

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

210132Orig1s000

STATISTICAL REVIEW(S)

Addendum to the Statistical Review

NDA/BLA #: 210132
Supplement #:
Drug Name: Estradiol (E2) and Progesterone (P) Capsules:
Combination of estradiol and progesterone
Indication(s): Treatment of moderate to severe vasomotor symptoms
associated with menopause in women with a uterus
Applicant: TherapeuticsMD, INC.
Date(s): Submitted: 12/28/2017
PDUFA: 10/28/2018
Review Priority: Standard

Biometrics Division: Division of Biometrics III
Statistical Reviewer: Jia Guo, Ph.D.
Concurring Reviewers: Mahboob Sobhan, Ph.D.

Medical Division: Division of Bone, Reproductive and Urologic Products, HFD-580
Clinical Team: Theresa Van Der Vlugt, MD. Clinical reviewer
Shelley Slaughter, MD. Clinical team leader
Project Manager: Kimberly Shiley

Keywords: NDA review, subgroup, ANCOVA, MMRM

Reference is made to Statistical Review of NDA210132 submitted to DARRTS on 10/16/2018. This addendum pertains to Division of Biometrics 3/Office of Biostatistics' perspective on the subgroup analysis findings in the Black/African-American patients. Per Division of Biometrics 3/Office of Biostatistics, the subgroup analysis in one phase 3 trial does not establish that the drug is ineffective in Black/African-American patients, but the subgroup findings suggesting lack of efficacy in Black/African-American patients could be described in labeling.

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/

JIA GUO
10/26/2018

MAHBOOB SOBHAN
10/26/2018



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 #: 210132
Supplement #:
Drug Name: Estradiol (E2) and Progesterone (P) Capsules:
Combination of estradiol and progesterone
Indication(s): Treatment of moderate to severe vasomotor symptoms
associated with menopause in women with a uterus
Applicant: TherapeuticsMD, INC.
Date(s): Submitted: 12/28/2017
PDUFA: 10/28/2013
Review Priority: Standard
Biometrics Division: Division of Biometrics III
Statistical Reviewer: Jia Guo, Ph.D.
Concurring Reviewers: Mahboob Sobhan, Ph.D.
Medical Division: Division of Bone, Reproductive and Urologic Products, HFD-
580
Clinical Team: Theresa Van Der Vlugt, MD. Clinical reviewer
Shelley Slaughter, MD. Clinical team leader
Project Manager: Kimberly Shiley
Keywords: NDA review, subgroup, ANCOVA, MMRM

Table of Contents

1	EXECUTIVE SUMMARY	4
2	INTRODUCTION	6
2.1	OVERVIEW	6
2.2	DATA SOURCES	6
3	STATISTICAL EVALUATION	6
3.1	DATA AND ANALYSIS QUALITY	6
3.2	EVALUATION OF EFFICACY	6
3.2.1	<i>Study Design and Endpoints</i>	7
3.2.2	<i>Statistical Methodologies</i>	9
3.2.3	<i>Patient Disposition, Demographic and Baseline Characteristics</i>	10
3.2.4	<i>Results and Conclusions</i>	11
3.3	EVALUATION OF SAFETY	14
4	FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	14
4.1	GENDER, RACE, AGE, AND GEOGRAPHIC REGION	14
4.2	OTHER SPECIAL/SUBGROUP POPULATIONS	24
5	SUMMARY AND CONCLUSIONS	24
5.1	CONCLUSIONS AND RECOMMENDATIONS	24
	APPENDICES	25
	DEMOGRAPHICS	25
	APPLICANT'S ADDITIONAL ANALYSES	29

LIST OF TABLES

Table 1: List of all studies included in analysis.....	6
Table 2: Subject Disposition for the Entire 52-Week Study (Safety Population).....	10
Table 3: Change from Baseline in the Mean Number of Weekly Moderate and Severe VMS at Week 4 and Week 12 (MITT-VMS Population).....	11
Table 4: Change from Baseline in the Mean Weekly Severity Scores of VMS at Week 4 and Week 12 (MITT-VMS Population).....	13
Table 5: Incidence of Endometrial Hyperplasia at 12 Months (ES Population).....	14
Table 6: Change from Baseline in the Number of Weekly Moderate to Severe VMS at Week 4 and Week 12 by Race (MITT-VMS Population).....	15
Table 7: Change from Baseline in the Severity of Weekly Moderate to Severe VMS at Week 4 and Week 12 by Race (MITT-VMS Population).....	16
Table 8: Change from baseline in Weekly VMS frequency at Weeks 4 and 12 - ANCOVA analysis.....	21
Table 9: Change from baseline in Weekly VMS frequency at Weeks 4 and 12 - ANCOVA analysis adjusting for more baseline characteristics.....	22
Table 10: Change from baseline in Weekly VMS frequency at Weeks 4 and 12 - MMRM analysis.....	22
Table 11: Change from baseline in Weekly VMS frequency at Weeks 4 and 12 - MMRM analysis adjusting for more baseline characteristics.....	22
Table 12: Change from baseline in Weekly VMS severity at Weeks 4 and 12 - ANCOVA analysis.....	23
Table 13: Change from baseline in Weekly VMS severity at Weeks 4 and 12 - ANCOVA analysis adjusting for more baseline characteristics.....	23
Table 14: Change from baseline in Weekly VMS severity at Weeks 4 and 12 - MMRM analysis.....	23
Table 15: Change from baseline in Weekly VMS severity at Weeks 4 and 12 - MMRM analysis adjusting for more baseline characteristics.....	24
Table 16: Selected Demographic and Baseline Characteristics for Safety Population.....	25
Table 17: Gynecological History for Safety Population.....	26
Table 18: Selected Demographic and Baseline Characteristics for MITT-VMS.....	27
Table 19: Gynecological History for MITT-VMS Population.....	28
Table 20: Applicant’s analysis on Change from Baseline in Weekly Frequency of Moderate to Severe VMS at Week 12 – unadjusted ANCOVA Model (MITT-VMS Population) by race (white and Black/ African American only)....	29
Table 21: Applicant’s analysis on Change from Baseline in Weekly Frequency of Moderate to Severe VMS at Week 12 – adjusted ANCOVA Model (MITT-VMS Population) by race (white and Black/ African American only).....	30

LIST OF FIGURES

Figure 1: Mean Change from Baseline in the Weekly Frequency of Moderate to Severe VMS (MITT-VMS Population).....	12
Figure 2: Change from Baseline in the Weekly Severity of Moderate to Severe VMS (MITT-VMS Population)....	13
Figure 3: CDF for Change from Baseline in Weekly VMS frequency at Week 4 by Race and Treatment (MITT-VMS population).....	19
Figure 4: CDF for Change from Baseline in Weekly VMS frequency at Week 12 by Race and Treatment (MITT-VMS population).....	19
Figure 5: CDF for Change from Baseline in Weekly VMS Severity at Week 4 by Race and Treatment (MITT-VMS population).....	20
Figure 6: CDF for Change from Baseline in Weekly VMS Severity at Week 12 by Race and Treatment (MITT-VMS population).....	20

1 EXECUTIVE SUMMARY

In this submission, the Applicant is seeking approval of (b) (4) 1 mg estradiol (E2) in combination with 100 mg progesterone (P) in postmenopausal women with an intact uterus to reduce the frequency and severity of vasomotor symptoms (VMS) and to manage the incidence of endometrial hyperplasia. To support this claim, the safety and efficacy data from one phase 3, double-blind, randomized, placebo-controlled clinical trial (TXC12-05) was submitted. This statistical review evaluates the adequacy of the submitted information to support the safety and efficacy of (b) (4) 1 mg E2 combined with 100 mg P.

A total of 1845 subjects were enrolled in the study, of which 756 subjects who experienced a minimum daily frequency of ≥ 7 (≥ 50 per week) moderate to severe hot flushes were randomized in a 1:1:1:1:1 allocation ratio to one of the following 5 groups (Referred to as VMS substudy).

- Treatment 1: Combined 1 mg E2/100 mg P
- Treatment 2: Combined 0.5 mg E2/100 mg P
- Treatment 3: Combined 0.5 mg E2/50 mg P
- Treatment 4: Combined 0.25 mg E2/50 mg P
- Treatment 5: Placebo.

Other enrolled patients who did not participate the VMS substudy were randomized to one of the active treatment groups in a 1:1:1:1 allocation ratio (Treatment Groups 1 through 4). All subjects, including VMS substudy participants received blinded treatment for 12 months for safety evaluation.

The primary efficacy was evaluated based on the following four pre-specified co-primary endpoints, i.e. change in the *frequency* and *severity* of moderate to severe VMS per day from baseline to *Week 4* and *Week 12*.

The primary safety endpoint was the incidence of endometrial hyperplasia at 12 months and was evaluated on the endometrium hyperplasia population (ES population).

Per FDA analysis, study TXC12-05 showed that

1. The 1 mg E2/100 mg P and 0.5 mg E2/100 mg P combinations demonstrated statistically significant reductions from baseline in the weekly frequency and severity of moderate to severe VMS at Week 4 and Week 12 compared to placebo. The 1 mg E2/100 mg P dose achieved the clinical meaningful threshold of reducing at least 14 hot flushes per week starting from Week 5 and maintained through Week 12. However, the 0.5mg E2/100mg P dose did not achieve this threshold until Week 9.
2. Despite the significant efficacy results in the overall study population and similar results in the subgroup of White subjects, there was no efficacy seen in the other major subgroup of Black/African American subjects on the VMS frequency reduction at Weeks 4 and 12 and reduction in VMS severity was only seen at Week 12, not at Week 4. However, the study was not powered to demonstrate efficacy by subgroup of race.

3. The incidence rate of endometrial hyperplasia in each active treatment group was less than a clinically acceptable rate of 4% (limit of the upper bound of the 95% confidence interval).

From a statistical perspective, the data supports the efficacy of 1 mg E2/100 mg P dose for the treatment of moderate to severe vasomotor symptoms associated with menopause. However, the efficacy was not observed in the subgroup of Black/African American subjects. The review team recommends that this finding needs to be included in the label, should the Division decide to approve this product.

2 INTRODUCTION

2.1 Overview

The Applicant, Therapeutics MD, INC. seeks approval of TX-001HR, i.e. estradiol ((b) (4) 1 mg) in combination with progesterone (100 mg) in postmenopausal women with an intact uterus to reduce the frequency and severity of vasomotor symptoms and to manage the incidence of endometrial hyperplasia.

Per the Applicant, “TX-001HR is a soft gel formulation containing solubilized estradiol which is chemically and biologically identical to endogenous 17 β -estradiol with micronized progesterone” The Applicant has submitted one phase 3 clinical study (TXC12-05) to support this submission. Table 1 presents a brief summary of the study addressed in this review.

Table 1: List of all studies included in analysis

Study	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
TXC12-05	Phase 3, double-blind, Randomized, multicenter, placebo-controlled	12 months	15 days	Randomized: 1 mg E2/100 mg P: 418 0.5 mg E2/100 mg P: 426 0.5 mg E2/50 mg P: 422 0.25 mg E2/50 mg P: 427 Placebo: 152	Female 40-65 years old with an intact uterus For VMS substudy: ≥ 7 per day (average) (or ≥ 50 per week) moderate to severe hot flushes

Source: Reviewer’s summary based on study protocol.

2.2 Data Sources

The study data, reports and additional information for these studies were submitted electronically. The submitted SAS data sets for all studies were complete and well documented. These items are located in the Electronic Document Room at \\Cdseub1\EVSPROD\NDA210132 under the submissions dated 12/28/2017 and 7/12/2018.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The Applicant submitted both the tabulation data and analysis data for the study TXC12-05. Data sets 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

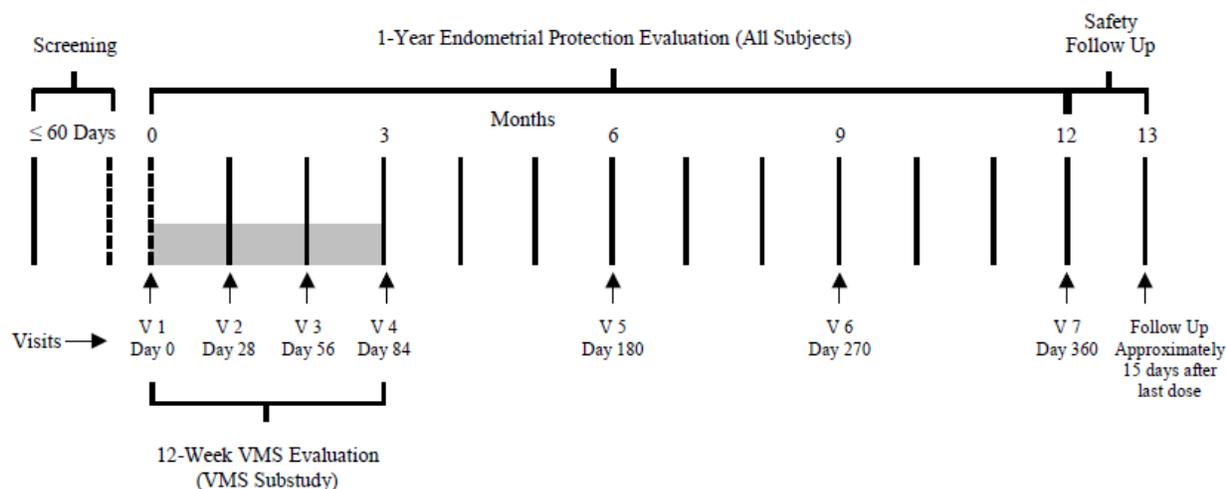
The efficacy evaluation of TX-001HR is based on the study TXC12-05.

3.2.1 Study Design and Endpoints

3.2.1.1 Study Design

Study TXC12-05 is a phase 3, randomized, double-blind, placebo-controlled, parallel group, multicenter study of determine the safety and efficacy of estradiol/progesterone combinations in postmenopausal subjects between the ages of 40 and 65 years with an intact uterus having vasomotor symptoms associated with menopause.

The study was comprised of a screening period (approximately 60 days before randomization), a 12-month double-blind treatment period and a follow-up period (approximately 15 days after last dose). Subjects who experienced a minimum daily frequency of ≥ 7 (or ≥ 50 per week) moderate to severe hot flushes and not taking any medications that may affect vasomotor symptoms at baseline participated in a VMS substudy during the first 12 weeks of treatment and then continued participation on the same treatment for an additional nine (9) months. Study schema was presented in Figure 1.



The study treatments are

- Treatment 1: Combined 1 mg E2/100 mg P
- Treatment 2: Combined 0.5 mg E2/100 mg P
- Treatment 3: Combined 0.5 mg E2/50 mg P
- Treatment 4: Combined 0.25 mg E2/50 mg P
- Treatment 5: Placebo.

Subjects in the VMS substudy were randomized within each study site to one of the treatment groups in a 1:1:1:1:1 allocation ratio (Treatment Groups 1 through 5). Subjects not participating in the VMS substudy were randomized to one of the active treatment groups in a 1:1:1:1 allocation ratio (Treatment Groups 1 through 4). All subjects, including VMS substudy participants received blinded treatment for 12 months for safety evaluation.

3.2.1.2 Endpoints

3.2.1.2.1 Primary Efficacy Endpoints

The co-primary efficacy endpoints were defined as:

- Mean change in frequency of moderate to severe vasomotor symptoms from Baseline to Week 4 in an active treatment group compared with placebo;
- Mean change in frequency of moderate to severe vasomotor symptoms from Baseline to Week 12 in an active treatment group compared with placebo;
- Mean change in severity of moderate to severe vasomotor symptoms at Baseline to Week 4 in an active treatment group compared with placebo;
- Mean change in severity of moderate to severe vasomotor symptoms at Baseline to Week 12 in an active treatment group compared with placebo.

The VMS frequency and severity are recorded using daily diaries.

For each subject, the baseline frequency and severity of VMS was calculated using the most recent seven consecutive days of data prior to randomization. The eligibility for the VMS Substudy was determined by this baseline frequency of hot flushes.

The weekly frequency of moderate to severe hot flushes during treatment period was calculated from the daily diary records using a forward counting process of 7-day intervals beginning with the baseline date. The weekly number of moderate to severe hot flushes for each assessment week (Baseline, and Weeks 1 through 12) were derived as:

Weekly Frequency = total number of moderate and severe hot flushes for the subject week

Diary data extending beyond 12 weeks (84 days) were excluded from this calculation.

The weekly severity of vasomotor symptoms is derived as:

- Baseline Weekly Severity Score = (number of moderate hot flushes for 7 days) x 2 + (number of severe hot flushes for 7 days) x 3 / (total number of moderate to severe hot flushes over 7 days).
- On Treatment Weekly Severity Score = [(number of mild hot flushes for 7 days) x 1 + (number of moderate hot flushes for 7 days) x 2 + (number of severe hot flushes for 7 days) x 3] / (total number of mild, moderate and severe hot flushes over 7 days).

The calculation of weekly severity was agreed with the agency a priori.

A weekly severity score of zero (0) was assigned for subjects reporting no hot flushes for a given assessment week.

3.2.1.2.2 Primary Safety Endpoint

The primary safety endpoint is the subject incidence proportion of endometrial hyperplasia at 12 months (to demonstrate a hyperplasia proportion that is $\leq 1\%$ with an upper bound of the one-sided 95 percent confidence interval [CI] for that proportion that does not exceed 4%).

3.2.2 Statistical Methodologies

All primary efficacy analysis was conducted on the MITT-VMS population, which was defined MITT subjects in the VMS substudy (for MITT definition, refer to section 3.2.3).

3.2.2.1 Analysis of Co-Primary Efficacy Endpoints

For each co-primary endpoint, a Mixed Model Repeated Measures (MMRM) analysis was applied to the 12 weekly change scores. The model included baseline value as covariate, treatment, study week, and treatment-by-study week interaction as fixed factors, and subject as the repeated measure unit.

Descriptive statistics were reported for each endpoint. Least square (LS) estimates of mean differences from placebo for each dose and week with the associated 95% CIs were derived.

A gatekeeping testing procedure was used to control the overall type I error for the four combination doses and the co-primary efficacy endpoints. The testing started by examining the highest dose (1 mg E2/100 mg P) for each of the co-primary endpoints at 0.05 level, 2-sided. If the four p-values for the co-primaries were significant ($p \leq 0.05$) then the hypothesis testing would continue to the next dose (0.5 mg E2/100 mg P) for each of the co-primaries, as described above. If at any point the hypothesis testing yielded a non-significant result, the testing would be stopped.

3.2.2.2 Analysis of Primary Safety Endpoint

The primary analysis population for endometrium hyperplasia is the endometrium hyperplasia population (ES population). An ES subject at Month 12 was one who was randomly assigned and took at least one dose of study medication, with no exclusionary protocol violation (as detailed herein) and had a pretreatment, evaluable endometrial biopsy, which was negative at baseline, and an evaluable biopsy at Month 12, or who had developed endometrial hyperplasia at any time during the study, post-Baseline.

The incidence rate of endometrial hyperplasia at Month 12 was calculated as follows:

Incidence rate = # of new subjects with biopsies positive for endometrial hyperplasia during the study, but post-Baseline / # of ES subjects at Month 12

A confidence interval approach was used to determine if the hyperplasia incidence was acceptable. For each active treatment group, the incidence of hyperplasia at Month 12 and the associated asymptotic upper 95% one-sided confidence limit were calculated.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

In study TXC12-05, a total of 1845 subjects were randomized. The Applicant pre-defined the following populations in both studies,

- MITT population: all consented and randomized subjects who had valid baseline hot flush diary data, received at least 1 dose of their randomized treatment, and had at least 1 day of on-treatment daily diary data.
- Safety population: all randomized subjects who received at least 1 dose of their randomized treatment and had at least 1 post-treatment safety assessment.

1835 subjects were included in the safety population. The disposition for the Safety population is summarized in Table 2. In the Safety population, 1275 (69.5%) completed the study and 560 (30.5%) subjects discontinued prematurely. Overall, the most common reasons for early discontinuations were: AE (8.9%), subject withdrew consent (8.2%), lost to follow-up (7.5%), protocol deviation (3.2%), lack of efficacy (1.9%), Investigator/Sponsor decision (0.4%), and other (0.3%).

Table 2: Subject Disposition for the Entire 52-Week Study (Safety Population)

	1 mg E2/ 100 mg P (N=415)	0.5mg E2/ 100 mg P (N=424)	0.5 mg E2/ 50 mg P (N=421)	0.25 mg E2/ 50 mg P (N=424)	Placebo (N=151)	Total (N=1835)
Number of subjects completed, n (%)	284 (68.4)	305 (71.9)	312 (74.1)	281 (66.3)	93 (61.6)	1275 (69.5)
Number of subjects discontinued, n (%)	131 (31.6)	119 (28.1)	109 (25.9)	143 (33.7)	58 (38.4)	560 (30.5)
Adverse Event	46 (11.1)	33 (7.8)	34 (8.1)	41 (9.7)	10 (6.6)	164 (8.9)
Investigator/Sponsor Decision	1 (0.2)	3 (0.7)	2 (0.5)	2 (0.5)	0	8 (0.4)
Lack of Efficacy	5 (1.2)	4 (0.9)	4 (1.0)	10 (2.4)	12 (7.9)	35 (1.9)
Lost to Follow-up	27 (6.5)	30 (7.1)	26 (6.2)	38 (9.0)	17 (11.3)	138 (7.5)
Protocol Deviation	15 (3.6)	6 (1.4)	12 (2.9)	20 (4.7)	6 (4.0)	59 (3.2)
Subject Withdrew Consent	36 (8.7)	42 (9.9)	29 (6.9)	31 (7.3)	13 (8.6)	151 (8.2)
Other	1 (0.2)	1 (0.2)	2 (0.5)	1 (0.2)	0	5 (0.3)

Source: Study report, Table 10.

The MITT-VMS population was the primary population for the efficacy assessment. Of the 766 subjects randomized to the VMS Substudy, 726 (94.8%) subjects met the criteria for MITT population to be included in the MITT-VMS population with 89.1% completing through Week 12.

The demographics and baseline characteristics of the treatment groups are summarized in the Appendix (Table 12 to Table 15) for Safety population and MITT-VMS population respectively. In both study populations, approximately 65% of subjects were white and 31% were Black/African American. The mean age of subjects was 54.6 year. At baseline, the mean BMI was approximately 27 kg/m².

3.2.4 Results and Conclusions

3.2.4.1 Results for Co-Primary Efficacy Endpoints

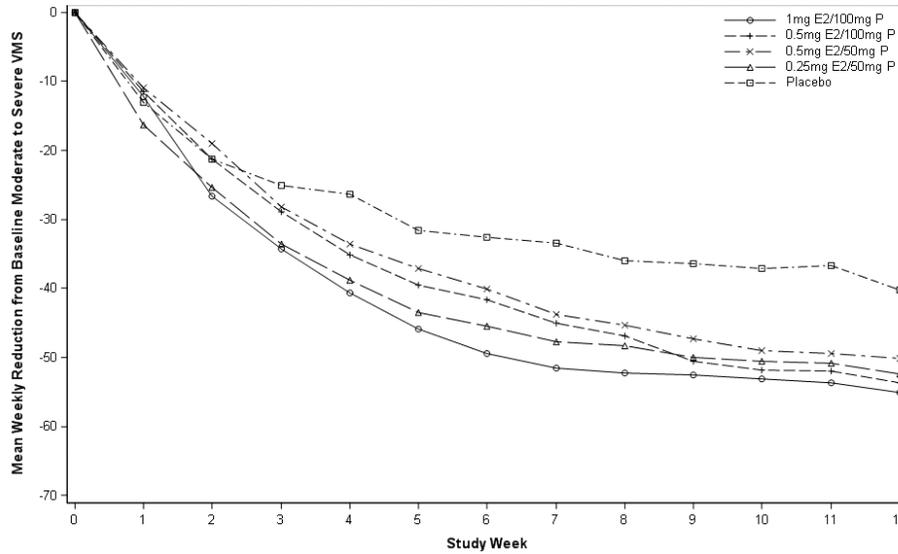
The four co-primary efficacy endpoints were analyzed using MMRM models as pre-specified in the protocol. Results are summarized in Table 3 and Table 4, and depicted in Figure 1 and Figure 2. For the co-primary efficacy endpoints, the baseline mean weekly number of moderate to severe VMS ranged from 72.1 to 77.0. The 1 mg E2/100 mg P and 0.5 mg E2/100 mg P groups both showed statistically significant decrease in the weekly moderate to severe VMS frequency from baseline compared to the placebo group at Weeks 4 and 12 (Week 4: -12.81 for 1 mg E2/100 mg P; -8.07 for 0.5 mg E2/100 mg P; Week 12: -16.58 for 1 mg E2/100 mg P; -15.07 for 0.5 mg E2/100 mg P). By Week 4, neither dose achieved the clinical threshold for hormonal products compared to placebo of a reduction of at least 14 in the weekly frequency of VMS. Compared to placebo, 1 mg E2/100 mg P achieved more than 14 hot flushes reduction per week from Week 5 and maintained to Week 12; and 0.5 mg E2/100 mg P achieved more than 14 hot flushes reduction per week from Week 9 and maintained to Week 12.

Table 3: Change from Baseline in the Mean Number of Weekly Moderate and Severe VMS at Week 4 and Week 12 (MITT-VMS Population)

	1 mg E2/ 100 mg P (N=141)	0.5 mg E2/ 100 mg P (N=149)	0.5 mg E2/ 50 mg P (N=147)	0.25 mg E2/ 50 mg P (N=154)	Placebo (N=135)
Baseline					
Mean (SD)	74.4 (35.26)	72.1 (27.76)	75.9 (28.04)	77.0 (30.42)	72.4 (23.26)
Week 4 (n)	134	144	142	152	126
Mean (SD) change from Baseline	-40.6 (30.59)	-35.1 (29.14)	-33.6 (30.64)	-38.9 (31.04)	-26.4 (27.05)
LS Mean (SE) difference from placebo	-12.81 (3.30)	-8.07 (3.25)	-4.81 (3.26)	-10.40 (3.22)	---
MMRM P-value vs placebo	< 0.001	0.013	0.141	0.001	---
Week 12 (n)	124	129	124	135	115
Mean (SD) change from Baseline	-55.1 (31.36)	-53.7 (31.93)	-50.2 (31.35)	-52.4 (33.90)	-40.2 (29.79)
LS Mean (SE) difference from placebo	-16.58 (3.44)	-15.07 (3.39)	-10.79 (3.41)	-11.71 (3.36)	---
MMRM P-value vs placebo	< 0.001	< 0.001	0.002	< 0.001	---

Source: Study report, Table 19. LS mean difference from placebo and P-values are obtained from MMRM model, which included baseline frequency as covariate, treatment, study week, and treatment-by-study week interaction as fixed factors, and subject as the repeated measure unit.

Figure 1: Mean Change from Baseline in the Weekly Frequency of Moderate to Severe VMS (MITT-VMS Population)



Source: Study report, Figure 7.

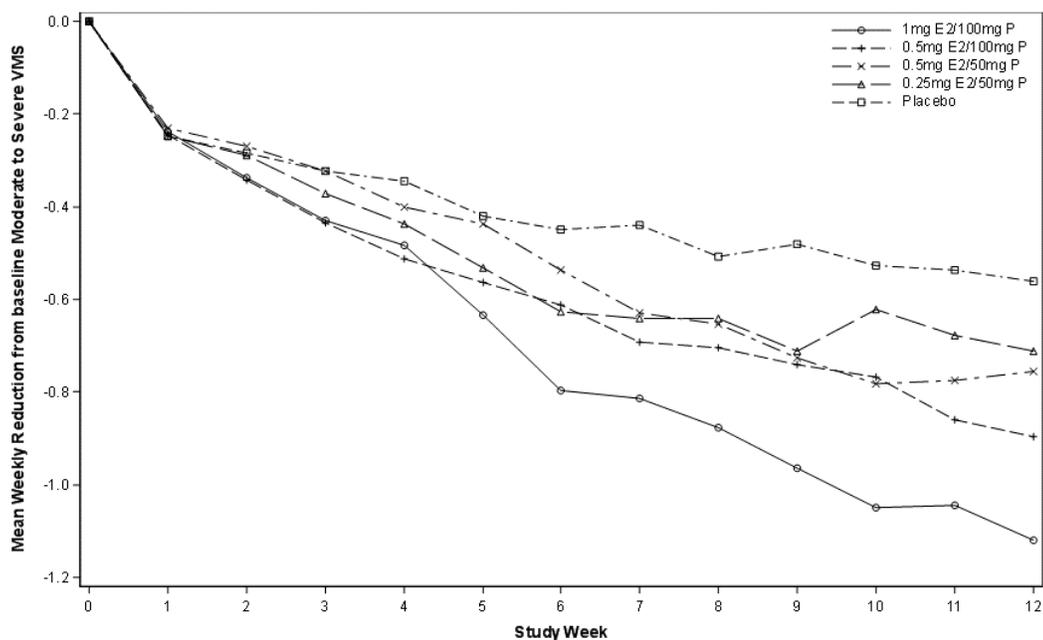
The baseline mean severity of weekly moderate to severe VMS was approximately 2.5 in each treatment group. The 1 mg E2/100 mg P and 0.5 mg E2/100 mg P groups both showed statistically significant decrease in the severity of weekly moderate to severe VMS from baseline compared to the placebo group at Weeks 4 and 12 (Week 4: -0.13 for 1 mg E2/100 mg P; -0.17 for 0.5 mg E2/100 mg P; Week 12: -0.57 for 1 mg E2/100 mg P; -0.39 for 0.5 mg E2/100 mg P).

Table 4: Change from Baseline in the Mean Weekly Severity Scores of VMS at Week 4 and Week 12 (MITT-VMS Population)

	1 mg E2/ 100 mg P (N=141)	0.5 mg E2/ 100 mg P (N=149)	0.5 mg E2/ 50 mg P (N=147)	0.25 mg E2/ 50 mg P (N=154)	Placebo (N=135)
Baseline					
Mean (SD)	2.54 (0.320)	2.51 (0.249)	2.50 (0.231)	2.51 (0.262)	2.52 (0.246)
Week 4 (n)	134	144	142	152	126
Mean (SD) change from Baseline	-0.48 (0.547)	-0.51 (0.563)	-0.40 (0.469)	-0.44 (0.514)	-0.34 (0.386)
LS Mean (SE) difference from placebo	-0.13 (0.061)	-0.17 (0.060)	-0.05 (0.060)	-0.10 (0.059)	---
MMRM P-value vs placebo	0.031	0.005	0.401	0.100	---
Week 12 (n)	124	129	124	135	115
Mean (SD) change from Baseline	-1.12 (0.963)	-0.90 (0.783)	-0.76 (0.744)	-0.71 (0.806)	-0.56 (0.603)
LS Mean (SE) difference from placebo	-0.57 (0.100)	-0.39 (0.099)	-0.24 (0.100)	-0.16 (0.098)	---
MMRM P-value vs placebo	< 0.001	< 0.001	0.018	0.096	---

Source: Study report, Table 20. LS mean difference from placebo and P-values are obtained from MMRM model, which included baseline severity as covariate, treatment, study week, and treatment-by-study week interaction as fixed factors, and subject as the repeated measure unit.

Figure 2: Change from Baseline in the Weekly Severity of Moderate to Severe VMS (MITT-VMS Population)



Source: Figure 8 in study report.

Sensitivity analyses of the co-primary endpoints were conducted using LOCF imputation for missing data points, (e.g., from subjects who were withdrawn prematurely or discontinued from the treatment). The analyses results were similar to those noted in the MMRM analyses.

3.3 Evaluation of Safety

The clinical reviewer identified 3 cases of endometrial hyperplasia (subjects [REDACTED] (b) (6) [REDACTED] which were not counted by the Applicant. The one-sided upper 95% confidence limit for the incidence rate at Month 12 was less than 4% for all groups (1.97% for 1 mg E2/100 mg P; 1.83% for 0.5 mg E2/100 mg P; 1.81% for 0.5 mg E2/50 mg P; 1.34% for 0.25 mg E2/50 mg P; and 3.93% for the placebo group).

Table 5: Incidence of Endometrial Hyperplasia at 12 Months (ES Population)

	1 mg E2/ 100 mg P (N=281)	0.5 mg E2/ 100 mg P (N=303)	0.5 mg E2/ 50 mg P (N=306)	0.25 mg E2/ 50 mg P (N=274)	Placebo (N=92)
Hyperplasia incidence rate (%)	1/281 (0.36)	1/303 (0.33)	1/306 (0.33)	0/274 (0.00)	0/92 (0.00)
One-sided upper 95% CL	1.97%	1.83%	1.81%	1.34%	3.93%

Source: FDA's analysis using PROC FREQ in SAS.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

Efficacy of Estradiol and Progesterone combination doses were also explored by subgroups defined by race (White, Black/African American, Others), age, (<55 years and ≥55 years), BMI (<25, 25 to <30, and ≥30 kg/m²), parity (nulliparous or parous). Analyses of each co-primary efficacy endpoint by subgroups were performed using the same MMRM model described previously in section 3.2.2.1 with additional terms for subgroup and treatment by subgroup interaction as appropriate.

4.1 Gender, Race, Age, and Geographic Region

The study was conducted in the U.S. and enrolled female subjects only and most subjects were not Hispanic or Latino; therefore, analysis by gender, ethnicity and geographical region was not performed.

Most subjects were either White or Black/African American. Of note, there were very low numbers of subjects in the 'Other' category (American Indian or Alaska Native, Native Hawaiian or Pacific Islander), and thus limiting further analysis including the latter categories.

Baseline, the mean change from baseline, and the LS mean change from placebo in the frequency of moderate to severe VMS at Weeks 4 and 12 for White and Black/African American subgroups are shown in Table 6. Baseline frequency was similar for the two racial groups. However, 1 mg E2/100 mg P or 0.5 mg E2/100 mg P showed no effect on frequency reduction in the Black/African American compared to placebo.

Table 6: Change from Baseline in the Number of Weekly Moderate to Severe VMS at Week 4 and Week 12 by Race (MITT-VMS Population)

	1 mg E2/ 100 mg P (N=141)	0.5mg E2/ 100 mg P (N=149)	0.5 mg E2/ 50 mg P (N=147)	0.25 mg E2/ 50 mg P (N=154)	Placebo (N=135)
Black or African American					
Baseline (n)	45	48	43	48	41
Mean (SD)	75.4 (47.93)	70.0 (19.56)	77.2 (28.55)	77.6 (28.11)	74.6 (27.01)
Week 4 (n)	43	46	41	46	37
Mean (SD) change from Baseline	-32.5 (32.47)	-30.9 (26.03)	-31.1 (39.69)	-40.5 (30.19)	-32.1 (34.94)
LS Mean (SE) difference from placebo	-0.55 (6.60)	1.10 (6.49)	3.73 (6.67)	-6.87 (6.50)	---
95% CI	(-13.57, 12.46)	(-11.70, 13.90)	(-9.41, 16.88)	(-19.69, 5.95)	
MMRM P-value vs placebo	0.933	0.866	0.576	0.292	---
Week 12 (n)	39	43	32	41	34
Mean (SD) change from Baseline	-45.3 (31.56)	-50.1 (26.09)	-42.1 (38.56)	-54.2 (29.53)	-48.8 (29.27)
LS Mean (SE) difference from placebo	0.24 (6.59)	-5.68 (6.46)	1.85 (6.72)	-7.71 (6.50)	---
95% CI	(-12.76, 13.23)	(-18.42, 7.06)	(-11.39, 15.10)	(-20.53, 5.11)	
MMRM P-value vs placebo	0.971	0.380	0.783	0.237	---
White					
Baseline (n)	95	99	99	102	91
Mean (SD)	73.8 (27.83)	72.9 (31.04)	75.6 (28.49)	75.6 (31.15)	71.9 (21.73)
Week 4 (n)	90	96	97	102	87
Mean (SD) change from Baseline	-44.0 (28.93)	-37.2 (30.69)	-35.5 (25.95)	-37.3 (31.51)	-24.4 (22.81)
LS Mean (SE) difference from placebo	-18.07 (3.74)	-12.35 (3.70)	-9.42 (3.69)	-11.45 (3.66)	---
95% CI	(-25.43, -10.72)	(-19.62, -5.08)	(-16.67, -2.17)	(-18.63, -4.27)	
MMRM P-value vs placebo	< 0.001	< 0.001	0.011	0.002	---
Week 12 (n)	84	84	88	92	79
Mean (SD) change from Baseline	-59.4 (30.47)	-56.1 (34.63)	-53.5 (28.01)	-50.7 (35.04)	-36.7 (29.66)
LS Mean (SE) difference from placebo	-23.98 (3.96)	-20.01 (3.92)	-16.57 (3.90)	-13.35 (3.87)	---
95%CI	(-31.76, -16.19)	(-27.72, -12.31)	(-24.24, -8.90)	(-20.95, -5.74)	
MMRM P-value vs placebo	< 0.001	< 0.001	< 0.001	< 0.001	---

Source: Study report, Table 34. The "Other" race subgroup was not presented.

Baseline severity of moderate to severe VMS was similar across racial groups. At Week 4, there was no treatment effect on severity reduction in Black/African American subjects treated with 1mg E2/100 mg P or 0.5 mg E2/100 mg P; and at Week 12, the treatment effect was much smaller in Black/African American subjects vs. White subjects in the 1 mg E2/100 mg P group compared to placebo group (see Table 7).

Table 7: Change from Baseline in the Severity of Weekly Moderate to Severe VMS at Week 4 and Week 12 by Race (MITT-VMS Population)

	1 mg E2/ 100 mg P (N=141)	0.5mg E2/ 100 mg P (N=149)	0.5 mg E2/ 50 mg P (N=147)	0.25 mg E2/ 50 mg P (N=154)	Placebo (N=135)
Black or African American					
Baseline (n)	45	48	43	48	41
Mean (SD)	2.52 (0.427)	2.48 (0.245)	2.54 (0.241)	2.49 (0.267)	2.52 (0.203)
Week 4 (n)	43	46	41	46	37
Mean (SD) change from Baseline	-0.35 (0.546)	-0.49 (0.421)	-0.43 (0.482)	-0.52 (0.560)	-0.51 (0.427)
LS Mean (SE) difference from placebo	0.14 (0.103)	-0.01 (0.101)	0.08 (0.104)	-0.05 (0.101)	---
95% CI	(-0.06, 0.34)	(-0.21, 0.19)	(-0.12, 0.29)	(-0.25, 0.15)	
MMRM P-value vs placebo	0.181	0.946	0.428	0.634	---
Week 12 (n)	39	43	32	41	34
Mean (SD) change from Baseline	-1.06 (0.991)	-0.93 (0.716)	-0.53 (0.594)	-0.81 (0.780)	-0.63 (0.656)
LS Mean (SE) difference from placebo	-0.38 (0.171)	-0.36 (0.167)	0.06 (0.176)	-0.22 (0.169)	---
95% CI	-0.71, -0.04	-0.69, -0.03	-0.29, 0.40	-0.55, 0.11	
MMRM P-value vs placebo	0.028	0.033	0.747	0.197	---
White					
Baseline (n)	95	99	99	102	91
Mean (SD)	2.56 (0.255)	2.53 (0.247)	2.48 (0.231)	2.51 (0.257)	2.52 (0.267)
Week 4 (n)	90	96	97	102	87
Mean (SD) change from Baseline	-0.54 (0.537)	-0.53 (0.622)	-0.39 (0.472)	-0.41 (0.496)	-0.28 (0.347)
LS Mean (SE) change from placebo	-0.25 (0.075)	-0.25 (0.074)	-0.09 (0.073)	-0.12 (0.073)	---
95% CI	(-0.39, -0.10)	(-0.39, -0.10)	(-0.24, 0.05)	(-0.27, 0.02)	
MMRM P-value vs placebo	0.001	< 0.001	0.199	0.091	---
Week 12 (n)	84	84	88	92	79
Mean (SD) change from Baseline	-1.13 (0.954)	-0.90 (0.820)	-0.83 (0.767)	-0.68 (0.824)	-0.53 (0.579)
LS Mean (SE) difference from placebo	-0.66 (0.125)	-0.43 (0.124)	-0.34 (0.123)	-0.16 (0.122)	---
95% CI	(-0.90, -0.42)	(-0.67, -0.19)	(-0.59, -0.10)	(-0.40, 0.08)	
MMRM P-value vs placebo	< 0.001	< 0.001	0.005	0.188	---

Source: Study Report, Table 35. The “Other” racial subgroup was not presented.

Applicant’s Additional Analyses

The potential reasons for the disparity in efficacy between White and Black/African American subjects were further investigated by the Applicant. Per FDA’s request, the Applicant conducted additional analyses to assess the potential impact of outliers, baseline demographics and subject characteristics, compliance with drug, diary completion and study sites.

The results for these analyses ruled out the possibility that the disparity is due to the impact from any data outliers or any of the above factors.

The potential effect of study drug compliance for Week 12, defined as the number of capsules taken between Study Day 1 and 84 divided by the number of capsules expected for the respective treatment period (168), was also investigated. Compliance was calculated for all subjects regardless of whether the subject completed or discontinued the study during the respective time periods. The diary compliance was determined based on data recorded by subjects in daily diaries.

Overall, numerically more Black/African American subjects were < 80% compliant with taking study drug at Week 12 than White subjects based on diary data (16.4% versus 11.9%). The mean compliance rate at Week 12 was 92% for White subjects versus 89% for Black/African American subjects. Since compliance rates were based on diary data, the percentage of subjects who completed their diaries at Week 12 was also evaluated. Approximately 91% of White subjects completed their diaries compared with 88% of Black/African American women. Overall, slightly more Black/African American women were < 80% compliant in completing their diaries than White women (16% versus 12%). The rate of discontinuation was similar between the White and Black/African American groups (28.0% versus 29.8%).

For the baseline demographics and characteristics, the mean age between White women (54.9 years) and Black/African American women (54.1 years) were similar. The Black/African American subgroup had a mean BMI of 27.93 kg/m² and the White subgroup had a mean BMI of 26.05 kg/m². Additionally, more Black/African American women were current smokers than White women (34.2% versus 18.9%) for the overall population. And, more White women reported current use of alcohol than Black/African American women (63.2% versus 50.7%). Baseline estradiol concentrations were numerically higher in Black/African American women than in White women (~6.6 versus 5.5 pg/mL).

Based on the individual analyses performed, no single factor appeared to independently explain the differences seen by race subgroups in the co-primary endpoints. To further explore the impact of multiple covariates on efficacy, the Applicant used an analysis of covariance (ANCOVA) interaction model to assess the impact of the above factors that were found to be different in the race subgroups (Baseline BMI categories: <25, 25 to <30, ≥30 kg/m², current smoking and alcohol use, and baseline estradiol concentration categories: <15, ≥15 pg/mL). The endpoints assessed were the change from baseline to Weeks 4 and 12 in weekly frequency by race (White and Black/African American). The ANCOVA model examined the treatment effect within each racial subgroup (least square mean difference) as well as between the subgroups (comparison of least square mean difference for White versus Black/African American).

In the ANCOVA analyses without adjusting the baseline characteristics, P-values the treatment by race interaction term at Week 4 and Week 12 were 0.1983 and 0.0119 and the P-values for the same interaction after adjusting these baseline variables, were 0.4593 and 0.2491. The detailed results for Week 12 are presented in Table 16 and Table 17. At Week 12, in Black/African American, the treatment effect in the 1 mg E2/100 mg P on the frequency reduction was -10.2 (-28.7, 8.2) after adjusting for other baseline characteristics compared to -1.0 (-12.1, 10.1) without adjustment. While in white subjects, the treatment effect in the 1 mg E2/100 mg P on the frequency reduction was -23.8 (-41.4, -6.2) after adjusting for other baseline characteristics compared to -20.9 (-28.3, -13.5) without adjustment.

The Applicant concluded that *“A significant treatment by race interaction was observed for change from Baseline to Week 12 in weekly frequency. However, when controlling for confounding factors utilizing multivariate analysis, Baseline BMI, Baseline smoking, Baseline alcohol use, and Baseline estradiol concentration and their interactions with the treatment group, there was no longer a statistically significant difference between racial subgroups. These data suggest that the impact of Race on efficacy is multifactorial and not dependent on a single covariate.”*

FDA’s additional analyses

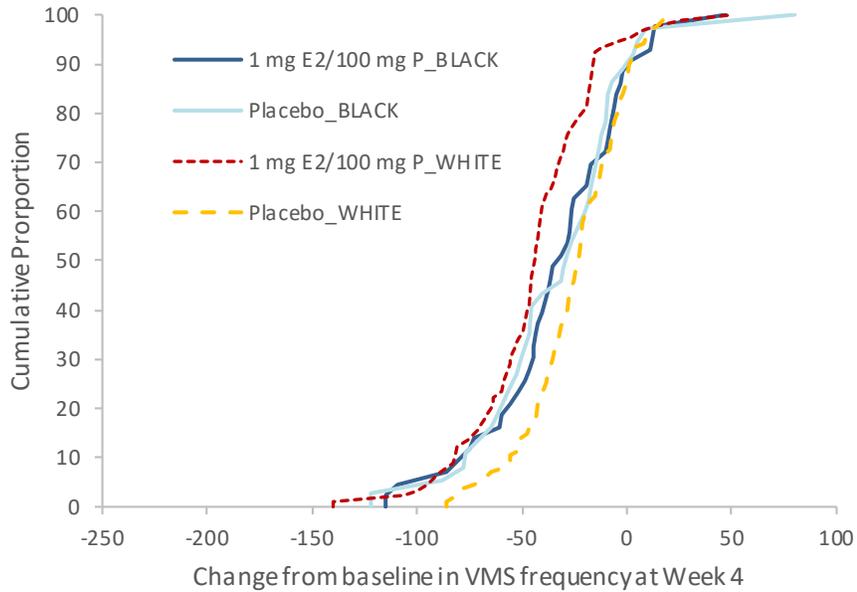
The FDA reviewer also used Bayesian shrinkage estimation approach to estimate the treatment effects in the racial subgroups, which confirmed the disparity that was observed between the racial groups.

The FDA reviewer found that Applicant’s conclusion may not be appropriate because the impact of baseline BMI and estradiol concentration may not be completely captured in the Applicant’s analyses due to the categorization of these two variables instead of using the original continuous data. Overall, the 1 mg E2/100 mg P achieved statistical and clinical significance on all four efficacy endpoints and this combination was the only potential candidate that was considered for approval by the agency. Therefore, to rule out the effect of interaction due to other combination dose groups with race, only the White and Black/ African American treated 1 mg E2/100 mg P and placebo were considered in FDA’s analyses.

Figure 3 to Figure 6 are the cumulative distribution function (CDF) curves of each of the co-primary efficacy endpoint by race in 1mg E2/100 mg P and placebo groups respectively. It appears that in at least part of the graph the placebo effect in Black/African American subjects was higher compared to White subjects with respect to reduction in frequency. For Black subjects, except the change from baseline in severity at Week 12, the CDF curves do not appear to separate from each other, which indicates that the distributions are similar for the two treatment groups.

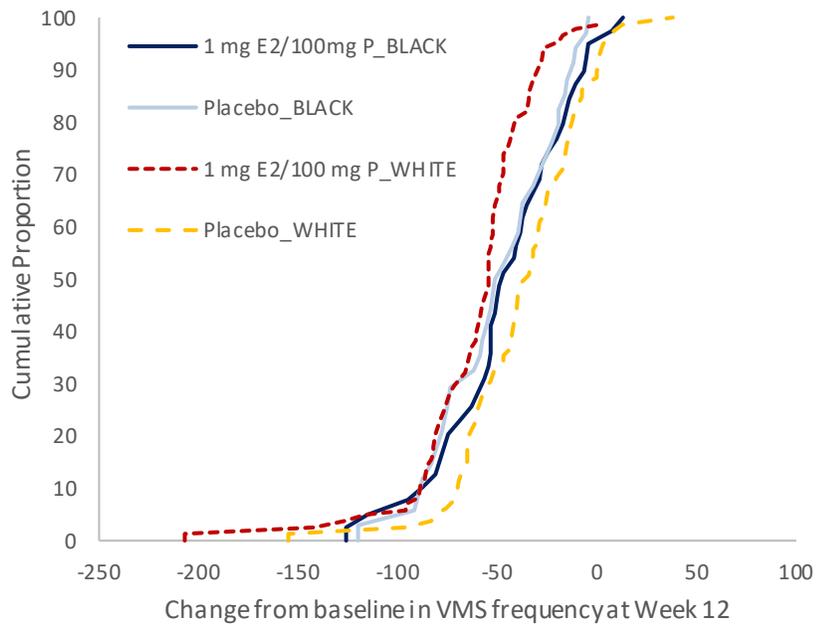
The CDF curves also showed that the White subjects treated with 1mg E2/100 mg P consistently had more reduction on VMS frequency compared to the Black subjects taking the same treatment. The same trend was observed for VMD severity reduction at Week 4, but not at Week 12. Therefore, we do not agree with the Applicant that “the disparity in efficacy by Race, predominantly due to the high placebo response” as claimed.

Figure 3: CDF for Change from Baseline in Weekly VMS frequency at Week 4 by Race and Treatment (MITT-VMS population)



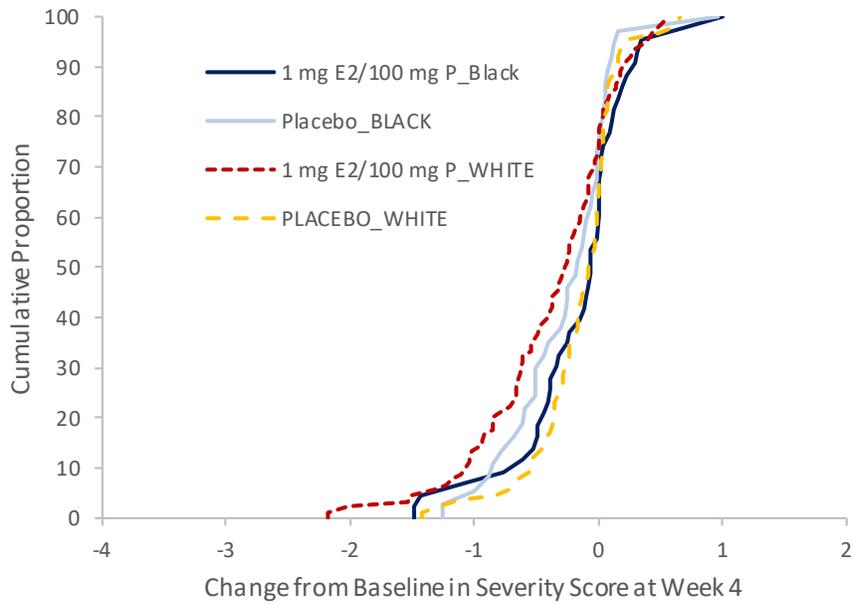
Source: FDA’s analysis

Figure 4: CDF for Change from Baseline in Weekly VMS frequency at Week 12 by Race and Treatment (MITT-VMS population)



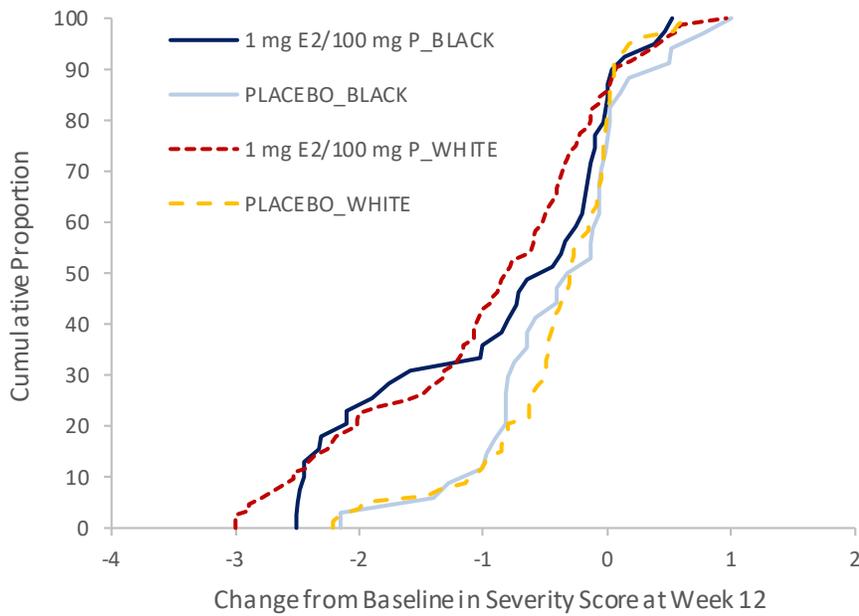
Source: FDA’s analysis

Figure 5: CDF for Change from Baseline in Weekly VMS Severity at Week 4 by Race and Treatment (MITT-VMS population)



Source: FDA’s analysis

Figure 6: CDF for Change from Baseline in Weekly VMS Severity at Week 12 by Race and Treatment (MITT-VMS population)



Source: FDA’s analysis

FDA reviewer conducted the same ANCOVA analyses for the frequency co-primary endpoints for White and Black/African American subjects in 1 mg E2/100 mg P and placebo group, similar to the Applicant's analyses except that the continuous baseline BMI and estradiol level were used instead of the categories. In addition, MMRM analyses were also carried out for sensitivity assessment. For each type of the above analysis, both unadjusted and adjusted (for the four factors and their interaction with treatment) models are considered.

Both ANCOVA and MMRM analyses show consistently that after adjusting for the four baseline factors and their interactions with treatment, there is still strong disparity in the efficacy between White and Black/African American subjects on the frequency endpoints by race.

At Week 4, in Black/African American subjects, the treatment effect in the 1 mg E2/100 mg P on the frequency reduction was -2.5 (-14.4, 9.3) after adjusting for other baseline characteristics compared to -3.4 (-14.7, 7.9) without adjustment. While in white subjects, the treatment effect in the 1 mg E2/100 mg P on the frequency reduction was -12.3 (-22.0, -2.6) after adjusting for other baseline characteristics compared to -18.7 (-26.2, -11.1) without adjustment. At Week 12, in Black/ African American subjects, the treatment effect in the 1 mg E2/100 mg P on the frequency reduction was 1.0 (-9.9, 11.9) after adjusting for other baseline characteristics compared to -1.4 (-12.1, 9.2) without adjustment of additional covariates. While in white subjects, the treatment effect in the 1 mg E2/100 mg P on the frequency reduction was -11.6 (-20.6, -2.5) after adjusting for other baseline characteristics compared to -20.7 (-27.9, -13.7) without adjustment of additional covariates.

The same analyses were conducted for the severity co-primary endpoints as well. Consistently, in the Black/African American subgroup, (see Table 12 to Table 15), reduction in VMS severity was only seen at Week 12 and it was similar to the reduction in the white subgroup, but not at Week 4.

Based on all analyses, we conclude that the imbalances in the four baseline factors did not explain the treatment differences as the Applicant claimed. The underlying cause for the disparity of the treatment effect in White and Black/African American subjects remains unknown.

Table 8: Change from baseline in Weekly VMS frequency at Weeks 4 and 12 - ANCOVA analysis

Subgroup		Week 4		Week 12	
		1 mg E2 / 100 mg P	Placebo	1 mg E2 / 100 mg P	Placebo
Black	LS mean	-34.3	-30.9	-48.3	-46.9
	95% CI	-42.0, -26.6	-39.2, -22.6	-55.6, -41.1	-54.6, -39.1
	LS mean Difference	-3.4		-1.4	
	95% CI	-14.7, 7.9		-12.1, 9.2	
White	LS mean	-43.3	-24.6	-58.2	-37.4
	95% CI	-48.7, -38.0	-30.0, -19.3	-63.1, -53.2	-42.5, -32.3
	LS mean Difference	-18.7		-20.7	
	95% CI	-26.2, -11.1		-27.9, -13.7	
P-value for Trt*Race		0.0274		0.0032	

For each week, the ANCOVA model included baseline frequency, treatment, race, and treatment by race interaction.

Source: FDA's analysis.

Table 9: Change from baseline in Weekly VMS frequency at Weeks 4 and 12 - ANCOVA analysis adjusting for more baseline characteristics

		Week 4		Week 12	
		1 mg E2 / 100 mg P	Placebo	1 mg E2 / 100 mg P	Placebo
Black	LS mean	-32.6	-30.0	-45.8	-46.8
	95% CI	-40.6, -25.4	-38.7, -21.3	-53.3, -38.3	-54.7, -38.9
	LS mean Difference	-2.5		1.0	
	95% CI	-14.4, 9.3		-9.9, 11.9	
White	LS mean	-39.0	-26.7	-52.0	-40.4
	95% CI	-45.9, -32.2	-33.7, -19.8	-58.4, -45.7	-47.0, -33.9
	LS mean Difference	-12.3		-11.6	
	95% CI	-22.0, -2.6		-20.6, -2.5	
P-value for Trt*Race		0.1970		0.0754	

For each week, the ANCOVA model included baseline frequency, treatment, race, current alcohol use (Yes/No), tobacco use (Yes, /No), baseline BMI and baseline estradiol level and the interactions with treatment.

Source: FDA's analysis.

Table 10: Change from baseline in Weekly VMS frequency at Weeks 4 and 12 - MMRM analysis

Subgroup		Week 4		Week 12	
		1 mg E2 / 100 mg P	Placebo	1 mg E2 / 100 mg P	Placebo
Black	LS mean	-33.0	-32.4	-44.5	-44.5
	95% CI	-40.9, -25.1	-40.9, -25.1	-52.9, -36.1	-52.9, -36.1
	LS mean Difference	-0.5		-0.1	
	95% CI	-12.0, 10.9		-12.3, 12.2	
White	LS mean	-43.0	-24.8	-59.8	-35.6
	95% CI	-48.4, -37.5	-30.3, -19.3	-65.6, -54.0	-41.5, -29.7
	LS mean Difference	-18.2		-24.2	
	95% CI	-25.9, -10.4		-33.0, -15.9	
P-value for Trt*Race: 0.0032					

MMRM model includes Treatment, Week (1-12), race, Treatment-by-race interaction, Treatment-by-Week interaction, Treatment-by-race-by-Week interaction as factors, Baseline as covariate, and Subject as repeated measures unit.

Source: FDA's analysis.

Table 11: Change from baseline in Weekly VMS frequency at Weeks 4 and 12 - MMRM analysis adjusting for more baseline characteristics

Subgroup		Week 4		Week 12	
		1 mg E2 / 100 mg P	Placebo	1 mg E2 / 100 mg P	Placebo
Black	LS mean	-32.7	-30.9	-44.3	-42.9
	95% CI	-40.8, -24.6	-39.4, -22.3	-52.9, -35.7	-52.1, -33.8
	LS mean Difference	-1.9		-1.4	
	95% CI	-13.6, 9.9		-13.9, 11.2	
White	LS mean	-40.1	-26.4	-57.0	-37.2
	95% CI	-46.5, -33.7	-32.7, -20.0	-63.6, -50.3	-43.9, -30.5
	LS mean Difference	-13.7		-19.8	
	95% CI	-22.7, -4.7		-29.2, -10.3	
P-value for Trt*Race: 0.0510					

MMRM model includes Treatment, Week (1-12), race, current alcohol use (Yes/No), tobacco use (Yes, /No), baseline BMI and baseline estradiol level and the interactions with treatment and the interaction between treatment with week and race, Baseline as covariate, and Subject as repeated measures unit.

Source: FDA's analysis.

Table 12: Change from baseline in Weekly VMS severity at Weeks 4 and 12 - ANCOVA analysis

Subgroup		Week 4		Week 12	
		1 mg E2 / 100 mg P	Placebo	1 mg E2 / 100 mg P	Placebo
Black	LS mean	-0.17	-0.28	-0.85	-0.40
	95% CI	-0.20, -0.04	-0.42, -0.13	-1.10, -0.60	-0.67, -0.12
	LS mean Difference	0.11		-0.46	
	95% CI	-0.09, 0.30		-0.83, -0.08	
White	LS mean	-0.36	-0.14	-0.96	-0.40
	95% CI	-0.45, -0.27	-0.24, -0.05	-1.13, -0.78	-0.57, -0.12
	LS mean Difference	-0.22		-0.56	
	95% CI	-0.35, -0.09		-0.81, -0.31	
P-value for Trt*Race		0.0066		0.6363	

For each week, the ANCOVA model included baseline frequency, treatment, race, and treatment by race interaction.

Source: FDA's analysis.

Table 13: Change from baseline in Weekly VMS severity at Weeks 4 and 12 - ANCOVA analysis adjusting for more baseline characteristics

		Week 4		Week 12	
		1 mg E2 / 100 mg P	Placebo	1 mg E2 / 100 mg P	Placebo
Black	LS mean	-0.13	-0.26	-0.74	-0.39
	95% CI	-0.27, 0.01	-0.41, -0.11	-1.0, -0.48	-0.67, -0.11
	LS mean Difference	0.13		-0.35	
	95% CI	-0.07, 0.34		-0.74, 0.03	
White	LS mean	-0.25	-0.14	-0.62	-0.42
	95% CI	-0.36, -0.13	-0.26, -0.02	-0.84, -0.40	-0.65, -0.19
	LS mean Difference	-0.10		-0.20	
	95% CI	-0.27, 0.07		-0.52, 0.12	
P-value for Trt*Race		0.0757		0.5327	

For each week, the ANCOVA model included baseline frequency, treatment, race, current alcohol use (Yes/No), tobacco use (Yes/No), baseline BMI and baseline estradiol level and the interactions with treatment.

Source: FDA's analysis.

Table 14: Change from baseline in Weekly VMS severity at Weeks 4 and 12 - MMRM analysis

Subgroup		Week 4		Week 12	
		1 mg E2 / 100 mg P	Placebo	1 mg E2 / 100 mg P	Placebo
Black	LS mean	-0.16	-0.25	-0.82	-0.36
	95% CI	-0.29, -0.03	-0.38, -0.11	-1.07, -0.56	-0.63, -0.09
	LS mean Difference	0.09		-0.45	
	95% CI	-0.11, 0.28		-0.82, -0.09	
White	LS mean	-0.35	-0.15	-1.00	-0.38
	95% CI	-0.45, -0.27	-0.24, -0.05	-1.18, -0.83	-0.55, -0.20
	LS mean Difference	-0.21		-0.63	
	95% CI	-0.34, -0.08		-0.88, -0.38	
P-value for Trt*Race: 0.0291					

MMRM model includes Treatment, Week (1-12), race, Treatment-by-race interaction, Treatment-by-Week interaction, Treatment-by-race-by-Week interaction as factors, Baseline as covariate, and Subject as repeated measures unit.

Source: FDA's analysis.

Table 15: Change from baseline in Weekly VMS severity at Weeks 4 and 12 - MMRM analysis adjusting for more baseline characteristics

Subgroup		Week 4		Week 12	
		1 mg E2 / 100 mg P	Placebo	1 mg E2 / 100 mg P	Placebo
Black	LS mean	-0.14	-0.24	-0.80	-0.34
	95% CI	-0.27, -0.001	-0.38, -0.10	-1.05, -0.55	-0.61, -0.07
	LS mean Difference	0.10		-0.46	
	95% CI	-0.09, 0.29		-0.83, -0.09	
White	LS mean	-0.31	-0.12	-0.96	-0.35
	95% CI	-0.41, -0.22	-0.21, -0.02	-1.14, -0.79	-0.53, -0.17
	LS mean Difference	-0.20		-0.61	
	95% CI	-0.33, -0.06		-0.86, -0.37	
P-value for Trt*Race: 0.0341					

MMRM model includes Treatment, Week (1-12), race, current alcohol use (Yes/No), tobacco use (Yes, /No), baseline BMI and baseline estradiol level and the interactions with treatment and the interaction between treatment with week and race, Baseline as covariate, and Subject as repeated measures unit.

Source: FDA's analysis.

4.2 Other Special/Subgroup Populations

In both studies, analyses of each co-primary efficacy endpoint were also performed for subgroups based on baseline BMI (<25, ≥25 to <30 and >30 kg/m²) and parity (nulliparous, parous). The treatment effects are similar across the subgroups.

5 SUMMARY AND CONCLUSIONS

5.1 Conclusions and Recommendations

The data from study TXC12-05 showed that

1. The 1 mg E2/100 mg P and 0.5 mg E2/100 mg P combinations demonstrated statistically significant reductions from baseline in the weekly frequency and severity of moderate to severe VMS at Week 4 and Week 12 compared to placebo. The 1 mg E2/100 mg P dose achieved the clinical meaningful threshold of reducing at least 14 hot flushes per week starting from Week 5 and maintained through Week 12. However, the 0.5mg E2/100mg P dose did not achieve this threshold until Week 9.
2. Despite the significant efficacy results in the overall study population and similar results in the subgroup of White subjects, there was no efficacy seen in the other major subgroup of Black/African American subjects on the VMS frequency reduction at Weeks 4 and 12 and reduction in VMS severity was only seen at Week 12, not at Week 4. However, the study was not powered to demonstrate efficacy by subgroup of race.
3. The incidence rate of endometrial hyperplasia in each active treatment group was less than a clinically acceptable rate of 4% (limit of the upper bound of the 95% confidence interval).

From a statistical perspective, the data supports the efficacy of 1 mg E2/100 mg P dose for the treatment of moderate to severe vasomotor symptoms associated with menopause. However, the efficacy in the Black/African American subgroup was not observed, the review team recommends that this finding needs to be included in the label, should the Division decide to approve this product.

APPENDICES

Demographics

Table 16: Selected Demographic and Baseline Characteristics for Safety Population

	1 mg E2/ 100 mg P (N=415)	0.5 mg E2/ 100 mg P (N=424)	0.5 mg E2/ 50 mg P (N=421)	0.25 mg E2/ 50 mg P (N=424)	Placebo (N=151)	Total (N=1835)
Age (years)						
Mean (SD)	54.7 (4.37)	54.5 (4.52)	54.9 (4.27)	54.4 (4.04)	54.5 (4.32)	54.6 (4.31)
Median	55	54	55	54	54	54
Min, Max	40, 65	43, 66	41, 65	43, 65	45, 65	40, 66
Race, n(%)						
White	271 (65.3)	281 (66.3)	276 (65.6)	273 (64.4)	100 (66.2)	1201 (65.4)
Black or African American	134 (32.3)	136 (32.1)	133 (31.6)	140 (33.0)	46 (30.5)	589 (32.1)
Other ^a	10 (2.4)	7 (1.6)	12 (2.8)	11 (2.6)	5 (3.3)	45 (2.4)
Weight (kg)	n = 415	n = 424	n = 421	n = 423	n = 151	n = 1834
Mean (SD)	72.1 (12.32)	71.7 (13.07)	72.2 (11.79)	72.1 (11.93)	71.4 (11.48)	72.0 (12.21)
Median	72	72	72	73	71	72
Min, Max	41, 106	39, 105	41, 110	41, 100	45, 98	39, 110
Height (cm)	n = 415	n = 424	n = 421	n = 423	n = 151	n = 1834
Mean (SD)	163.7 (6.61)	163.6 (6.95)	164.1 (6.47)	164.0 (6.47)	163.5 (6.11)	163.8 (6.58)
Median	163	164	165	163	164	164
Min, Max	147, 181	137, 183	148, 180	147, 185	146, 180	137, 185
BMI (kg/m²)	n = 415	n = 424	n = 421	n = 423	n = 151	n = 1834
Mean (SD)	26.81 (4.122)	26.67 (4.344)	26.74 (3.977)	26.72 (4.005)	26.63 (3.870)	26.72 (4.091)
Median	26.	27.	26.6	26.8	26.	26.
Min, Max	14.0, 34.2	15.2, 34.5	15.5, 34.4	16.9, 34.3	16.0, 34.1	14.0, 34.5

^a Other includes: Other (n=20), Asian (n=12), American Indian or Alaska Native (n=6), Native Hawaiian or Pacific Islander (n=5), and Unknown (n=2).

Source: Study report, Table 13.

Table 17: Gynecological History for Safety Population

	1 mg E2/ 100 mg P (N=415)	0.5 mg E2/ 100 mg P (N=424)	0.5 mg E2/ 50 mg P (N=421)	0.25 mg E2/ 50 mg P (N=424)	Placebo (N=151)	Total (N=1835)
Time Since Last Menstrual Period (years)						
Mean (SD)	5.8 (4.86)	6.0 (5.10)	5.7 (4.55)	5.6 (4.93)	6.0 (5.30)	5.8 (4.90)
Median	4.5	4.6	4.5	4.2	4.1	4.4
Min, Max	0.5, 32.5	0.5, 30.3	0.5, 24.5	0.2, 30.6	0.5, 28.7	0.2, 32.5
Bilateral Oophorectomy, n (%)						
No	411 (99.0)	418 (98.6)	418 (99.3)	421 (99.3)	151 (100.0)	1819 (99.1)
Yes	4 (1.0)	6 (1.4)	3 (0.7)	3 (0.7)	0	16 (0.9)
Parity, n (%)						
Nulliparous	70 (16.9)	66 (15.6)	62 (14.7)	51 (12.0)	25 (16.6)	274 (14.9)
Parous	345 (83.1)	358 (84.4)	359 (85.3)	373 (88.0)	126 (83.4)	1561 (85.1)
Number of Pregnancies, n (%)						
0	36 (8.7)	38 (9.0)	28 (6.7)	32 (7.5)	15 (9.9)	149 (8.1)
≥ 1	379 (91.3)	386 (91.0)	393 (93.3)	392 (92.5)	136 (90.1)	1686 (91.9)
Number of Vaginal Births, n (%)						
0	34 (9.0)	28 (7.3)	34 (8.7)	19 (4.8)	10 (7.4)	125 (7.4)
≥ 1	345 (91.0)	358 (92.7)	359 (91.3)	373 (95.2)	126 (92.6)	1561 (92.6)

Source: Study Report, Table 14.

Table 18: Selected Demographic and Baseline Characteristics for MITT-VMS

	1 mg E2/ 100 mg P (N=141)	0.5 mg E2/ 100 mg P (N=149)	0.5 mg E2/ 50 mg P (N=147)	0.25 mg E2/ 50 mg P (N=154)	Placebo (N=135)	Total (N=726)
Age (years)						
Mean (SD)	54.7 (4.80)	54.9 (4.45)	54.8 (4.63)	54.5 (3.78)	54.3 (4.29)	54.6 (4.39)
Median	55	55	55	54	54	54
Min, Max	40, 65	45, 65	41, 65	45, 65	45, 65	40, 65
Race, n (%)						
White	95 (67.4)	99 (66.4)	99 (67.3)	102 (66.2)	91 (67.4)	486 (66.9)
Black or African American	45 (31.9)	48 (32.2)	43 (29.3)	48 (31.2)	41 (30.4)	225 (31.0)
Other ^a	1 (0.7)	2 (1.3)	5 (3.4)	4 (2.6)	3 (2.2)	15 (2.1)
Weight (kg)						
Mean (SD)	71.7 (12.47)	72.7 (13.19)	72.0 (11.43)	71.1 (11.68)	71.7 (11.24)	71.9 (12.01)
Median	71	72	72	72	71	71
Min, Max	41, 106	45, 105	49, 100	45, 95	45, 98	41, 106
Height (cm)						
Mean (SD)	164.3 (6.96)	163.7 (7.49)	164.4 (6.51)	163.9 (6.41)	163.8 (6.05)	164.0 (6.70)
Median	164	163	165	163	165	164
Min, Max	149, 180	137, 183	151, 180	150, 185	146, 180	137, 185
BMI (kg/m²)						
Mean (SD)	26.45 (3.935)	27.05 (4.333)	26.57 (3.943)	26.42 (3.983)	26.64 (3.817)	26.63 (4.006)
Median	26.1	27.0	26.6	26.2	26.8	26.8
Min, Max	14.0, 34.2	18.2, 34.5	18.0, 34.4	17.0, 34.3	16.0, 34.1	14.0, 34.5

Source: Study report, Table 16.

Table 19: Gynecological History for MITT-VMS Population

	1 mg E2/ 100 mg P (N=141)	0.5 mg E2/ 100 mg P (N=149)	0.5 mg E2/ 50 mg P (N=147)	0.25 mg E2/ 50 mg P (N=154)	Placebo (N=135)	Total (N=726)
Time Since Last Menstrual Period (years)						
Mean(SD)	6.1 (5.53)	6.5 (5.43)	6.0 (4.82)	5.2 (4.75)	5.7 (4.92)	5.9 (5.10)
Median	4.4	5.2	5.3	4.0	3.9	4.6
Min, Max	0.5, 32.5	0.5, 30.3	0.5, 24.5	0.2, 30.6	0.5, 28.7	0.2, 32.5
Bilateral Oophorectomy, n (%)						
No	138 (97.9)	146 (98.0)	146 (99.3)	153 (99.4)	135 (100.0)	718 (98.9)
Yes	3 (2.1)	3 (2.0)	1 (0.7)	1 (0.6)	0	8 (1.1)
Parity, n (%)						
Nulliparous	21 (14.9)	25 (16.8)	20 (13.6)	25 (16.2)	24 (17.8)	115 (15.8)
Parous	120 (85.1)	124 (83.2)	127 (86.4)	129 (83.8)	111 (82.2)	611 (84.2)
Number of Pregnancies, n (%)						
0	8 (5.7)	16 (10.7)	8 (5.4)	16 (10.4)	15 (11.1)	63 (8.7)
≥ 1	133 (94.3)	133 (89.3)	139 (94.6)	138 (89.6)	120 (88.9)	663 (91.3)
Number of Vaginal Births, n (%)						
0	133	133	139	138	120	663
0	13 (9.8)	9 (6.8)	12 (8.6)	9 (6.5)	9 (7.5)	52 (7.8)
≥ 1	120 (90.2)	124 (93.2)	127 (91.4)	129 (93.5)	111 (92.5)	611 (92.2)

Source: Study report, Table 17.

Applicant's additional Analyses

Table 20: Applicant's analysis on Change from Baseline in Weekly Frequency of Moderate to Severe VMS at Week 12 – unadjusted ANCOVA Model (MITT-VMS Population) by race (white and Black/ African American only)

Race Group	Statistics	1 mg E2/ 100 mg P	0.5 mg E2/ 100 mg P	0.5 mg E2/ 50 mg P	0.25 mg E2/ 50 mg P	Placebo
Black	LS Mean	-49.2	-52.4	-40.8	-51.7	-48.2
	95% CI	(-56.8, -41.6)	(-59.6, -45.2)	(-49.1, -32.4)	(-59.1, -44.3)	(-56.3, -40.1)
	LS Mean difference vs Placebo	-1.0	-4.2	7.4	-3.6	---
	95% CI for LSM difference	(-12.1, 10.1)	(-15.1, 6.7)	(-4.3, 19.0)	(-14.5, 7.4)	---
	P-value	0.8555	0.4472	0.2135	0.5229	---
White	LS Mean	-59.4	-56.3	-52.0	-49.6	-38.5
	95% CI	(-64.6, -54.2)	(-61.4, -51.1)	(-57.1, -47.0)	(-54.6, -44.7)	(-43.8, -33.1)
	LS Mean difference vs Placebo	-20.9	-17.8	-13.5	-11.2	---
	95% CI for LSM difference	(-28.3, -13.5)	(-25.2, -10.4)	(-20.9, -6.2)	(-18.4, -3.9)	---
	P-value	< 0.0001	< 0.0001	0.0003	0.0026	---
P-value for treatment by race interaction: 0.0119						

Source: clinical-information-amendment.pdf, submitted on 7/12/2018, Tables 13 and 14.

Table 21: Applicant’s analysis on Change from Baseline in Weekly Frequency of Moderate to Severe VMS at Week 12 – adjusted ANCOVA Model (MITT-VMS Population) by race (white and Black/ African American only)

Race Group	Statistics	1 mg E2/ 100 mg P	0.5 mg E2/ 100 mg P	0.5 mg E2/ 50 mg P	0.25 mg E2/ 50 mg P	Placebo
Black	LS Mean	-39.5	-49.2	-43.8	-51.3	-29.3
	95% CI	(-51.1, -27.9)	(-60.9, -37.5)	(-57.1, -30.5)	(-63.2, -39.4)	(-43.6, -15.0)
	LS Mean difference vs Placebo	-10.2	-19.9	-14.5	-22.0	---
	95% CI for LSM difference	(-28.7, 8.2)	(-38.5, -1.4)	(-34.0, 5.1)	(-40.7, -3.4)	---
	P-value	0.2777	0.0349	0.1467	0.0208	---
White	LS Mean	-47.3	-53.0	-51.5	-49.5	-23.5
	95% CI	(-58.0, -36.7)	(-64.3, -41.7)	(-62.9, -40.2)	(-61.5, -37.5)	(-37.5, -9.6)
	LS Mean difference vs Placebo	-23.8	-29.4	-28.0	-26.0	---
	95% CI for LSM difference	(-41.4, -6.2)	(-47.4, -11.5)	(-46.0, -10.0)	(-44.4, -7.6)	---
	P-value	0.0081	0.0014	0.0023	0.0058	---
P-value for treatment by race interaction: 0.2491						

Source: clinical-information-amendment.pdf, submitted on 7/12/2018, Tables 15 and 16.

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/

JIA GUO
10/16/2018

MAHBOOB SOBHAN
10/16/2018