/100 (82) 0.4940	0/100 (84) 0.5000
	M-Malignant Lymphoma	3/100 (84) 0.8800	4/100 (86) 0.5131	2/100 (82) 0.4886	1/100 (84) 0.6898
Jejunum	B-Leiomyoma	0/100 (84) 0.2523	0/100 (84) NC	0/99 (81) NC	1/100 (84) 0.5000
Kidney	B-Adenoma, Tubule Cell	0/100 (84) 0.2515	0/100 (84) NC	0/100 (82) NC	1/100 (84) 0.5000
Liver	B-Adenoma, Hepatocellular	2/100 (84) 0.8183	4/100 (84) 0.3409	2/100 (82) 0.6806	1/100 (84) 0.5000
	B-Adenoma, Hepatocholangioc*	0/100 (84) 0.4985	1/100 (85) 0.5030	1/100 (82) 0.4940	0/100 (84) NC
	B-Adenoma, Hepatocellular/ B-Adenoma, Hepatocholang	2/100 (84) 0.8599	5/100 (85) 0.2265	3/100 (82) 0.4886	1/100 (84) 0.5000
	M-Carcinoma, Hepatocellular	0/100 (84) 0.2515	0/100 (84) NC	0/100 (82) NC	1/100 (84) 0.5000
	B-Adenoma, Hepatocellular/ B-Adenoma, Hepatocholang/ M-Carcinoma, Hepatocellular	2/100 (84) 0.6960	5/100 (85) 0.2265	3/100 (82) 0.4886	2/100 (84) NC
Lung	B-Adenoma, Bronchiolo-Alveo*	0/100 (84) 0.2515	0/100 (84) NC	1/100 (82) 0.4940	0/100 (84) NC
Lymph Node, Mesenteric	B-Hemangioma	0/100 (84) 0.5963	1/100 (84) 0.5000	2/100 (82) 0.2425	0/100 (84) NC
	M-Hemangiosarcoma	5/100 (84) 0.8541	0/100 (84) 0.9706	2/100 (82) 0.7684	1/100 (84) 0.8949
	B-Hemangioma/ M-Hemangiosarcoma	5/100 (84) 0.8790	1/100 (84) 0.8949	4/100 (83) 0.4924	1/100 (84) 0.8949

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

Table 2A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Rats (Continued)

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	10 mg Low (L) P - C vs. L	30 mg Mid (M) P - C vs. M	100 mg High (H) P - C vs. H
Mammary Gland	B-Adenoma	1/98 (82)	0/97 (81)	0/97 (79)	0/94 (79)
		0.7445	0.4969	0.4907	0.4907
	M-Carcinoma	12/98 (84)	6/97 (82)	3/97 (80)	0/94 (79)
		1.0000	0.8841	0.9825	0.9998
	B-Adenoma/M-Carcinoma	13/98 (84)	6/97 (82)	3/97 (80)	0/94 (79)
		1.0000	0.9211	0.9898	0.9999
	B-Fibroadenoma	18/98 (84)	17/97 (84)	12/97 (81)	4/94 (79)
		0.9996	0.5000	0.8156	0.9982
Oral Mucosa	M-Malignant Melanoma	0/100 (84)	1/100 (85)	0/100 (82)	0/100 (84)
		0.4955	0.5030	NC	NC
Ovary	B-Adenoma, Rete Ovarii	0/100 (84)	0/100 (84)	1/100 (82)	0/100 (84)
		0.2515	NC	0.4940	NC
	B-Adenoma, Tubulostromal	0/100 (84)	0/100 (84)	3/100 (82)	0/100 (84)
		0.5819	NC	0.1183	NC
	B-Cystadenoma	0/100 (84)	0/100 (84)	0/100 (82)	1/100 (84)
		0.2515	NC	NC	0.5000
	M-Carcinoma	0/100 (84)	0/100 (84)	2/100 (82)	0/100 (84)
		0.4403	NC	0.2425	NC
	B-Adenoma, Rete Ovarii/B-Adenoma, Tubulostromal/B-Cystadenom /M-Carcinoma	0/100 (84)	0/100 (84)	6/100 (82)	1/100 (84)
		0.3606	NC	0.0132 $	0.5000
	B-Luteoma	0/100 (84)	0/100 (84)	1/100 (82)	0/100 (84)
		0.2515	NC	0.4940	NC
	M-Granulosa Cell Tumor, Mal*	0/100 (84)	1/100 (84)	0/100 (82)	0/100 (84)
		0.4970	0.5000	NC	NC
	M-Malignant Granulosa/Theca*	0/100 (84)	0/100 (84)	0/100 (82)	1/100 (84)
		0.2515	NC	NC	0.5000
	B-Luteoma/M-Granulosa Cell Tumor, Mal*/M-Malignant Granulosa/Theca*	0/100 (84)	1/100 (84)	1/100 (82)	1/100 (84)
		0.2964	0.5000	0.4940	0.5000
	M-Dysgerminoma	0/100 (84)	0/100 (84)	1/100 (82)	0/100 (84)
		0.2515	NC	0.4940	NC
	M-Fibrosarcoma	0/100 (84)	0/100 (84)	0/100 (82)	1/100 (84)
		0.2515	NC	NC	0.5000
	M-Schwannoma	0/100 (84)	1/100 (85)	0/100 (82)	0/100 (84)
		0.4955	0.5030	NC	NC

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

Table 2A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Rats (Continued)

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	10 mg Low (L) P - C vs. L	30 mg Mid (M) P - C vs. M	100 mg High (H) P - C vs. H
Pancreas	B-Adenoma, Islet Cell	0/100 (84)	1/100 (85)	0/100 (82)	1/99 (84)
		0.3131	0.5030	NC	0.5000
	M-Carcinoma, Islet Cell	0/100 (84)	0/100 (84)	0/100 (82)	1/99 (83)
		0.2492	NC	NC	0.4970
	B-Adenoma, Islet Cell/ M-Carcinoma, Islet Cell	0/100 (84)	1/100 (85)	0/100 (82)	2/99 (84)
		0.1091	0.5030	NC	0.2485
Pituitary	B-Adenoma, Pars Distalis	56/100 (93)	54/100 (92)	51/100 (90)	39/100 (87)
		0.9883	0.5242	0.6319	0.9725
	B-Adenoma, Pars Intermedia	2/100 (84)	0/100 (84)	1/100 (82)	0/100 (84)
		0.8274	0.7515	0.4909	0.7515
	B-Adenoma, Pars Distalis/B-Adenoma, Pars Intermedi	57/100 (93)	54/100 (92)	52/100 (90)	39/100 (87)
		0.9912	0.5832	0.6306	0.9806
	B-Ganglioneuroma, Benign	0/100 (84)	1/100 (84)	0/100 (82)	0/100 (84)
		0.4970	0.5000	NC	NC
	M-Carcinoma	0/100 (84)	1/100 (84)	1/100 (82)	0/100 (84)
		0.5000	0.5000	0.4940	NC

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

Table 2A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Rats (Continued)

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	10 mg Low (L) P - C vs. L	30 mg Mid (M) P - C vs. M	100 mg High (H) P - C vs. H
Skin/Subcutis	B-Fibroma	1/100 (84)	0/100 (84)	2/100 (82)	1/100 (84)
		0.3703	0.5000	0.4909	NC
	M-Fibrosarcoma	0/100 (84)	1/100 (85)	0/100 (82)	0/100 (84)
		0.4955	0.5030	NC	NC
	M-Fibrosarcoma, Pleomorphic	0/100 (84)	1/100 (85)	0/100 (82)	0/100 (84)
		0.4955	0.5030	NC	NC
	B-Fibroma/M-Fibrosarcoma/M-Fibrosarcoma, Pleomorph	1/100 (84)	2/100 (85)	2/100 (82)	1/100 (84)
		0.5541	0.5045	0.4909	NC
	B-Granular Cell Tumor	1/100 (84)	0/100 (84)	0/100 (82)	0/100 (84)
		0.7485	0.5000	0.4940	0.5000
	B-Keratoacanthoma	0/100 (84)	1/100 (84)	0/100 (82)	0/100 (84)
		0.4970	0.5000	NC	NC
	B-Melanoma, Amelanotic	0/100 (84)	0/100 (84)	1/100 (82)	0/100 (84)
		0.2515	NC	0.4940	NC
	M-Hemangiosarcoma	0/100 (84)	1/100 (84)	0/100 (82)	0/100 (84)
		0.4970	0.5000	NC	NC
	M-Malignant Schwannoma	3/100 (85)	0/100 (84)	0/100 (82)	1/100 (85)
		0.6777	0.8750	0.8704	0.6897
	M-Sarcoma	0/100 (84)	0/100 (84)	0/100 (82)	1/100 (85)
		0.2537	NC	NC	0.5030
Spleen	B-Hemangioma	0/100 (84)	0/100 (84)	0/100 (82)	1/100 (84)
		0.2515	NC	NC	0.5000
	M-Hemangiosarcoma	0/100 (84)	1/100 (84)	0/100 (82)	0/100 (84)
		0.4970	0.5000	NC	NC
	B-Hemangioma/ M-Hemangiosarcoma	0/100 (84)	1/100 (84)	0/100 (82)	1/100 (84)
		0.3134	0.5000	NC	0.5000
Stomach, Nonglandular	B-Papilloma, Squamous Cell	0/100 (84)	0/100 (84)	1/100 (82)	0/100 (84)
		0.2515	NC	0.4940	NC
Thymus	B-Thymoma	5/98 (82)	5/99 (84)	6/99 (82)	10/99 (84)
		0.0584	0.3884	0.5000	0.1506
	M-Malignant Thymoma	0/98 (82)	0/99 (84)	2/99 (82)	1/99 (83)
		0.2023	NC	0.2485	0.5030
	B-Thymoma/ M-Malignant Thymoma	5/98 (82)	5/99 (84)	8/99 (82)	11/99 (84)
		0.0401	0.3884	0.2824	0.1025

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

Table 2A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Rats (Continued)

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	10 mg Low (L) P - C vs. L	30 mg Mid (M) P - C vs. M	100 mg High (H) P - C vs. H
Thyroid	B-Adenoma, C-Cell	10/100 (85)	5/100 (85)	8/100 (82)	5/100 (84)
		0.8163	0.8604	0.5663	0.8550
	M-Carcinoma, C-Cell	0/100 (84)	2/100 (84)	2/100 (82)	0/100 (84)
		0.7207	0.2485	0.2425	NC
	B-Adenoma, C-Cell/ M-Carcinoma, C-Cell	10/100 (85)	7/100 (85)	10/100 (82)	5/100 (84)
		0.8734	0.6948	0.5601	0.8550
	B-Adenoma, Follicular Cell	3/100 (84)	2/100 (84)	4/100 (82)	9/100 (85)
		0.0094	0.5000	0.4866	0.0685
	M-Carcinoma, Follicular Cell	3/100 (84)	0/100 (84)	1/100 (82)	2/100 (84)
		0.4367	0.8772	0.6806	0.5000
	B-Adenoma, Follicular Cell/ M-Carcinoma, Follicular	6/100 (84)	2/100 (84)	5/100 (82)	10/100 (85)
		0.0299	0.8615	0.4832	0.2232
Tooth, Other	M-Carcinoma, Squamous Cell	0/100 (84)	0/100 (84)	0/100 (82)	1/100 (84)
		0.2515	NC	NC	0.5000
	M-Odontoma, Ameloblastic	0/100 (84)	1/100 (85)	0/100 (82)	0/100 (84)
		0.4955	0.5030	NC	NC
	M-Tumor, Periodontal	0/100 (84)	0/100 (84)	0/100 (82)	1/100 (84)
		0.2515	NC	NC	0.5000

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

Table 2A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Rats (Continued)

Organ name	Tumor name	0 mg Vehicle (C) P - Trend	10 mg Low (L) P - C vs. L	30 mg Mid (M) P - C vs. M	100 mg High (H) P - C vs. H
Uterus	B-Adenoma	0/100 (84)	2/100 (84)	1/100 (82)	0/100 (84)
		0.6870	0.2485	0.4940	NC
	M-Carcinoma	6/100 (84)	1/100 (84)	4/100 (82)	5/100 (84)
		0.3291	0.9414	0.6116	0.5000
	B-Adenoma/M-Carcinoma	6/100 (84)	3/100 (84)	5/100 (82)	5/100 (84)
		0.4651	0.7522	0.4832	0.5000
	B-Polyp, Endometrial Stromal	10/100 (84)	4/100 (84)	6/100 (83)	8/100 (85)
		0.4328	0.9195	0.7769	0.6084
	M-Sarcoma, Endometrial Stro*	0/100 (84)	0/100 (84)	0/100 (82)	1/100 (84)
		0.2515	NC	NC	0.5000
	B-Polyp, Endometrial Stromal/ M-Sarcoma, Endometria	10/100 (84)	4/100 (84)	6/100 (83)	9/100 (85)
		0.3155	0.9195	0.7769	0.5110
	M-Fibrosarcoma, Pleomorphic	0/100 (84)	1/100 (85)	0/100 (82)	1/100 (84)
		0.3131	0.5030	NC	0.5000
	M-Hemangiosarcoma	1/100 (84)	0/100 (84)	0/100 (82)	0/100 (84)
		0.7485	0.5000	0.4940	0.5000
	M-Leiomyosarcoma	0/100 (84)	1/100 (85)	0/100 (82)	0/100 (84)
		0.4955	0.5030	NC	NC
	M-Schwannoma	3/100 (84)	0/100 (84)	1/100 (82)	1/100 (84)
		0.6735	0.8772	0.6806	0.6898
Vagina	B-Polyp	1/99 (83)	2/100 (84)	0/100 (82)	1/100 (84)
		0.5779	0.5045	0.4970	0.2515
	B-Tumor, Granular Cell, Ben*	1/99 (83)	0/100 (84)	0/100 (82)	0/100 (84)
		0.7508	0.5030	0.4970	0.5030
Whole body	B-Hemangioma	0/100 (84)	1/100 (84)	2/100 (82)	1/100 (84)
		0.3242	0.5000	0.2425	0.5000
	M-Hemangiosarcoma	6/100 (85)	2/100 (84)	2/100 (82)	1/100 (84)
		0.9496	0.8575	0.8494	0.9395
	B-Hemangioma/ M-Hemangiosarcoma	6/100 (85)	3/100 (84)	4/100 (83)	2/100 (84)
		0.8723	0.7462	0.6118	0.8575
Zymbal Gland	M-Carcinoma	3/100 (86)	0/100 (84)	0/100 (82)	0/100 (84)
		0.9837	0.8728	0.8682	0.8728

& X/YY (ZZ): X=number of tumor bearing animals; YY=unweighted total number of animals observed; ZZ=mortality weighted total number of animals;
NC = Not calculable.

Table 3A: Intercurrent Mortality Rate in Male Mice

	Vehicle Control		50 mg/kg/day Low		150 mg/kg/day Mid		450 mg/kg/day High		Positive Control	
Week / Type of Death	No. of Death	Cum %	No. of Death	Cum %	No. of Death	No. of Death	Cum %	No. of Death	Cum %	No. of Death
0 - 13	1	4.00					1	4.00		
14 - 27			1	4.00					10	
Terminal sacrifice	24	96.00	24	96.00	25	100.00	24	96.00		
Total	25		25		25		25			

Test	All Dose Groups	Vehicle Control vs. Low	Vehicle Control vs. Mid	Vehicle Control vs. High	Vehicle Control vs. Positive Control
Dose-Response (Likelihood Ratio)	0.9560	0.9885	0.2390	0.9885	<.0001**
Homogeneity (Log-Rank)	0.7978	0.9885	0.3173	0.9885	<.0001**

#All Cum. % Cumulative Percentage except for Terminal sacrifice;
* = Significant at 5% level; ** = Significant at 1% level.

Table 3B: Intercurrent Mortality Rate in Female Mice

	Vehicle Control		50 mg/kg/day Low		150 mg/kg/day Mid		450 mg/kg/day High		Positive Control	
Week / Type of Death	No. of Death	Cum %	No. of Death	Cum %	No. of Death	No. of Death	Cum %	No. of Death	Cum %	No. of Death
0 - 13			2	8.00						
14 - 27	2	8.00					1	4.00	10	
Terminal sacrifice	23	92.00	23	92.00	25	100.00	24	96.00		
Total	25		25		25		25			

Test	All Dose Groups	Vehicle Control vs. Low	Vehicle Control vs. Mid	Vehicle Control vs. High	Vehicle Control vs. Positive Control
Dose-Response (Likelihood Ratio)	0.4643	0.9667	0.0935	0.5521	<.0001**
Homogeneity (Log-Rank)	0.5124	0.9667	0.1531	0.5557	<.0001**

#All Cum. % Cumulative Percentage except for Terminal sacrifice;
* = Significant at 5% level; ** = Significant at 1% level.

Table 4A: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Male Mice

Organ name	Tumor name	Vehicle (VC) 0 mg P - Trend	Low (L) 50 mg P - VC vs. L	Mid (M) 150 mg P - VC vs. M	High (H) 450 mg P - VC vs. H	Positive (PC) 0 mg P - VC vs. PC
Duodenum	M-Carcinoma	0/25 (24) NC	0/25 (24) NC	0/25 (25) NC	0/25 (24) NC	1/10 (1) 0.0400
Epididymis	M-Hemangiosarcoma	0/25 (24) 0.7526	1/25 (24) 0.5000	0/25 (25) NC	0/25 (24) NC	0/10 (0) NC
Hemolympho-Reticular System	M-Malignant Lymphoma	0/25 (24) NC	0/25 (24) NC	0/25 (25) NC	0/25 (24) NC	10/10 (10) 0.0000
Kidney	B-Hemangioma	0/25 (24) 0.5052	0/25 (24) NC	1/25 (25) 0.5102	0/25 (24) NC	0/10 (0) NC
Liver	B-Hemangioma	0/25 (24) 0.5052	0/25 (24) NC	1/25 (25) 0.5102	0/25 (24) NC	0/10 (0) NC
Lung	B-Adenoma, Bronchiolo-Alveo*	2/25 (24) 0.5772	3/25 (24) 0.5000	0/25 (25) 1.0000	2/25 (24) NC	0/10 (0) NC
Marrow, Femur	M-Hemangiosarcoma	0/25 (24) 0.7500	1/25 (24) 0.5000	0/24 (24) NC	0/25 (24) NC	0/10 (0) NC
Rectum	M-Hemangiosarcoma	0/25 (24) 0.2474	0/25 (24) NC	0/25 (25) NC	1/25 (24) 0.5000	0/10 (0) NC
Skin/Subcutis	B-Papilloma, Squamous Cell	0/25 (24) NC	0/25 (24) NC	0/25 (25) NC	0/25 (24) NC	4/10 (4) 0.0000
	M-Carcinoma, Squamous Cell	0/25 (24) NC	0/25 (24) NC	0/25 (25) NC	0/25 (24) NC	2/10 (2) 0.0031
	B-Papilloma, Squamous Cell/ M-Carcinoma, Squamous Cell	0/25 (24) NC	0/25 (25) NC	0/25 (25) NC	0/25 (24) NC	4/10 (6) 0.0005
Stomach, Nonglandular	B-Papilloma, Squamous Cell	0/25 (24) NC	0/25 (24) NC	0/25 (25) NC	0/25 (24) NC	7/10 (7) 0.0000
Thymus	B-Thymoma	0/25 (24) 0.1050	1/24 (23) 0.4894	1/25 (25) 0.5102	2/23 (22) 0.2232	0/10 (0) NC

& X/YY (ZZ): X=number of tumor bearing animals; YY=mortality weighted total number of animals; ZZ=unweighted total number of animals observed;
NC = Not calculable.

Table 4B: Tumor Rates and P-Values for Trend and Pairwise Comparisons in Female Mice

Organ name	Tumor name	Vehicle (VC) 0 mg P - Trend	Low (L) 50 mg P - VC vs. L	Mid (M) 150 mg P - VC vs. M	High (H) 450 mg P - VC vs. H	Positive (PC) 0 mg P - VC vs. PC
Bone, Other	M-Hemangiosarcoma	0/25 (23)	1/25 (24)	0/25 (25)	0/25 (24)	0/10 (0)
		0.7604	0.5106	NC	NC	NC
Harderian Gland	B-Adenoma	1/25 (23)	0/25 (23)	0/25 (25)	0/25 (24)	0/10 (0)
		1.0000	1.0000	1.0000	1.0000	NC
Hemolympho-Reticular System	M-Malignant Lymphoma	0/25 (23)	0/25 (23)	0/25 (25)	0/25 (24)	9/10 (9)
		NC	NC	NC	NC	0.0000
Kidney	B-Hemangioma	0/25 (23)	2/25 (23)	0/25 (25)	0/25 (24)	0/10 (0)
		0.8249	0.2444	NC	NC	NC
Lung	B-Adenoma, Bronchiolo-Alveo*	1/25 (23)	0/25 (23)	0/25 (25)	2/25 (24)	0/10 (0)
		0.1562	1.0000	1.0000	0.5163	NC
	M-Carcinoma, Bronchiolo-Alv*	0/25 (23)	1/25 (23)	0/25 (25)	0/25 (24)	0/10 (0)
		0.7579	0.5000	NC	NC	NC
	B-Adenoma, Bronchiolo-Alveo*/ M-Carcinoma, Bronchiolo-Alv*	1/25 (24)	1/25 (23)	0/25 (25)	2/25 (24)	0/10 (3)
		0.2365	0.7447	1.0000	0.5000	1.0000
	M-Mesothelioma	1/25 (24)	0/25 (23)	0/25 (25)	0/25 (24)	0/10 (0)
		1.0000	1.0000	1.0000	1.0000	NC
Ovary	M-Hemangiosarcoma	1/25 (23)	0/25 (23)	0/25 (25)	0/25 (24)	0/10 (0)
		1.0000	1.0000	1.0000	1.0000	NC
Skin/Subcutis	B-Papilloma, Squamous Cell	0/25 (23)	0/25 (23)	0/25 (25)	0/25 (24)	1/10 (1)
		NC	NC	NC	NC	0.0417
	M-Carcinoma, Squamous Cell	0/25 (23)	0/25 (23)	0/25 (25)	1/25 (25)	1/10 (1)
		0.2604	NC	NC	0.5208	0.0417
	B-Papilloma, Squamous Cell/ M-Carcinoma, Squamous Cell	0/25 (24)	0/25 (23)	0/25 (25)	1/25 (25)	2/10 (4)
		0.2577	NC	NC	0.5102	0.0159
Spleen	M-Hemangiosarcoma	1/25 (23)	0/25 (23)	0/25 (25)	2/25 (24)	0/10 (0)
		0.1562	1.0000	1.0000	0.5163	NC
Stomach, Glandular	M-Carcinoma	0/25 (23)	1/25 (23)	0/25 (25)	0/25 (24)	0/10 (0)
		0.7579	0.5000	NC	NC	NC
Stomach, Nonglandular	B-Papilloma, Squamous Cell	0/25 (23)	0/25 (23)	0/25 (25)	0/25 (24)	3/10 (3)
		NC	NC	NC	NC	0.0004
Thymus	B-Thymoma	1/22 (21)	2/25 (23)	2/24 (24)	4/23 (22)	0/9 (0)
		0.0819	0.5349	0.5511	0.1869	NC
	M-Mesothelioma	1/22 (22)	0/25 (23)	0/24 (24)	0/23 (22)	0/9 (0)
		1.0000	1.0000	1.0000	1.0000	NC

& X/YY (ZZ): X=number of tumor bearing animals; YY=mortality weighted total number of animals; ZZ=unweighted total number of animals observed;
NC = Not calculable.

Figure 1: Kaplan-Meier Survival Functions for Female Rats

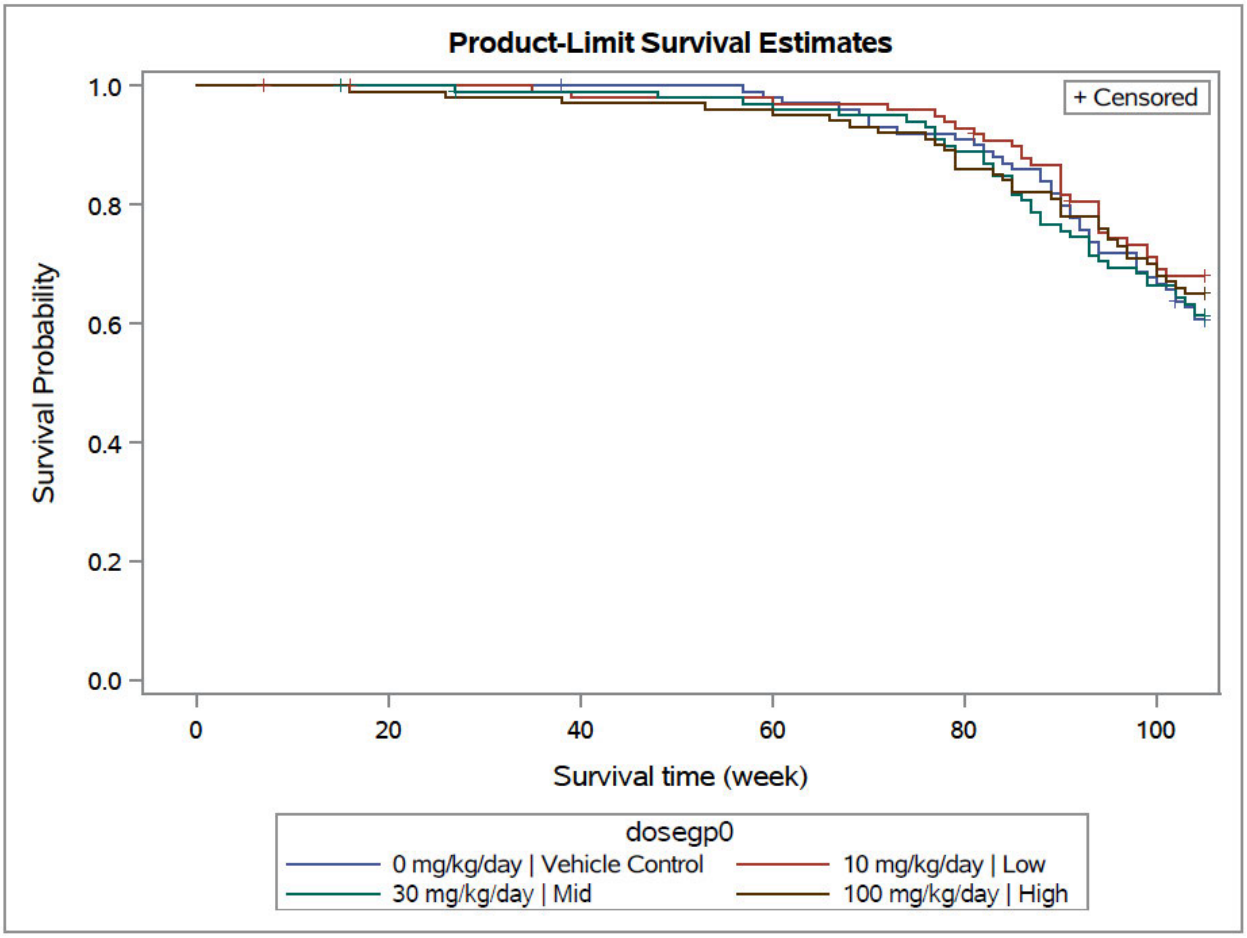


Figure 2A: Kaplan-Meier Survival Functions for Male Mice

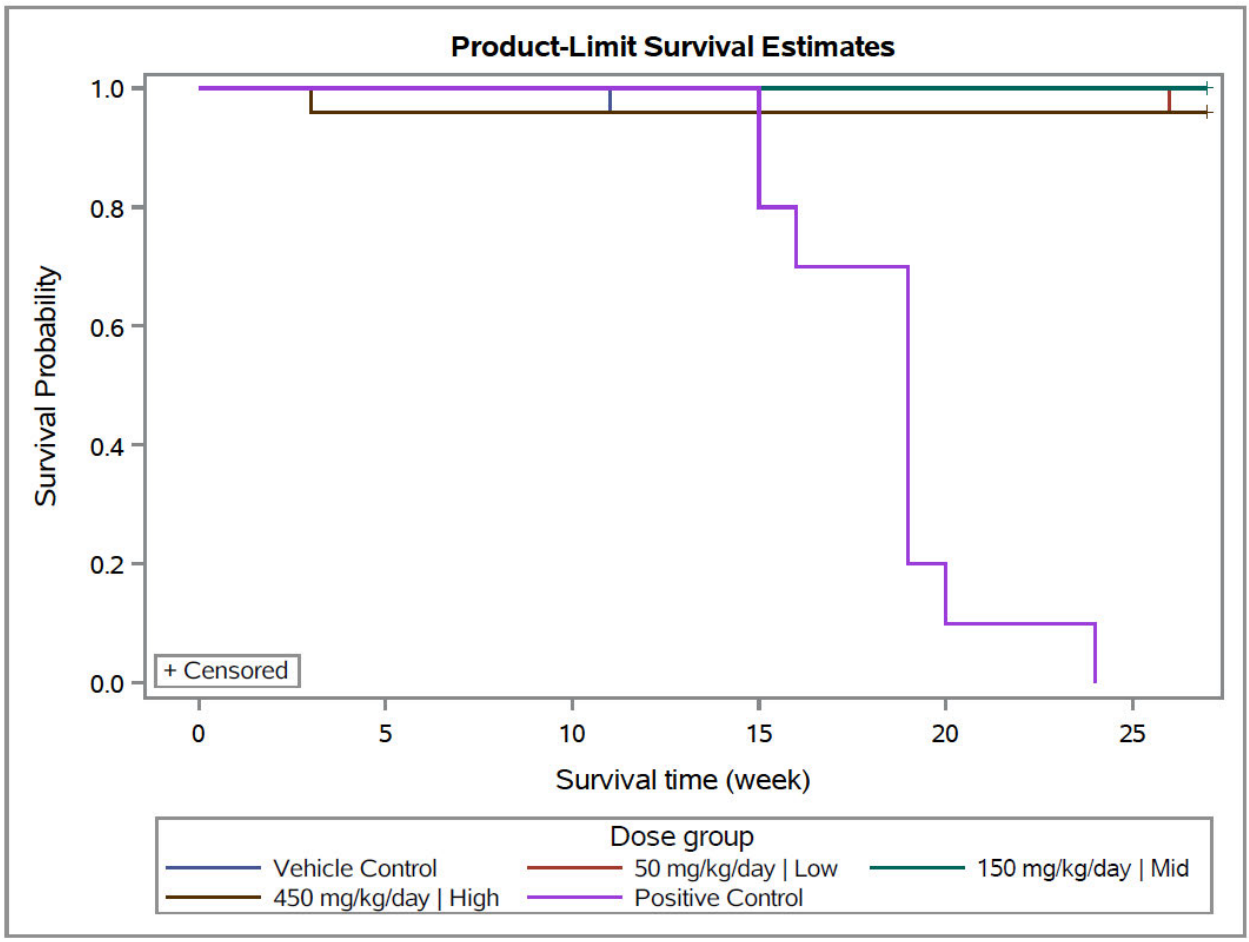
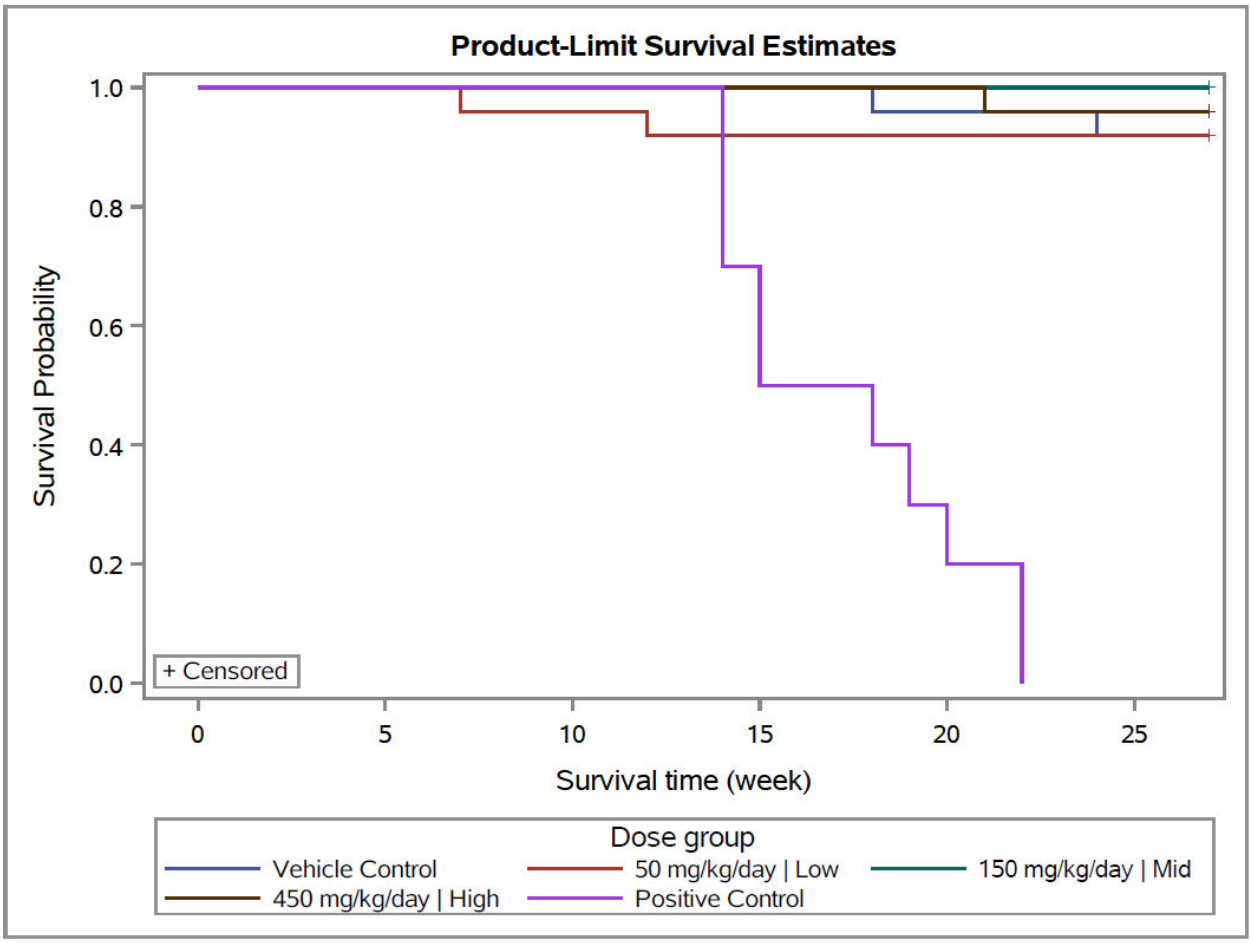


Figure 2B: Kaplan-Meier Survival Functions for Female Mice



6. References

Bailer, A.J, Portier, C.J. (1988). “Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples.” *Biometrics*, 44, 417-431.

Bieler, G.S. and Williams, R.L. (1993). “Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity”. *Biometrics* 49, 793-801.

Guidance for Industry. Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Statues of Pharmaceuticals (Draft Guidance). U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). May 2001.

Mantel N. (1980). Assessing laboratory evidence for neoplastic activity. *Biometrics*. 36: 381-399.

Heyse J.F., Rom D. (1988). Adjusting for Multiplicity of Statistical Tests in the Analysis of Carcinogenicity Studies. *Biometrical Journal*. 30: 883-896.

Harter H.I. (1957). Error Rates and Sample Sizes for Range Tests in Multiple Comparisons. *Biomtrics*. 13: 511-536.

Haseman, J. (1983). “A re-examination of false-positive rates for carcinogenesis studies”, *Fundamental and Applied Toxicology*, 3: 334-339.

Lin K.L. and Ali M.W. (1994). Statistics in the pharmaceutical industry, 2nd ed., Marcel Dekker, pp.19-57.

Lin K.K. (2000) Carcinogenicity Studies of Pharmaceuticals. In: *Encyclopedia of Biopharmaceutical Statistics*, ed. Shein-Chung Chow, Marcel Dekker, New York.

Lin K.K. and Rahman A.M. (1998). ” Overall false positive rates in tests for linear trend in tumor incidence in animal carcinogenicity studies of new drugs”, *Journal of Biopharmaceutical Statistics*, 8(1), 1-15.

Peto, R., M.C. Pike, N.E. Day, R.G. Gray, P.N. Lee, S. Parish, J. Peto, Richards, and J.Wahrendorf. (1980) “Guidelines for sample sensitive significance test for carcinogenic effects in long-term animal experiments”, Long term and short term screening assays for carcinogens: A critical appraisal, International agency for research against cancer monographs, *Annex to supplement, World Health Organization, Geneva,* 311-426.

Rahman, A.M., and Lin, K.K. (2008), "A Comparison of False Positive Rates of Peto and Poly-3 Methods for Long-Term Carcinogenicity Data Analysis Using Multiple Comparison Adjustment Method Suggested by Lin and Rahman", *Journal of Biopharmaceutical Statistics*, 18:5, 849-858.

Rahman, A.M., and Lin, K.K. (2009), "Design and Analysis of Chronic Carcinogenicity Studies of Pharmaceuticals in Rodents", in “Design and Analysis of Clinical Trials with Time-to-Event Endpoints", K.E Peace, Editor, Chapman & Hall/CRC, Taylor & Francis Group, LLC, Boca Raton, FL, London, and New York.

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

HEPEI CHEN
10/26/2022 01:15:46 PM

KARL K LIN
10/26/2022 02:58:09 PM
Concur with review.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

NDA #:	216,578
Drug Name:	(b) (4) (fezolinetant) tablet
Indication(s):	Treatment of moderate to severe vasomotor symptoms associated with menopause
Applicant:	Astellas Pharma US, Inc.
Date(s):	Stamp date: 6/22/2022 Consult date: 8/4/2022 Completion date: 10/20/2022
Safety Endpoints:	Malignancy, bone fracture, glucose level shift to high, liver test elevation, creatine kinase elevation, endometrial safety outcome
Review Priority:	Priority
Biometrics Division:	Division of Biometrics VII
Statistical Reviewer:	Joo-Yeon Lee, PhD
Concurring Reviewer:	Clara Kim, PhD, Team Leader Mat Soukup, PhD, Deputy Division Director
Consulting Division:	Division of Urology, Obstetrics and Gynecology
Consulting Team:	Theresa H Van Der Vlugt, MD Regina Zopf, MD Shelly Slaughter, MD, PhD, Team Leader
Project Manager:	Samantha Bell

Keywords: malignancy, bone fracture, glucose level, liver toxicity, creatine kinase elevation, endometrial safety, exposure-adjusted incidence rates, exposure-adjusted incidence rate difference, risk difference

1 INTRODUCTION

Astellas Pharma US, Inc. (the Applicant) submitted an NDA for fezolinetant, indicated for the treatment of moderate to severe vasomotor symptoms (VMS) associated with menopause. Fezolinetant, a first in class product, is a nonhormonal, selective neurokinin 3 receptor antagonist that blocks neurokinin B binding on the kisspeptin, neurokinin B, and dynorphin neuron to modulate neuronal activity in the thermoregulatory center.

The Division of Urology, Obstetrics and Gynecology (DUOG) consulted the Division of Biometrics VII (DB VII) to assess: i) whether subjects with different exposure duration to study drug can be pooled; and ii) whether the placebo arm from one trial can be compared to active treatment arms from another trial. Additionally, DUOG requested assessment of a potential malignancy safety signal (Section 4).

This review is limited to our responses to the consult questions and corresponding analysis results of selected safety outcomes based on a discussion with DUOG (August 16, 2022). This review includes the summary of the methods used for the safety analyses, the results from the reviewer's analyses and responses to DUOG's questions. For the full safety review of fezolinetant, refer to the clinical review by DUOG.

2 SUMMARY OF SAFETY ANALYSES

2.1 STUDIES

The Applicant's safety analyses included three 52-week phase 3 placebo-controlled randomized double-blind clinical trials and two 12-week phase 2 placebo-controlled trials. Because women are expected to use fezolinetant chronically, per discussion with DUOG, safety was assessed on the three phase 3 trials, summarized in Table 1. Note that long-term (52 weeks) comparative safety assessment was only assessed in study 2693-CL-304.

Table 1: Summary of Study Design and Population in Three Phase 3 Trials.

Study Number, Region(s) Involved, Location within CTD	Study Design	Study Population
2693-CL-0301 US, Canada and Europe Module 5.3.5.1	Randomized, multicenter, placebo-controlled, 12-week double-blind, followed by an active treatment extension period	Female participants suffering from moderate to severe VMS associated with menopause
2693-CL-0302 US, Canada and Europe Module 5.3.5.1	Randomized, multicenter, placebo-controlled, 12-week double-blind, followed by an active treatment extension period	Female participants suffering from moderate to severe VMS associated with menopause
2693-CL-0304 US, Canada and Europe Module 5.3.5.1	Randomized, multicenter, placebo-controlled, double-blind, 52-week long-term safety study	Female participants suffering from VMS associated with menopause

Source: Applicant's report, summary of clinical safety, page 12.

2.2 SAFETY ANALYSIS POOL

The safety population includes all subjects who were randomized and received at least one dose of study drug or placebo. Among the five Applicant-defined pooled safety populations, the populations of interest were the two populations (POP2 and POP4) that included the 52-week phase 3 trial. The pooled safety populations of interest are described as follows:

- POP2: all randomized subjects who received at least one dose of study drug from the three 52-week phase 3 studies (2693-CL-0301, 2693-CL-0302 and 2693-CL-0304). All three studies were placebo-controlled that assessed fezolinetant 30 mg and fezolinetant 45 ng. However, in studies 2693-CL-0301 and 2693-CL-0302, subjects randomized to placebo were re-randomized to either fezolinetant 30 mg or fezolinetant 40 mg at week 12. Therefore, the treatment duration for the two active treatment arms were 40 weeks and 52 weeks.
- POP4: all randomized subjects who received at least one dose of study drug in study 2693-CL-0304, a three-arm, placebo-controlled 52-week phase 3 study. The three arms were placebo, fezolinetant 30 mg, and fezolinetant 45 mg.

Reviewer's comments: *Because of the cross-over study element of studies 2693-CL-301 and 2693-CL-302, comparing placebo and fezolinetant arms at 52 weeks would violate the intention-to-treat (ITT) principle. Furthermore, by design, subjects re-randomized to fezolinetant arm(s) at week 12 knew that they were on treatment, which might make studies 2693-CL-301and 2693-CL-302 inherently too different from study 2693-CL-304 to make comparisons. Therefore, analyses for POP2 are presented for descriptive purposes and only POP4 should be used for comparative assessment at week 52.*

2.3 SAFETY OUTCOMES

Per discussion with DUOG, the safety outcomes of interest were liver test elevations, bone fractures, glucose level shift to high from low or normal at baseline, Creatine Kinase (CK) elevation, endometrial outcomes and malignancy outcomes. The outcomes of liver test elevation, bone fractures, CK elevation and malignancy outcomes were treatment emergent adverse events (TEAE) based on the Medical dictionary for regulatory activities (MedDRA) preferred term (PT). AE data were collected from the signing of informed consent until 21 days after the last dose of the study drug. The outcome of glucose level shift was assessed using laboratory data. Endometrial outcomes were assessed by endometrial biopsy data using the endometrial health (EH) set which was a subset of the safety population. The EH set consists of all randomized subjects who met the following conditions:

- Received at least 1 dose of study drug
- Had the postbaseline biopsy done within 30 days after the last dose of study drug
- Had an acceptable biopsy at baseline (at least 1 endometrial biopsy with satisfactory tissue and no read of hyperplasia, disordered proliferative pattern or malignancy)

- Had a satisfactory endometrial biopsy result after or on day 326
- Had a postbaseline final diagnosis of hyperplasia, disordered proliferative pattern or malignancy prior to day 326.

Endometrial biopsies were reviewed by three independent pathologists. The initial evaluation of the biopsy was conducted by the primary pathologist. Then digital images of the biopsy uploaded to a database by the primary pathologist were then assessed by secondary and tertiary pathologists independently. Each pathologist was blinded to the treatment group information as well as other reviewer's results. The final diagnosis across the three pathologists' readings was established using a concordance rubric aligned with FDA guidance[1]. The concordance of two of three pathologists' readings determined the final diagnosis classification. If all three individual pathologist readings were discordant, the final diagnosis was classified based on the worst diagnosis.

2.4 STATISTICAL ANAYLSIS

This section describes the reviewer's approach to analysis for POP2 and POP4.

2.4.1 POP2

The analyses of POP2 are descriptive. The reviewer's analyses include the number and exposure-adjusted incidence rates. Exposure-adjusted incidence rate, defined as number of subjects with an event per 100 subject-years, accounts for the differences in follow-up time, and drug-exposure time among studies or treatment arms. The exposure for each subject is the time from first dose to the first onset of an event for those who had event or to the date of last collection of AE (last dose + 21 days) for those who did not. A 95% confidence interval (CI) for the exposure-adjusted incidence rates were calculated using normal approximation.

The reviewer combined the three studies by the exposure. In other words, the first 12 weeks of placebo subjects in studies 2693-CL-301 and 2693-CL-302 contributed to the placebo person-years, but their exposure time after week 12 contributed to the person-year of the treatment to which they were re-randomized. The reviewer's analysis accounted for study variability using meta-analysis, specifically the inverse-variance method.

For endometrial outcomes, per the FDA guidance [1], an event rate of $\leq$1% with an upper limit of the one-sided 95% CI $\leq$ 4% was considered a success. Clopper-Pearson exact method for the binomial proportion was used for the confidence interval for endometrial outcomes.

Reviewer Comments: *We found multiple flaws in the Applicant's analyses. The Applicant did not take study variability into consideration in the analyses of POP2 and simply pooled data from each study. When pooling data from multiple studies, the analysis should account for study variability. In addition, the Applicant's pre-specified definition of exposure did not account for*

[1] Guidance for Industry: Estrogen and Estrogen/Progestin Drug Products to Treat Vasomotor Symptoms and Vulvar and Vaginal Atrophy Symptoms – Recommendations for Clinical Evaluation

the timing and occurrence of an event. Therefore, only the results from the reviewer's analyses are presented. Note that the reviewer verified the Applicant's analyses results.

2.4.2 POP4

POP4 only contained the 52-week placebo-controlled trial (2693-CL-0304). The descriptive analysis methods are the same as that of POP2. Comparisons are based on exposure-adjusted incidence rates difference (EAIRD) with the 95% confidence interval (normal approximation).

3 RESULTS

Table 2 presents the number of patients and duration of drug exposure by treatment arm. POP2 included 952, 1103 and 1100 subjects in the placebo, fezolinetant 30 mg and fezolinetant 45 mg arms, respectively. The duration of drug exposure was similar between the fezolinetant 30 mg and fezolinetant 45 mg arms but the placebo arm had a shorter duration, because subjects in studies 2693-CL-301 and 2693-CL-302 were only exposed to placebo for 12 weeks. POP4 included 610, 611 and 609 subjects in the placebo, fezolinetant 30 mg and fezolinetant 45 mg arms, respectively. The median duration of drug exposure was the same across the three arms (364 days). Table 3 presents the subject disposition in POP4. Slightly more patients in the placebo arm discontinued the treatment.

Table 2: Summary of Drug Exposure.

Population			Placebo	Fezolinetant 30 mg	Fezolinetant 45 mg
POP2	Number of Subjects		952	1103 [1]	1100 [2]
	Duration of drug exposure (days)	Mean (SD)	210.7(144.1)	293.5(116.1)	302.9(108.1)
		Median	138.0	364.0	364.0
		Total (year)	549.1	886.3	912.1
POP4	Number of Subjects		610	611	609
	Duration of drug exposure (days)	Mean (SD)	284.4(130.8)	301.2(119.8)	302.4(117.9)
		Median	364.0	364.0	364.0
		Total (year)	475.0	503.8	504.2

[1] includes 951 subjects originally assigned to fezolinetant 30 mg and 152 subjects re-randomized to fezolinetant 30 mg from placebo.
[2] includes 949 subjects originally assigned to fezolinetant 45 mg and 151 subjects re-randomized to fezolinetant 45 mg from placebo.
Source: Reviewer's table using ADSL.xpt.

Table 3: Subject Disposition (POP4)

Category	Placebo (n = 610)	Fezolinetant 30 mg (n = 611)	Fezolinetant 45 mg (n = 609)	Total (n = 1830)
Completed	410 (67.2%)	451 (73.8%)	444 (72.9%)	1305 (71.3%)
Treatment discontinuation †	200 (32.8%)	160 (26.2%)	165 (27.1%)	525 (28.7%)
Primary reason for study intervention discontinuation				
Adverse event	27 (4.4%)	34 (5.6%)	28 (4.6%)	89 (4.9%)
Death	0	0	0	0
Lost to follow-up	39 (6.4%)	30 (4.9%)	33 (5.4%)	102 (5.6%)
Protocol deviation	1 (0.2%)	6 (1.0%)	5 (0.8%)	12 (0.7%)
Withdrawal by subject	119 (19.5%)	79 (12.9%)	85 (14.0%)	283 (15.5%)
Other	14 (2.3%)	11 (1.8%)	14 (2.3%)	39 (2.1%)

Source: The Applicant's clinical study report of 2693-CL-304, page 28.

The results of each safety outcome are presented in the rest of this section.

Malignancy: Table 4 presents the results from the analyses of malignancy outcome. The exposure-adjusted incidence rates were 0.2, 0.65 and 1.15 in placebo, fezolinetant 30 mg and fezolinetant 45 mg, respectively in POP2. The exposure-adjusted incidence rate of malignancy outcome showed a dose-dependent increased risk in POP4. Note that the lower bound of the 95% confidence interval of the exposure-adjusted incidence rate difference (EAIRD) comparing fezolinetant 45 mg and placebo arms was greater than the null value of 0. Figure 1 presents the cumulative incidence rates over time in POP4, which also shows a clear separation among arms.

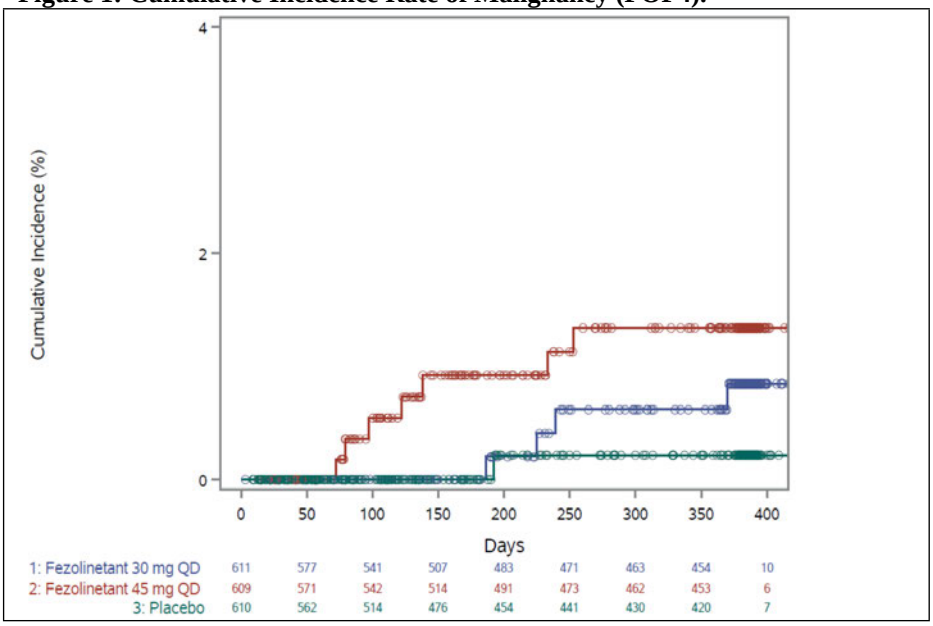
Table 4: Number of Subjects with Events (%) and Exposure-Adjusted Incidence Rates and Differences of Malignancy.

POP 2 (CL-301, CL-302 and CL-304)			
	Placebo N=952 PY=599.4	Fezolinetant 30 mg N=1103 PY=941.0	Fezolinetant 45 mg N=1100 PY=966.9
Subjects with Events	1	6	11
Exposure-adjusted Incidence Rates per 100 PY, (95% CI)	0.20 (0.03, 1.39)	0.65 (0.29, 1.44)	1.15 (0.64, 2.07)

POP 4 (CL-304)			
	Placebo N=610 PY=509.5	Fezolinetant 30 mg N=611 PY=537.9	Fezolinetant 45 mg N=609 PY=538.2
Subjects with Events	1	4	7
Proportion, % (95% CI)	0.16 (0, 0.91)	0.65 (0.18, 1.67)	1.15 (0.46, 2.35)
Crude Risk Difference, % (95% CI)	———	0.49 (-0.22, 1.21)	0.99 (0.08, 1.89)
Exposure-Adjusted Incidence Rates (EAIR) per 100 PY, (95% CI)	0.20 (0.0, 0.58)	0.74 (0.02, 1.47)	1.30 (0.33, 2.27)
EAIR_Difference per 100 PY (95% CI)	———	0.55 (-0.28, 1.37)	1.10 (0.07, 2.14)

Source: Reviewer's table using ADAE.xpt and ADSL.xpt.
N, number of subjects;PY, person-year

Figure 1: Cumulative Incidence Rate of Malignancy (POP4).



Source: Reviewer's figures using ADAE.xpt and ADSL.xpt.

Bone fractures: There were 11 subjects, 15 subjects and 15 subjects with bone fractures in placebo, fezolinetant 30 mg and fezolinetant 45 mg, respectively in POP2. No increased risk was shown in the fezolinetant arms compared to placebo in POP4 (Table 5).

Table 5: Number of Subjects with Events (%) and Exposure-Adjusted Incidence Rates and Differences of Bone Fracture.

POP 2 (CL-301, CL-302 and CL-304)			
	Placebo N=952 PY=583.5	Fezolinetant 30 mg N=1103 PY=921.9	Fezolinetant 45 mg N=1100 PY=951.1
Subjects with Events	11	15	15
Exposure-adjusted Incidence Rates per 100 PY, (95% CI)	1.99 (1.10, 3.60)	1.64 (0.99, 2.72)	1.60 (0.97, 2.66)

POP 4 (CL-304)			
	Placebo N=610 PY=505.7	Fezolinetant 30 mg N=611 PY=533.7	Fezolinetant 45 mg N=609 PY=533.1
Subjects with Events	10	9	10
Proportion, % (95% CI)	1.64 (0.79, 2.99)	1.47 (0.68, 2.78)	1.64 (0.79, 3.00)
Crude Risk Difference, % (95% CI)	———	-0.17 (-1.55, 1.22)	0.0 (-1.42, 1.43)
Exposure-Adjusted Incidence Rates (EAIR) per 100 PY, (95% CI)	1.98 (0.75, 3.20)	1.69 (0.58, 2.79)	1.88 (0.71, 3.04)
EAIR_Difference per 100 PY (95% CI)	———	-0.29 (-1.94, 1.36)	-0.10 (-1.79, 1.59)

Source: Reviewer's table using ADAE.xpt and ADSL.xpt.
N, number of subjects; PY, person-year

Liver test elevation: In POP2, the exposure-adjusted incidence rates were 7.31, 8.26 and 6.62 in placebo, fezolinetant 30 mg and fezolinetant 45 mg, respectively. Although the incidence rate of elevated liver test was higher in the fezolinetant 30 mg arm compared to placebo in POP 4, the placebo and both fezolinetant arms' 95% CIs overlapped, and the 95 % CI for EAIRD included 0.

Table 6: Number of Subjects with Events (%) and Exposure-Adjusted Incidence Rates and Differences of Liver Test Elevation.

POP 2 (CL-301, CL-302 and CL-304)			
	Placebo N=952 PY=588.5	Fezolinetant 30 mg N=1103 PY=914.7	Fezolinetant 45 mg N=1100 PY=943.8
Subjects with Events	39	66	61
Exposure-adjusted Incidence Rates, per 100 PY, (95% CI)	7.31 (5.34, 10.00)	8.26 (6.49, 10.51)	6.62 (5.15, 8.51)

POP 4 (CL-304)			
	Placebo N=610 PY=496.3	Fezolinetant 30 mg N=611 PY=520.9	Fezolinetant 45 mg N=609 PY=524.9
Subjects with Events	30	35	32
Proportion, % (95% CI)	4.92 (3.34, 6.95)	5.73 (4.02, 7.88)	5.25 (3.62, 7.34)
Crude Risk Difference, % (95% CI)	———	0.81 (-1.71, 3.33)	0.34 (-2.13, 2.80)
Exposure-Adjusted Incidence Rates (EAIR) per 100 PY, (95% CI)	6.04 (3.88, 8.21)	6.72 (4.49, 8.95)	6.10 (3.98, 8.21)
EAIR_Difference per 100 PY (95% CI)	———	0.67 (-2.43, 3.78)	0.05 (-2.97, 3.07)

Source: Reviewer's table using ADAE.xpt and ADSL.xpt.
N, number of subjects;PY, person-year

Creatine Kinase elevation: In POP2, the exposure-adjusted incidence rates were 0.59, 1.66 and 2.86 in placebo, fezolinetant 30 mg and fezolinetant 45 mg, respectively. In POP4, the exposure-adjusted incidence rates for CK elevation were higher in the fezolinetant arms compared to placebo without a dose effect but the 95% CI for EAIRD included 0(Table 7).

Table 7: Number of Subjects with Events (%) and Exposure-Adjusted Incidence Rates and Differences of CK Elevation.

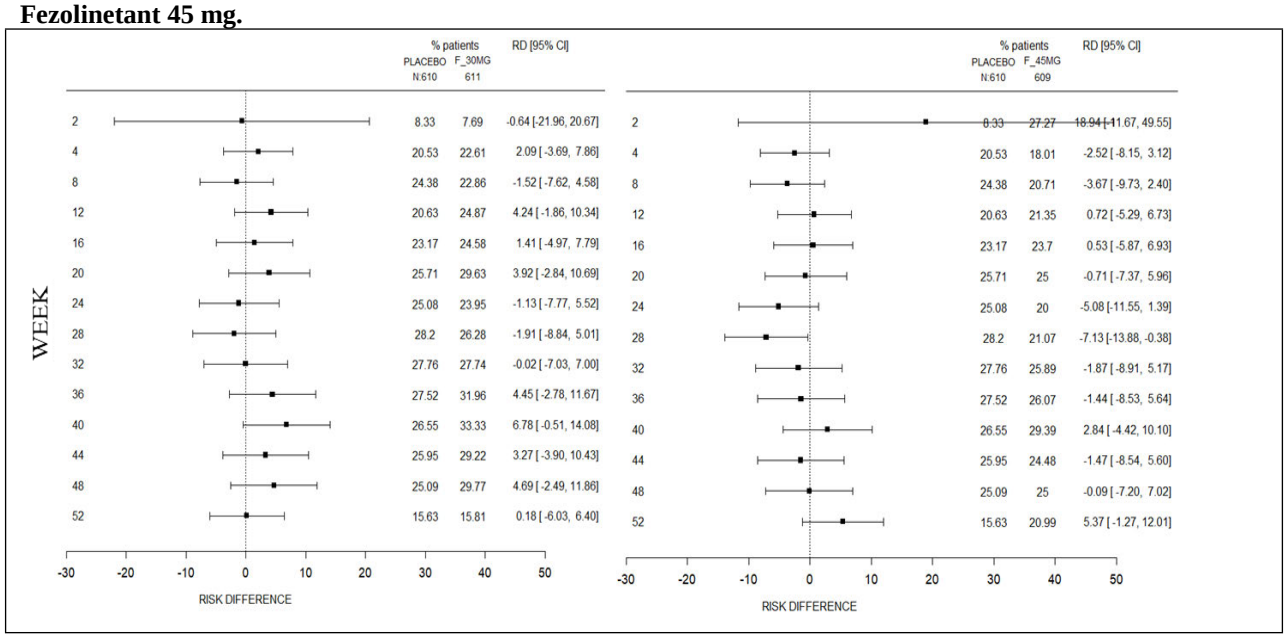
POP 2 (CL-301, CL-302 and CL-304)			
	Placebo N=952 PY=603.0	Fezolinetant 30 mg N=1103 PY=941.7	Fezolinetant 45 mg N=1100 PY=962.8
Subjects with Events	3	15	23
Exposure-adjusted Incidence Rates, per 100 PY, (95% CI)	0.59 (0.19, 1.83)	1.66 (1.00, 2.75)	2.86 (1.90, 4.31)
POP 4 (CL-304)			
	Placebo N=610 PY=509.2	Fezolinetant 30 mg N=611 PY=535.5	Fezolinetant 45 mg N=609 PY=536.6
Subjects with Events	3	7	6
Proportion, % (95% CI)	0.49 (0.10, 1.43)	1.15 (0.46, 2.35)	0.99 (0.36, 2.13)
Crude Risk Difference, % (95% CI)	———	0.65 (-0.36, 1.66)	0.49 (-0.47, 1.45)
Exposure-Adjusted Incidence Rates (EAIR) per 100 PY, (95% CI)	0.59 (0.0, 1.26)	1.31 (0.34, 2.28)	1.12 (0.22, 2.01)
EAIR Difference per 100 PY (95% CI)	———	0.72 (-0.46. 1.89)	0.53 (-0.59, 1.64)

Source: Reviewer's table using ADAE.xpt and ADSL.xpt.
N, number of subjects; PY, person-year.

Glucose level shift: Glucose level shift from normal or low at baseline to high was examined by week in POP4. The normal glucose level was defined as 3.8857 mmol/L to 5.551 mmol/L. As

shown in Figure 2, there were no notable patterns of increased risks in the fezolinetant arm compared to placebo.

Figure 2: Risk Difference in Glucose Level Shift to High from Low or Normal at Baseline by Week (POP 4). Left: Comparison between Placebo and Fezolinetant 30mg; Right: Comparison between Placebo and Fezolinetant 45 mg.



Source: Reviewer's figure using ADLB.xpt and ADSL.xpt.
N: number of subjects; RD: Risk Difference; F_30MG, fezolinetant 30 mg; F_45MG, fezolinetant 45 mg.

Endometrial Outcomes: EH set consists of 210 subjects (fezolinetant 30 mg) and 204 subjects (fezolinetant 45 mg) in POP4 and 302 subjects (fezolinetant 30 mg) and 305 subjects (fezolinetant 45 mg) in POP2. Note that DUOG did not agree with the Applicant's justification to exclude one fezolinetant 45 mg subject in study 2693-CL-304 who had endometrial adenocarcinoma. Therefore, the reviewer's EH set analyses included an additional subject compared to the Applicant's report. No subjects in the placebo arm had hyperplasia, malignancy or disordered proliferative pattern (data not shown). The rates of hyperplasia and malignancy were less than 1% with upper limits of a one-sided 95% CI $\leq$ 4%, which met the pre-specified criteria in the FDA guidance in both POP2 and POP4 (Table 8).

Table 8: Number of Subjects with Final Diagnosis of Hyperplasia, Malignancy and Disordered Proliferative Pattern with Upper Limit of One-sided 95% CI in EH set.

POP 2 (CL-301, CL-302 and CL-304)			
		Fezolinetant 30 mg N=302	Fezolinetant 45 mg N=305
Hyperplasia	n (%) upper limit	0 1.0 %	1 (0.3) 1.6 %
Malignancy	n (%) upper limit	1 (0.3) 1.6 %	1 (0.3) 1.6 %
Disordered proliferative pattern	n (%) upper limit	4 (1.3) 3.0 %	3 (1.0) 2.5 %
POP 4 (CL-304)			
		Fezolinetant 30 mg N=210	Fezolinetant 45 mg N=204
Hyperplasia	n (%) upper limit	0 1.4 %	1 (0.5) 2.3 %
Malignancy	n (%) upper limit	0 2.2 %	1 (0.5) 2.3 %
Disordered proliferative pattern	n (%) upper limit	3 (1.4) 3.7 %	0 1.5 %

Source: Reviewer's table using ADMI.xpt and ADSL.xpt.
Upper limit: upper limit of one-sided 95% CI using Clopper-Pearson exact method.

4 RESPONSES TO CONSULT QUESTIONS

1. Given that in Trial 2693-CL-0301 and Trial 2693-CL-0302, women were either on fezolinetant for 40 weeks (inclusive of women who were on placebo for the initial 12-week efficacy portion of the trials, but switched to 30 or 45 mg of fezolinetant for the trial extension) or 52 weeks (women who participated in a fezolinetant arm for the first 12 weeks of the trial and participated in the additional 40 -week extension), please conduct the appropriate analysis of safety for this scenario. Indicate whether the women receiving fezolinetant for 40 weeks should be considered separately from those receiving fezolinetant for 52 weeks or whether data from the two timeframes of consideration can be pooled.

DB VII Response: The subjects originally assigned to fezolinetant 30 mg (45 mg) (52-week exposure) can be combined with the subjects re-randomized to fezolinetant 30 mg (45 mg) (40-week exposure) to descriptively estimate within-arm incidence rates that incorporate exposure time, such has EAIR. It should be noted that, even descriptively, these within-arm estimates may be biased, because of the cross-over study element of studies 2693-CL-301 and 2693-CL-302 and subjects re-randomized to fezolinetant arm(s) at 12 weeks knew that they were on treatment, which might make studies 2693-CL-301and 2693-CL-302 inherently too different from study 2693-CL-304. Furthermore, comparing placebo and fezolinetant arms at 52 weeks would violate the intention-to-treat (ITT) principle. Therefore, we do not recommend comparative analyses of fezolinetant 30 mg (45 mg) arm to placebo arm for the long-term safety assessment at 52 weeks in POP2 and interpretation of within-arm estimates of the incidence rates in POP2 should be interpreted with caution. Study 2693-CL-304 (POP4) should be the primary source for comparative assessment at weeks 52.

Note that we found multiple flaws in the Applicant's analyses. The Applicant did not take study variability into consideration in the analyses of POP2 and simply pooled data from each study. When pooling data from multiple studies, the analysis should account for study variability. In addition, the Applicant's pre-specified definition of exposure did not account for the timing and occurrence of an event. However, because of the descriptive nature of the analyses of POP2 and the small number of events, these flaws did not make a big impact on the results.

2. Trial 2693-CL-0304 was a 52-week placebo-controlled trial of the 30 and 45 mg dosage strength of fezolinetant vs. placebo. Given that there was no placebo comparator for 40 weeks of data for Trial 2693-CL-0301 and Trial 2693-CL-0302, would it be appropriate to compare the data from 52 weeks exposure to placebo in Trial 2693-CL-0304 to pooled data from 40 weeks exposure to fezolinetant or 52 weeks of exposure to fezolinetant in Trials 2693-CL-0301 and 2693-CL-0302? Discuss the limitation of this proposal for comparisons of pooled data as well as that for comparison of exposure to placebo in Trial 2693-CL-0304 to pooled data from 52 weeks of exposure or 40 weeks of exposure to fezolinetant in Trials 2693-CL-0301, 2693-CL-0302 and 2693-CL-0304.

DB VII Response:

See the response to Question 1.

3. We request your assistance in evaluating whether there is a potential safety signal of malignancy in women taking fezolinetant. In a pooled consideration of Trials 2693-CL-0301 2693-CL-0302, and 2693-CL-0304, there was 1 case of malignancy in placebo(a squamous cell carcinoma of the skin), 4 cases of malignancy in fezolinetant 30mg (1 squamous cell carcinoma of the skin, 1 squamous cell carcinoma of the oral cavity, and 2 basal cell carcinomas) and 10 cases of malignancy in fezolinetant 45 mg (1 squamous cell carcinoma of the skin, 1 squamous cell carcinoma of the oral cavity, 1 bone cancer, 2 colon cancer, 3 endometrial adenocarcinomas, 1 malignant melanoma in situ, 1 non-small cell carcinoma of the lung). Additionally, 1 case of unspecified lung neoplasm and 1 case of keratoacanthoma, a neoplasm with high malignant potential. Given the higher incidence of some of these malignancy in the older population of women in general (as compared to their younger counterparts), do you consider that these 15 to 16 cases of malignancies of various primary sites represent a signal for malignancy with this product? If so, how would you characterize the strength, i.e., weak, moderate, or strong signal) and the impact on approvability of the product. If approved discuss relevant labeling, including whether focused monitoring should be recommended?

DB VII Response: Per the response to Question 1, the comparative assessment of malignancy is most appropriate using POP4 (i.e. Study 2693-CL-0304). In POP4, the risk of malignancy was increased in the fezolinetant arms compared to placebo in a dose-dependent manner. However, from a statistical perspective, several factors add uncertainty to this observed finding:

1) The duration of Study 2693-CL-0304 was 52 weeks which may be too short to assess malignancy, which typically has a long latency;

2) Study 2693-CL-0304 was not designed to ascertain safety outcomes that occur more than 21 days after treatment discontinuation which may result in malignancy outcomes not being captured;

3) The small number of events result in uncertainty around the estimated risk difference.

We defer to the oncology clinical team to assess the potential safety signal for malignancy from a clinical perspective.

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

JOO YEON LEE
10/20/2022 06:20:19 AM

CLARA KIM
10/20/2022 08:52:39 AM

MATTHEW J SOUKUP
10/20/2022 09:00:46 AM