Efficacy Study Part
Statistical Analysis Plan (SAP)

Protocol Title:	A Randomized Double-blind Placebo Controlled Phase 3 Trial to evaluate the Efficacy and Safety of Estetrol for the Treatment of Moderate to Severe Vasomotor Symptoms in Postmenopausal Women (E4Comfort Study II)
Protocol Number:	MIT-Do001-C302
Protocol Version, Date	Final Version 8.0 (Amendment 7), 21 June 2022
ICON ID:	3803/0007
Document Version, Date:	Final 4.0, 10Oct2022

Prepared by:

ICON Clinical Research Services

On behalf of:

Estetra SRL

SIGNATURE PAGE

Prepared at ICON Clinical Research by:

[redacted]

Date
(DD Mmm YYYY)

Reviewed at ICON Clinical Research by:

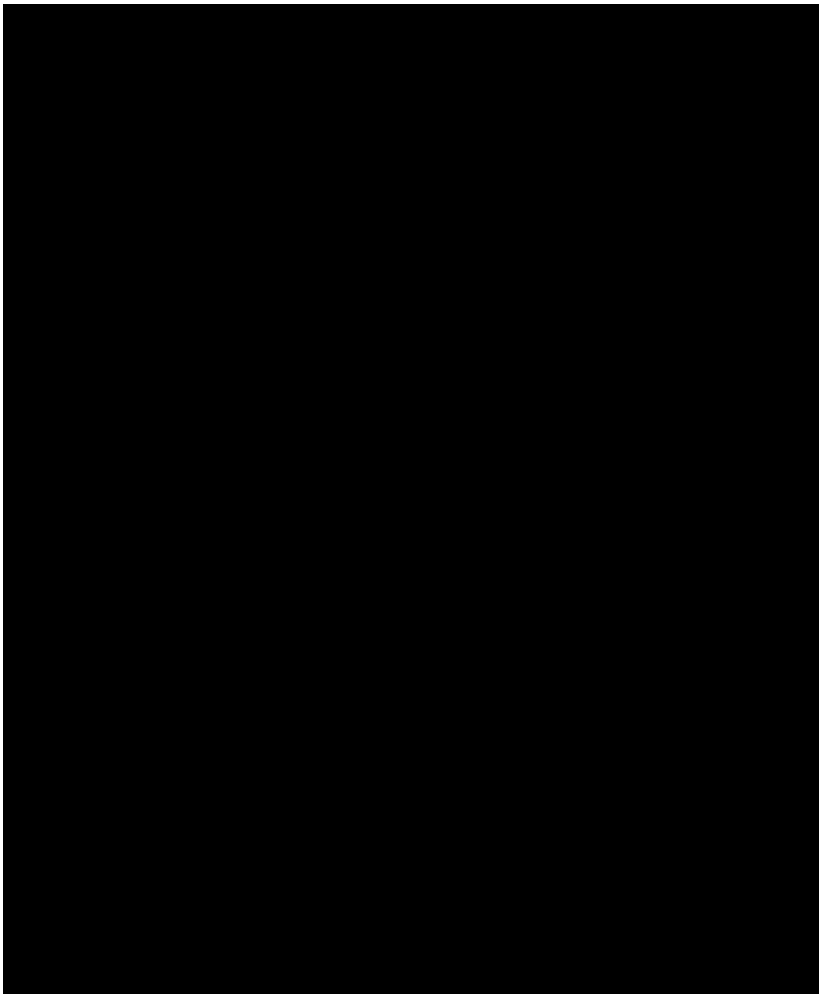
[redacted]

Date
(DD Mmm YYYY)

[redacted]

Date
(DD Mmm YYYY)

Approved at Estetra SRL by:



Date
(DD Mmm YYYY)

Date
(DD Mmm YYYY)

Date
(DD Mmm YYYY)

REVISION HISTORY

Version/Date	Version name	Section	Changes implemented
11Nov2020	Final 1.0		Initial version
09Jul2021		1 - 10	• Updated based on protocol version 7.0 (amendment 6.0) dated 26May2021 • Updated statistical sections to match with protocol
09Jul2021		5.3	• Intent-to-Treat (ITT) analysis set: Added additional information, "Based on handling of missing VMS diary information only subjects who have at least 4 days of VMS diary data for one on-treatment week following initiation of treatment will be included in primary analysis population." • Added additional analysis set "Endometrial Safety Analysis Set"
09Jul2021		5.8	• Removed the term "Violations" from the title and also form the text
09Jul2021		6.1	• Added derivations for the below variables ✓ Study Start Day ✓ Study Stop Day ✓ Therapy Start Day ✓ Therapy Stop Day ✓ Weekly treatment compliance ✓ Cycle ✓ AE Start Day ✓ AE Stop Day ✓ AE Duration • Updated derivations for efficacy endpoints • Added derivation for internal method • Added derivation for mean severity score
09Jul2021		6.2	Added section for "Missing MENQOL and CGI Data"
05Aug2021		1 - 10	Update per the latest protocol version 7.1
23Aug2021		1 - 10	Update per sponsor's comments
14Sep2021		1 - 10	Update per sponsor's comments
22Sep2021		1-10	Update per sponsor's comments
28Sep2021		1-10	Update per sponsor's comments
30Sep2021	2.0	6.1	Definition of Weekly frequency of moderate to severe VMS at Baseline is updated to consider the days from -7 to -1
30Sep2021	2.0	7.4.1	Removed Sex from the description as the population only includes female subjects
30Sep2021	2.0	1-10	Formatting updates
14Mar2022		7.7.6.6	Update section 7.7.6.6 per "Donesta_EMB_reading_process_20211122"
5 April 2022	3.0	7.6.3.2	Update MMRM to ANCOVA, per sponsor's comments
10Oct2022	4.0		• Added required abbreviations in list of abbreviations. • Changed 'patient' to 'subject'. • Added table numbers for all tables. • Changed 'population' to 'set'.
10Oct2022	4.0	1	• Updated protocol version and date. • Added latest CRF as reference.
10Oct2022	4.0	6.1	• Removed study start and stop date, post therapy medication and study duration definitions. • Updated treatment start and stop dates, study day, duration of exposure, study week, study cycle, AE start and stop

			days, TEAE, change from baseline and weekly frequency of VMS definitions.
10Oct2022	4.0	6.2.3, 6.2.4	• Updated text.
10Oct2022	4.0	7.2	• Added text to present subjects in efficacy study part in disposition table. • Updated text for denominator for percentage.
10Oct2022	4.0	7.4.2	• Added text to BI-RADS score in baseline characteristics table.
10Oct2022	4.0	7.4.4	• Updated prior and concomitant medications definition.
10Oct2022	4.0	7.6.3.4.1	• Updated MENQOL domain scores and total score definition.
10Oct2022	4.0	7.7	• Removed the text regarding endometrial events.
10Oct2022	4.0	7.7.1	• Added the list of event categories of VTE and cardiovascular events. • Updated PY definition.
10Oct2022	4.0	7.7.6.3	• Added text to present summary of breast density. • Removed text mentioning to present shift table for breast density.
10Oct2022	4.0	7.7.6.6	• Added text for endometrial events.
10Oct2022	4.0	7.7.6.6.1	• Added descriptions for biopsies and diagnoses.
10Oct2022	4.0	7.7.6.6.2	• Replaced ‘all visits’ to ‘post-baseline visits’ in the case of worst-case analysis for Final/Consensus diagnosis. • Added text for reader’s agreement analysis.
10Oct2022	4.0	7.7.6.8	• Added exclusion criteria for amenorrhea tables.
10Oct2022	4.0	Appendix 2	• Replaced the schedule of trial procedures table with the table in the latest protocol version 8.0.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse Event
AESI	Adverse Event of Special Interest
ANCOVA	analysis of covariance
APCr	activated Protein C resistance
APCsr	activated Protein C sensitivity ratio
aPTT	activated partial thromboplastin time
ATC	anatomical therapeutic chemical
BI-RADS	Breast Imaging-Reporting and Data System
BLQ	below the lower limit of quantification
BMI	body mass index
BP	blood pressure
CGI	clinical global impression
CI	confidence interval
CID	clinically important difference
CRF	Case Report Form
CRO	Contract Research Organization
CSR	Clinical Study Report
CTMS	clinical trial management system
CTX-1	C-terminal telopeptide type 1
E2	Estradiol
E4	Estetrol
EC	Ethics Committee
ECG	electrocardiogram
ED	early discontinuation
EoT	end of treatment
ETP	endogenous thrombin potential
FSH	Follicle stimulating hormone
HDL	high-density lipoprotein
HOMA-IR	homeostasis model assessment-estimated insulin resistance
HRQoL	health-related quality of life

ICF	informed consent form
ICH	International Conference on Harmonisation
IRB	Institutional Review Board
ITT	Intent-to-Treat
IWRS	Interactive Web Response System
LDH	lactate dehydrogenase
LDL	low-density lipoprotein
LOCF	last observation carried forward
max	maximum
MCID	minimally clinically important difference
MedDRA	Medical Dictionary for Regulatory Activities
MENQOL	menopause specific quality of life
min	minimum
mITT	Modified Intent-to-Treat
MMRM	Mixed-effect Models for Repeated Measures
NH	non-hysterectomized
P4	Progesterone
PAP	Papanicolaou
PINP	procollagen I N-propeptide
PP	Per-Protocol
PT	preferred term
PY	patient year
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	standard deviation
SHBG	Sex Hormone Binding Globulin
SI	International System of Units
SOC	system organ class
TEAE	Treatment Emergent Adverse Event
TFLs	tables, figures and listings
TS	treatment satisfaction
TSH	Thyroid stimulating hormone
TVUS	transvaginal ultrasound

ULQ	upper limit of quantification
VMS	vasomotor symptoms
VTE	venous thromboembolism events
WHO	World Health Organization

1 INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to provide detailed descriptions of the statistical methods, data derivations and data displays for the Efficacy part of study based on protocol MIT Do001-C302 version 8.0 "A Randomized Double-blind Placebo Controlled Phase 3 Trial to evaluate the Efficacy and Safety of Estetrol (E4) for the Treatment of Moderate to Severe Vasomotor Symptoms in Postmenopausal Women (E4Comfort Study II)" dated 21Jun2022 and Case Report Form (CRF) MIT-Do001-C302_aCRF_version 3.0 for Clinical Study Report (CSR) analysis. The table of contents and templates for the tables, figures and listings (TFLs) will be produced in a separate document.

Any deviations from this SAP will be described and justified in the CSR.

The preparation of this SAP has been based on International Conference on Harmonisation (ICH) E9 guidelines.

All data analyses and generation of TFLs will be performed using SAS Version 9.4® or higher.

2 STUDY OBJECTIVES AND ENDPOINTS

2.1 Primary Objective and Endpoints

To measure the effect of treatment with E4 15 mg or E4 20 mg compared to placebo on the frequency and severity of moderate to severe vasomotor symptoms (VMS) in postmenopausal women at 4 and 12 weeks.

Co-primary efficacy endpoints for primary objective #1

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

2.2 Secondary Objectives and Endpoints

1) To measure the effect of treatment with E4 15 mg or E4 20 mg compared to placebo on the frequency and severity of mild, moderate and severe VMS in postmenopausal women weekly up to 12 weeks.

Efficacy endpoints for secondary objective #1

- Change from baseline to week 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 in the weekly frequency and severity of moderate to severe VMS
- Change from baseline to weeks 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 in the weekly frequency and severity of mild, moderate, and severe VMS
- Percentage of subjects with 50% and 75% reduction from baseline in the weekly frequency of moderate to severe VMS at weeks 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12
- Percentage of subjects with 50% and 75% reduction from baseline in the weekly frequency of mild, moderate and severe VMS at weeks 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12

2) To measure the clinical meaningfulness of E4 15 mg or E4 20 mg compared to placebo on the reduction of VMS at weeks 4 and 12.

Efficacy endpoint for secondary objective #2

- Percentage of subjects with a clinically important difference (CID) compared to baseline in the weekly frequency of moderate to severe VMS after weeks 4 and 12 using the clinical global impression (CGI) questionnaire

3) To evaluate the effect of treatment with E4 15 mg or E4 20 mg compared to placebo on hemostasis, lipid and glucose metabolism and bone turnover.

Efficacy endpoint for secondary objective #3

- Change from baseline to week 12 and week 52 in prothrombin fragment 1 + 2, endogenous thrombin potential (ETP)-based activated Protein C sensitivity ratio (APCsr), activated partial thromboplastin time (aPTT) based activated Protein C resistance (APCr), anti-thrombin III, Protein-C, free Protein-S, Factor VIII, angiotensinogen, and Sex Hormone Binding Globulin (SHBG)
- Change from baseline to week 12 and week 52 in triglycerides, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, total cholesterol, lipoprotein (a), total cholesterol/HDL-cholesterol ratio, fasting glycemia, insulin, glycated hemoglobin and homeostasis model assessment-estimated insulin resistance (HOMA-IR)
- Change from baseline to week 12 and week 52 in procollagen I N-propeptide (PINP), C-terminal telopeptide type 1 (CTX-1), calcium, and vitamin D

4) To evaluate the effect of treatment with E4 15 mg or E4 20 mg compared to placebo on health-related quality of life (HRQoL) and treatment satisfaction (TS).

Efficacy endpoints for secondary objective #4

- Change from baseline to week 12 and week 52 in HRQoL using the menopause specific quality of life (MENQOL) questionnaire
- Total score in TS after 4, 12, and 52 weeks of treatment using the CGI questionnaire

5) To evaluate the general safety of treatment with E4 15 mg or E4 20 mg compared to placebo

Safety endpoints for secondary objective #5

- Frequency of treatment emergent adverse events (TEAEs) (including treatment emergent serious adverse events [SAEs])
- Frequency of changes in results in physical and, gynecological examination, vital signs, electrocardiogram (ECG), mammography and breast examination at each measured time point
- Frequency of changes in routine clinical laboratory tests results (hematology and chemistry) at each measured time point

6) To evaluate the effect of treatment with E4 15 mg or E4 20 mg on the endometrium in non-hysterectomized (NH) subjects compared to placebo

Safety endpoints for secondary objective #6

- Change from baseline to each measured time point in endometrial thickness measured by ultrasound
- Frequency of subjects in the different endometrial categories according to the Blaustein's pathology

7) To evaluate the effect of treatment with E4 15 mg or E4 20 mg on vaginal bleeding in NH subjects compared to placebo

Safety endpoints for secondary objective #7

- Frequency of women with vaginal bleeding and/or spotting during each 28-day cycle of treatment with E4 based on recording in the subject diary
- Number of days with bleeding and/or spotting during each 28-day cycle of treatment based on recording in the subject diary
- Frequency of women with amenorrhea (absence of any bleeding or spotting) during each 28-day cycle of treatment with E4 based on recording in the subject diary
- Cumulative rates of amenorrhea defined as the percentage of women who reported consecutive cycles of amenorrhea for a given cycle of time

2.3 Exploratory Objectives and Endpoints

1) To evaluate the effect of treatment with E4 15 mg or E4 20 mg compared to placebo on breast density

Safety endpoint for exploratory objective #1

- Change in breast density from digitized mammography readings from baseline to end of treatment (EoT) in subjects who had a paired digitized mammography

2) To evaluate the influence of E4 on the changes in the hemostasis parameters compared to placebo in women with and without inherited thrombophilia

Efficacy endpoint for exploratory objective #2

- Change in hemostasis parameters in women with and without inherited thrombophilia (Factor V Leiden mutation and prothrombin G20210A mutations).

3 STUDY DESIGN

3.1 General Study Design

This study is a Phase III study, with multiple centers in the USA and Canada. Eligible subjects will be post-menopausal women, ≥ 40 up to ≤ 65 years of age, seeking treatment for relief of moderate to severe VMS associated with menopause.

Efficacy Study part will be conducted on hysterectomized and NH healthy women (approximately 50% of the subjects are anticipated to have a history of hysterectomy). Subjects enrolled into the Efficacy Study part should be subjects (~300 hysterectomized and ~300 NH) with at least 7 moderate to severe bothersome VMS per day or at least 50 moderate to severe bothersome VMS a week in the last 7 consecutive days during the screening period.

The Efficacy Study part has a randomized, double-blinded placebo-controlled design to evaluate the primary efficacy (frequency and severity of VMS), secondary efficacy (effect on hemostasis, lipid and glucose metabolism, bone turnover, HRQoL and TS) and safety of E4 in both hysterectomized and NH postmenopausal women, as well as the effect of E4 on the endometrium in NH postmenopausal women.

A total of 600 subjects will be randomly allocated (1:1:1) to one of the 3 arms and will receive E4 15 mg (Arm 1), E4 20 mg (Arm 2) or Placebo (Arm 3). During the first 12 weeks, the effect of E4 on VMS will be evaluated. Thereafter, treatment will proceed for a total duration of 12 months (53 weeks), to continue the evaluation of secondary efficacy, safety and the effect on the endometrium. For endometrial protection, all NH subjects will receive treatment with 200 mg Progesterone (P4) once daily for 14 consecutive days, after completion of the E4/placebo treatment.

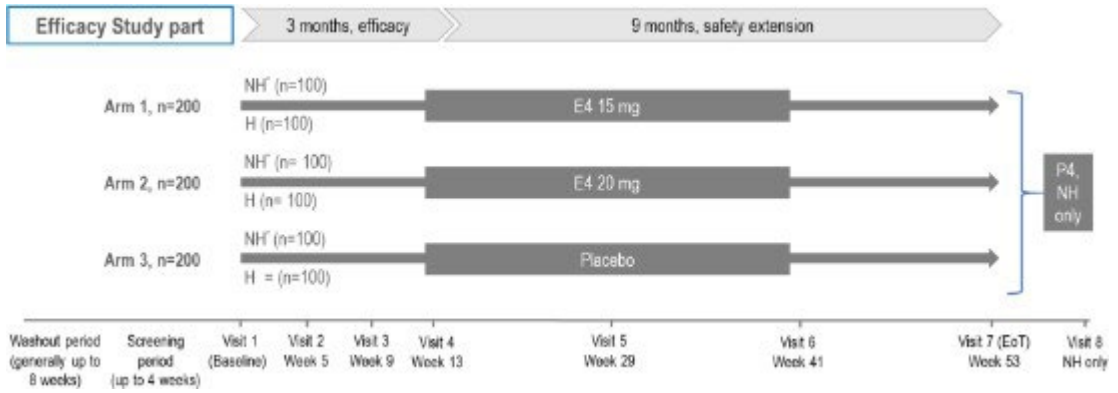
At least 9 visits (up to 15 months) are planned, this includes:

- Washout visit (occurring generally up to 8 weeks prior to Screening visit),
- Screening visit (occurring 4 weeks prior to Baseline [Visit 1]),
- Treatment visits (occurring within 53 weeks, Visit 2 to Visit 7), include:
    - Six On-treatment visits (Visits 2, 3, 4, 5, 6 and 7).
    - The last treatment visit (Visit 7) is the EoT Visit, which will also occur in case of early discontinuation (ED).
- For NH women an additional endometrial safety follow-up visit (Visit 8) is planned after completion of the 14 days of P4 treatment, which will also occur in case of ED.

Bothersome is defined as a score of 4 to 6 on question 1 of the modified MENQOL questionnaire at screening. The timing and number of VMS and modified MENQOL score needed for allocation will not be disclosed to the study subjects to avoid bias in the recording during the Screening process.

The trial schedule is summarized in Figure 1.

Figure 1: Trial Schedule



* For non-hysterectomized (NH) subjects，endometrial thickness will be measured after 3, 6, 9 and 12 months of treatment (Visits 4, 5, 6 and 7 respectively) by transvaginal ultrasound (TVUS) and subjects will record daily the vaginal bleeding events. If at any t ime after the first 4 weeks of treatment a subject presents with persistent and/or recurrent bleeding, confirmed by the investigator or presents with a bi-layer endometrial thickness >10 mm as assessed by TVUS, an endometrial biopsy will be performed. If the biopsy shows hyperplasia or worse, the study drug will be discontinued, and the subject will be withdrawn from the study. The subject will be treated with P4 200 mg once daily for 14 days. If the endometrial event has not resolved, treatment with a progestin will be started according to local practice/guidelines. All NH subjects will receive treatment with 200 mg P4 for two weeks after completion of E4/placebo treatment.

At the Washout visit, after signing the informed consent form (ICF), the inclusion and exclusion criteria will be checked as well as the prior medication list. If a subject uses medication prohibited by the protocol, then this subject will be asked to stop this medication. After a variable washout period (depending on type of prior medication) the subject will be asked to return to start screening. The Washout and Screening visits may occur at the same time for subjects who do not use prohibited medication.

The screening period requires at least one visit, to occur as per protocol, at the beginning of the period, but some assessments (e.g., mammography, biopsy, transvaginal ultrasound [TVUS], ECG and Papanicolaou [PAP] test) may require additional visits if these assessments cannot be done during the same visit.

During the first 12 weeks of treatment, primary efficacy assessments (the number and severity of VMS) will be recorded on a daily basis in a paper diary. At baseline (Visit 1), secondary efficacy assessments (hemostasis, lipid and glucose metabolism and bone turnover) will be performed and the MENQOL questionnaires will be completed. For NH subjects, endometrial thickness will be assessed after 3, 6, 9 and 12 months of treatment (Visits 4, 5, 6 and 7 respectively). Subjects will visit the trial center after 4 and 8 weeks of treatment (Visit 2 and Visit 3) for general safety assessments, for review of the diaries and for completion of the CGI questionnaire (Visit 2 only). After 12 weeks of treatment, subjects will visit the trial center for the final primary efficacy assessment (Visit 4). During this visit also secondary efficacy assessments and general safety assessments will be performed and the MENQOL and CGI questionnaires will be completed. For NH subjects, endometrial thickness will be assessed. Subjects will visit the trial center again after 6 months of treatment (Visit 5) for general safety assessments and endometrial thickness assessment (NH subjects only) and after 9 months of treatment (Visit 6) for general safety assessments and endometrial thickness assessment (NH

subjects only). After 12 months of treatment, subjects will visit the trial center for the EoT visit (Visit 7) during which secondary efficacy assessments and general safety assessments will be performed, the HRQoL and TS questionnaires will be completed and a mammography will be made. For NH subjects, endometrial thickness will be assessed and if they have completed at least 12 weeks of E4/placebo treatment, an endometrial biopsy will be performed. At Visits 1, 2, 3 and 4, blood samples will be taken to monitor estradiol (E2) levels and at Visits 2, 3 and 4 also to monitor E4 levels.

After the EoT procedures, all NH subjects will receive treatment with 200 mg P4 for 14 days. Approximately 10 days after completion of the progestin treatment, endometrial thickness will be measured (Visit 8).

3.2 Randomization and Blinding

3.2.1 Randomization

A randomization schedule will be generated within the Biometrics Department of the designated Contract Research Organization (CRO) by a statistician not involved in the trial. Subjects eligible for participation in the Efficacy Study part will be randomly allocated to one of the three treatment arms [E4 15 mg (Arm 1), E4 20 mg (Arm 2) or placebo (Arm 3)] in a 1:1:1 allocation ratio. The randomization schedule will be generated by means of the PLAN procedure of SAS®. The computer program uses the method of randomly permuted blocks. The block size will not be revealed before unblinding.

An Interactive Web Response System (IWRS) will be used for the randomization procedure. A subject ready for randomization will be entered in the IWRS. The respective next subject eligible for randomization will receive the lowest available randomization number at the trial center. A randomization confirmation including treatment assignment (i.e., kit number), date and time will be sent to the center, to the designated project managers or designee of the CRO and to the Sponsor's project manager.

The randomization schedule and the complete generation procedure will be filed at a secured place by the designated CRO until the trial database is unblinded. A copy of the list will be sent to the Pharmaceutical Supply for the purpose of assigning the kits to the subjects.

3.2.2 Blinding

This study will have a double-blind placebo-controlled design. Subjects will receive one of the two dosages of E4 or placebo in a blinded manner.

The Investigator, trial center personnel, CRO or Sponsor must maintain the blinding through database lock and must not break the blinding via IWRS or code break envelopes without a valid reason (e.g., in case of emergency). Emergency unblinding is to be done only when knowing the study drug is absolutely necessary for the management of an individual subject and where stopping the blinded medication is not sufficient in the judgment of the Investigator. The Medical Monitor is to be notified by the Investigator when unblinding of a subject's treatment is being considered or has already occurred.

3.3 Study Treatments and Assessments

After enrolled, the trial medication will be dispensed (see Section 9.4 in protocol) and subjects need to take trial medication (E4 or placebo, oral, once daily) on the day of randomization visit (= day 1 of week 1) and during the treatment period (Visit 2 to Visit 7) and only NH women will take 14 days of P4 treatment in Visit 8.

- E4 15 mg (Arm 1)
- E4 20 mg (Arm 2)
- Placebo (Arm 3)

Subjects will be instructed to take the trial medication at approximately the same time each day. If a subject has missed a dose, she should be instructed to take the dose as soon as she remembers. It is acceptable to take a dose up to 12 hours late. If taken more than 12 hours late the subjects should omit dosing for that day and continue with the next dose as usual.

An overview and time windows for each visit is presented in Appendix1; detailed description of procedures and assessments to be conducted during this study is summarized in Appendix2.

4 SAMPLE SIZE AND POWER

The sample size of the efficacy part of the Phase 3 study was estimated by simulation of the 15-mg dose based on the data collected in the phase 2 study.

The 4 co-primary endpoints were first recomputed and analyzed with the methodology planned for the Phase 3 study. The corresponding estimates of the different endpoints (change from baseline) are summarized in the Table 1 below; values are model estimates (standard error).

Table 1: Estimates (Standard Error) of Endpoints

VMS co-primary endpoint	E4 15 mg	Placebo	E4 15 mg - Placebo
Frequency			
Week 4	-45.45 (3.11)	-33.97 (2.91)	-11.46 (4.22)
Week 12	-55.64 (3.29)	-46.95 (3.17)	-8.69 (4.53)
Severity (FDA method)			
Week 4	-0.81 (0.09)	-0.39 (0.09)	-0.42 (0.13)
Week 12	-1.39 (0.10)	-0.81 (0.10)	-0.58 (0.14)

For each endpoint, 1000 Phase 2 studies were simulated using the Mixed-effect Models for Repeated Measures (MMRM) estimates. In addition, the dropout rate was directly incorporated into the simulation code. For hysterectomized women, the dropout rate was set to 20% as observed in the Phase 2 data. For NH women, the dropout rate was increased to 35%, accounting for the potential higher risk of dropout related to bleeding or other endometrial events.

Based on these simulations, the maximum number of subjects to be included in the efficacy part is 600. A sample size of 200 subjects per arm provides a power of at least 90% to simultaneously achieve a superiority (upper bound of a 95% 2-sided confidence interval [CI] lower than zero) over placebo for the four co-primary endpoints together. It is planned to enroll approximately 100 hysterectomized and 100 NH subjects per treatment arm.

5 ANALYSIS SETS

5.1 Enrolled Set

Enrolled Set will include all subjects enrolled in the Overall MIT-Do001-C302 study who provide informed consent (signed informed consent).

5.2 Randomized Set

The Randomized Set will include all subjects randomly allocated to one of the three treatment arms in Efficacy Study part (through IWRS system).

5.3 Intent-to-Treat (ITT) Set

The ITT Set consists of all randomized subjects who receive at least one dose of randomized study medication. The ITT Set will be the primary analysis population for the efficacy analyses and all analyses on this set will be based on randomized treatment. Based on handling of missing VMS diary information only subjects who have at least 4 days of VMS diary data for one on-treatment week following initiation of treatment will be included in primary analysis.

5.4 Modified Intent-to-Treat (mITT) Set

The mITT Set consists of all randomized subjects in the ITT Set excluding subjects presenting at least one of the following features:

- At least one post baseline E4 concentration level below the lower limit of quantification (BLQ)
- At least one post screening E2 concentration level is superior to 40 pg/mL

5.5 Safety Analysis Set

The safety analysis set consists of all randomized subjects who receive at least one dose of randomized study medication. The Safety Analysis Set will be used for all analyses of safety, tolerability and background characteristics and all analyses on this set will be based on treatment received.

5.6 Per-Protocol (PP) Set

The PP set consists of all subjects in the ITT Set who do not have major protocol deviations. The major protocol deviations will be defined at the time of the evaluability assessment between the database soft lock and hard lock before unblinding. The PP Set will be used for efficacy sensitivity analyses.

5.7 Protocol Deviations and Exclusions from Analysis Sets

A protocol deviation is defined as an instance of failure to follow, intentionally or unintentionally, the requirements of the protocol procedures or requirements.

All deviations and exclusions of subjects from analysis sets will be identified at a Classification Meeting just prior the database lock and the final analysis, through clinical review input provided by Sponsor and using the Protocol Deviation Logs, provided by ICON clinical trial

management system (CTMS). Further, deviations from protocol will be classified as major or minor.

6 STATISTICAL CONSIDERATIONS AND ANALYSIS

6.1 Derived Variables

The below table (Table 2) provides the list of derived variables for demographic and baseline characteristics, various duration derivations, baseline derivations and other important derivations applicable for this study.

Table 2: Derived Variables

Variables	Formula
Demographic and Baseline characteristics	
Body mass index (BMI) (kg/m^2)	Weight (kg)/ [height (m)] ^{2}
Study Treatment	
Treatment Start Date	The date of first dose of study drug.
Treatment Stop Date	The date of last dose of study drug.
Duration of exposure (Days)	Last treatment date – First treatment date + 1
Treatment compliance (%)	100 x [(total number of tablets dispensed) – (total number of tablets returned)]/ (duration of exposure in days)
Study day	Prior to Treatment Start Date: Date of interest – Treatment Start Date On or after Treatment Start Date: Date of interest – Treatment Start Date +1
Study Week	Study Week is defined as follows for the first 7 weeks

Week	Start (Study Day)	End (Study Day)
1	1	7
2	8	14
3	15	21
4	22	28
5	29	35
6	36	42
7	43	49

Study Cycle	28-day period starting from Treatment Start Day
Adverse Events	
AE Start Day	Prior to Treatment Start Date: AE start date – Treatment Start Date On or after Treatment Start Date: AE Start Date – Treatment Start Date + 1
AE Stop Day	Prior to Treatment Start Date: AE stop date – Treatment Start Date

	On or after Treatment Start Date: AE stop date – Treatment Start Date + 1
AE Duration	(AE Stop Date – AE Start Date) + 1
TEAE	AE Start Date known: TEAE if AE Start Date on or after Treatment Start Date and AE Start Date is before the date of the last visit where last visit is defined as end of treatment, early discontinuation or follow up visit (for NH subjects). AE Start Date unknown: TEAE if AE Stop Date is on or after Treatment Start Date.
Baseline Derivations	
Baseline	Unless otherwise noted, baseline is defined as the last non missing value, including results from repeated and unscheduled measurements, recorded prior to the first dose of study drug.
Change from baseline	Date of interest – Baseline
Efficacy Parameters	
Weekly frequency of moderate to severe VMS at Baseline	Sum of all recorded moderate to severe VMS experienced during days -7 to -1.
Weekly frequency of moderate to severe VMS at Week X	Sum of all recorded moderate to severe VMS experienced during the Week X.
Daily severity score of VMS at Baseline	[((2 x number of moderate VMS) + (3 x number of severe VMS))/ (total number of moderate + severe VMS)], if at least one moderate to severe VMS was recorded during the day. In case of documented absence of moderate to severe VMS during the day, the daily severity will be set to zero.
Daily severity score of VMS at Post-baseline	<u>FDA method (for primary and secondary efficacy objectives)</u> [((1 x number of mild VMS) + (2 x number of moderate VMS) + (3 x number of severe VMS))/ (total number of mild + moderate + severe VMS)], if at least one mild to severe VMS was recorded during the day. In case of documented absence of VMS during the day, the daily severity will be set to zero. <u>EMA method (for primary and secondary efficacy objectives)</u> [((2 x number of moderate VMS) + (3 x number of severe VMS))/ (total number of moderate + severe VMS)], if at least one moderate to severe VMS was recorded during the day. In case of documented absence of moderate to severe VMS during the day, the daily severity will be set to zero. The analysis including the mild VMS will be used for

	submission to the FDA and the analysis without the mild VMS will be used for submissionto the EMA. Internal method (for secondary efficacy objectives only) [((1 x number mild VMS) + (2 x number of moderate VMS) + (3 x number of severe VMS)) / (total number of mild + moderate + severe VMS)] if at least one mild to severe VMS is recorded during the day for baseline and post baseline visits. In case of documented absence of moderate to severe VMS during the day, the daily severity score will be set to zero.
Mean (weekly) severity score of VMS by FDA method at Baseline	Arithmetic mean of the daily severity score values of VMS (moderate or severe) observed from days -7 to -1.
Mean (weekly) severity score of VMS by FDA method at Week X	Arithmetic mean of the daily severity score values of VMS (mild to severe) observed during the Week X.
Mean (weekly) severity score of VMS by EMA method at Baseline	Arithmetic mean of the daily severity score values of VMS (moderate or severe) observed from days -7 to -1.
Mean (weekly) severity score of VMS by EMA method at Week X	Arithmetic mean of the daily severity score values of VMS (moderate or severe) observed during the Week x.
Mean (weekly) severity score of VMS by Internal method at Baseline	Arithmetic mean of the daily severity score values of VMS (mild to severe) observed from days -7 to -1.
Mean (weekly) severity score of VMS by Internal method at Week X	Arithmetic mean of the daily severity score values of VMS (mild to severe) observed during the last 7 consecutive days during the Week X.
Weekly Weighted Score of VMS by EMA Method at Baseline and Week X	[(2 x number of moderate VMS) + (3 x number of severe VMS)] for baseline and post baseline visits if at least one moderate to severe VMS was recorded and zero otherwise.
Weekly Weighted Score of VMS by FDA Method at Baseline and Week X	[(2 x number of moderate VMS) + (3 x number of severe VMS)] if at least one moderate to severe VMS was recorded and zero otherwise for baseline visits; [(1 x number of mild VMS) + (2 x number of moderate VMS) + (3 x number of severe VMS)] for post baseline visits if at least one mild to severe VMS was recorded and zero otherwise.
Weekly Weighted Score of VMS by Internal Method at Baseline and Week X	[(1 x number of mild VMS) + (2 x number of moderate VMS) + (3 x number of severe VMS) for baseline and post baseline visits if at least one mild to severe VMS was recorded and zero otherwise.
CID	The CID include the ratings of much improved and very much improved based on CGI questionnaire (Gerlinger, 2012).
Minimally Clinically Important Difference (MCID)	The MCID includes the rating of minimally improved based on CGI questionnaire (Gerlinger, 2012; Jaeschke, 1989).

Definition of Responders (Moderate to Severe)	Responders are defined as subjects with 50% and, separately, 75% reduction in the weekly frequency of moderate to severe VMS from Baseline to weeks from 1 to 12 (for assessment of clinical meaningfulness).
Definition of Responders (mild to Severe)	Responders are defined as subjects with 50% and, separately, 75% reduction in the weekly frequency of mild to severe VMS from Baseline to weeks from 1 to 12 (for assessment of clinical meaningfulness).
TS	TS is the percentage of subjects in each category in CGI questionnaire (very much improved, much improved, minimally improved, no change, minimally worse, much worse, and very much worse) which will be presented at week 4, week 12 and week 52 for the different treatment groups.

6.2 Handling of Missing Data and Outliers

6.2.1 Missing Data Analysis Methods

For days with missing VMS diary information, the number and intensity of VMS will be imputed with the mean values of the remaining days of the respective 7-day period, but only if at least 4 days with VMS data in this 7-day period are available. Otherwise, no calculation of the primary and related secondary efficacy variables will be done for this period.

No method of imputation is planned for primary efficacy variable analysis since an MMRM is used.

6.2.2 Missing Data Analysis Methods for Sensitivity Analyses

For the primary endpoints, last observation carried forward (LOCF) approach will be performed as a sensitivity analysis. In this approach, a subject with missing week 4 and week 12 value will be imputed with the last non-missing value of the subject prior to calculating change from baseline.

6.2.3 Missing Menopause Specific Quality of Life and Clinical Global Impression Data

For the MENQOL domains:

- If at least 50% of the items within a domain function score were answered (defined as at least 8 out of 16 questions are answered for physical domain, at least 4 out of 7 for psychosocial domain, and at least 2 out of 3 for vasomotor and sexual domains) the domain score will be calculated as the mean of the non-missing responses. For example: Vasomotor: if Q3 is missing (i.e. 2 items are not missing), then Score = (Q1+Q2)/2.
- If >50% of the items within a domain function score are missing, then the domain score will be considered as missing.

For the MENQOL total score:

- If at least 50% (at least 2 out of 4 domains) have non-missing scores, then the MENQOL total score will be calculated as the mean of the non-missing domains.
- If >50% (3 or 4 domains) are missing, then total score will be considered as missing.

For the CGI:

- Missing data will not be imputed.

6.2.1 Handling of Missing or Incomplete Dates

Imputation rules for missing or partial Adverse Event (AE) Start Date are defined below:

If only day of AE Start Date is missing:

If the AE start year and month are the same as that for the Treatment Start Date, then:

- If the full (or partial) AE Stop Date is not before the Treatment Start Date or AE Stop Date is missing, then impute the partial AE start day as the day of Treatment Start Date;
- Otherwise, impute the partial AE start day as 1.

If day and month of AE Start Date are missing:

If AE start year = year of Treatment Start Date, then:

- If the full (or partial) AE Stop Date is not before the Treatment Start Date or AE Stop Date is missing, then impute the partial AE start month and day as the month and day of Treatment Start Date;
- Otherwise, impute the partial AE start month as January and the day as 1.

If Year of AE Start Date is missing:

If the year of AE start is missing or AE Start Date is completely missing then:

- No imputation.

If the year of AE start is missing and AE Stop Date is not missing, then:

- If the AE Stop Date is before the Treatment Start Date then the AE should be considered as a pre-treatment AE.
- Otherwise, the AE will be considered as TEAE.

Imputation rules for missing or partial medication start/stop dates are defined below:

Missing or partial medication start date:

- If only day is missing, use the first day of the month.
- If day and month are both missing, use the first day of the year.
- If day, month and year are all missing, use the date of informed consent.

Missing or partial medication stop date:

- If only day is missing, use the last day of the month.
- If day and month are both missing, use the last day of the year.
- If day, month and year are all missing, assign 'continuing' status to stop date.

6.3 COVID-19 Impact

The impact of COVID-19 will be continuously monitored during the study and the primary endpoint is not impacted as this endpoint is recorded on a daily basis in a paper diary. At the time of writing the SAP, there is no relevant impact in term of screening failures, missing data and safety concern. If the impact becomes more significant in the future, post-hoc analysis may be performed as appropriate.

7 STATISTICAL METHODS

7.1 General Statistical Conventions

All statistical tests will be two-sided with a significance level of $\alpha = 0.05$, unless specified otherwise and will be performed using SAS Version 9.4 or higher.

CIs will be presented as 2-sided 95% CIs unless specified differently in specific analysis.

Categorical variables will be summarized using frequency counts (n) and percentages in each category, percentages will be computed for subjects with non-missing value (as the denominator). If there are missing values, missing values counts will be provided. All percentages will be rounded to one digit after decimal point. The frequency counts and percentages will be presented in the form XX (XX.X %), where the percentage is in the parentheses. If the percentage of a category is '100 %', it will be displayed as '100 %' (no decimals will be added after 100).

Continuous variables will be summarized using the number of subjects with non-missing values (n), mean, median, standard deviation (SD), minimum (min), maximum (max) and missing values counts. Missing values counts will be displayed only if there are missing values. All mean and median values will be formatted to one more decimal place than the measured value. SD values will be formatted to two more decimal places than the measured value. The min and max will be displayed to the same number of decimal places as the measured value. Unscheduled visits will not be summarized in tables.

Analyses will be performed by visit, irrespective of any time window deviations, unless otherwise specified. The ED visit will be separated from the scheduled EoT/- visit based on discontinuation reason. Unscheduled visits related to efficacy data will be clustered with scheduled visits based on the visit windows in Table 7 of Appendix 1. Similarly, the unscheduled visit results will be clustered with scheduled visits based on the visit windows defined in Table 8 (Appendix 1) for microscopic findings and laboratory endpoints and endometrial biopsy. Similarly, the scheduled visits for ECG, Procedures, Questionnaire, Vital Signs, Imaging and Findings will be based on visit window Table 9 in Appendix 1.

All subject data, including unscheduled data, will be presented in individual subject data listings. Unless otherwise stated, unscheduled visit results will be included in date/time chronological order, within listings only. The subject's age will be stated on each listing. Unless specified, no imputed dates or values will be presented in listings.

7.2 Subject Disposition

Overall subject disposition table will clearly describe disposition of subjects from enrollment to allocation to one of the study parts (i.e. Efficacy and Safety). It will summarize:

- Subjects who are screened (Enrolled Set).
- Frequency counts (n) and percentages of subjects who are screen failures and reasons for screen failures based on Enrolled Set.
- Subjects in Safety Study part (Included set)

- Subjects in Efficacy Study part (Randomized Set)

Another subject disposition table will be presented for the Randomized Set. This will summarize data by treatment arm and overall. Data will be also summarized by hysterectomy status (hysterectomized and non-hysterectomized) including

- Subjects in Efficacy Study part (Randomized Set)
- Subjects treated
- Subjects who completed the treatment
- Subjects who discontinued treatment and reasons for discontinuation

The denominator for percentage of subjects treated will be the number of subjects allocated to the Efficacy Study part (Randomized Set). The denominator for percentage of subjects who completed or discontinued treatment will be the number of subjects treated in the Efficacy Study part (Safety Analysis Set).

A table summarizing the analysis sets will be presented by hysterectomy status (hysterectomized and non-hysterectomized) and by treatment arm and overall for Randomized Set including

- Subjects included in Randomized Set
- Subjects in ITT Set and reason for exclusion
- Subjects in mITT Set and reason for exclusion
- Subjects in Safety Analysis Set and reason for exclusion
- Subjects in PP Set and reason for exclusion

In addition, summary table of cumulative number of days from baseline for each visit will be provided.

A listing of subject disposition will be provided for the Randomized Set, with the extent of participation in the study and the reason for discontinuation. In addition, hysterectomy status (hysterectomized and non-hysterectomized) will be flagged in the listing.

A listing of screen failures and reasons for screen failures will be provided for the Enrolled Set.

A listing of subjects included in each analysis set and reasons for exclusion will be provided for the Randomized Set.

7.3 Protocol Deviations

A summary of major protocol deviations will be presented by type of deviation for the ITT Set.

All protocol deviations will be listed for the Randomized Set.

7.4 Demographics and Baseline Characteristics

7.4.1 Demographics

Demographic data will be summarized using the Safety Analysis and ITT Sets.

Continuous variables such as age, body weight, height, BMI (at screening) will be summarized using descriptive statistics. Categorical variables such as race, ethnic origin and education level will be summarized using frequency counts (n) and percentages.

Listing of demographics will be produced for the Randomized Set.

7.4.2 Baseline Characteristics

Subject baseline characteristics including hysterectomy status (hysterectomized and non-hysterectomized), smoking habits, gynecological history, PAP test, Breast Imaging-Reporting and Data System (BI-RADS), modified MENQOL questionnaire, Follicle Stimulating Hormone (FSH), E2 and Thyroid Stimulating Hormone (TSH), lipid/glucose parameters (for inclusion) and the number of moderate to severe VMS per week during the Screening period.

Subject baseline characteristics will be summarized using the Safety Analysis and ITT Sets. The summary will include descriptive statistics for continuous measures and frequency counts (n) and percentages for categorical measures.

Tables summarizing subject baseline characteristics will also be presented by hysterectomy status (hysterectomized and non-hysterectomized).

Listing of baseline characteristics will be produced for the Randomized Set.

7.4.3 Medical History

A summary of medical history will be presented by system organ class (SOC) and preferred term (PT) using Medical Dictionary for Regulatory Affairs® (MedDRA) Version 25.0 or higher for the Safety Analysis and ITT Sets.

A listing including medical history and current medical conditions will be presented along with the SOC, PT by subjects for the Randomized Set.

7.4.4 Prior and Concomitant Medications

Medications used in this study will be coded by using the latest available version of the World Health Organization (WHO) Drug Dictionary Enhanced (Enhanced 2021 September version or higher).

Prior medications: A medication that has a start date prior to the date of first dose of study drug.

Concomitant medications: A medication that starts on or after the date of first dose of study drug.

If a medication starts before the date of first dose of study drug and is ongoing (i.e., does not have a stop before the date of first dose of study drug) then the medication will be counted as both a prior and a concomitant medication.

Prior and concomitant medications will be summarized by anatomical therapeutic chemical (ATC) level 2 and preferred name (level 5) for the Safety Analysis and ITT Sets. A listing of medication data will be produced for the Randomized Set.

Partial medication start and stop dates will be imputed as detailed in Handling of **Missing or Incomplete** .

7.5 Extent of Exposure

Treatment exposure (days) and compliance (%) of E4 will be assessed from study drug dispensing records.

Descriptive statistics will be used to summarize duration of exposure and compliance of E4 for the Safety Analysis and ITT Sets. Also, compliance will be summarized categorically as number of subjects with compliance categories < 80%, between 80% and 120% and > 120%.

Tables summarizing subject exposure to E4 will also be presented by hysterectomy status (hysterectomized and non-hysterectomized) for both analysis sets.

A listing of study drug administration of E4 and P4, along with derived variables including duration of exposure, treatment compliance and compliance category will be produced for the Randomized Set.

7.6 Efficacy Analyses

This section addresses separately the analyses to be conducted on the primary and secondary efficacy endpoints.

7.6.1 Analysis Methods

7.6.1.1 Analysis of Mixed-effects Model for Repeated Measures

The analysis of the efficacy endpoints collected over time will use an MMRM analysis on change from baseline. The MMRM will include treatment (E4 15 mg, E4 20 mg, Placebo), week, week*treatment, status (hysterectomized and non-hysterectomized), status*treatment and pooled trial centers as fixed effects and baseline as a covariate. If the pooled center effect included in the model is significant, a treatment by center interaction will be included in the model. Treatment effects will be assessed at week 4 and 12 using two separate MMRM (one on weekly frequency and another on weekly mean severity).

For the MMRM, autoregressive structure will be used as covariance structure. As this data has a repeated structure over time points, so it can be assumed that two measurements that are right next to each other in time are going to be strongly correlated but that as measurements get farther and farther apart they are less correlated. Hence, autoregressive structure will be the ideal option in that case.

If there is enough proof that the model with autoregressive structure used as covariance structure does not provide a good fit, unstructured and other covariance structures will be explored.

The Least Squares (LS) means, LS mean difference between each active arm and the placebo group (reference category), standard error, two-sided 95% CIs for the LS means and difference between treatment groups and p-value for treatment differences will be presented. The p-values for the effects will also be presented. The number of subjects in the analysis sets and number of subjects in the analysis will be provided by treatment group.

The sample SAS code for MMRM is:

```
proc mixed data=xxxx;
```

```
class <treatment group> <week> <status> <subject id>;
model <change from baseline> = <baseline result> <treatment group> <week> <week*treatment> <status> <status*treatment> <pooled trial centers>;
repeated <week> / type=AR(1) subject=<subject id>(<status*treatment>);
lsmeans <treatment group>/ diff cl alpha=0.05;
ods output lsmean=xxx diffs=xxx;
run;
```

7.6.1.2 Analysis of Covariance Model for Change from Baseline

The analysis of the primary efficacy endpoint will use an analysis of covariance (ANCOVA) model on change from baseline with treatment (E4 15 mg, E4 20 mg, Placebo), status (hysterectomized and non-hysterectomized), status*treatment and pooled trial centers as fixed effects and the baseline measurement as a covariate.

The model will be used to derive LS estimates of the treatment differences (E4 15 mg versus Placebo, E4 20 mg versus Placebo) in mean change and two-sided 95% CIs. t-statistics corresponding to the type III sums of squares for the differences in the LS means will be used to obtain p-values for treatment group comparisons (each active treatment group versus placebo). Moreover, two-sided 95% CIs for the mean change within each treatment group will be calculated. The p-values for the effects will also be presented. The number of subjects in the analysis set and number of subjects in the analysis will be provided by treatment group.

The sample SAS code for ANCOVA is:

```
proc mixed data=xxxx;
class <treatment group> <status>;
model <change from baseline> = <baseline result> <treatment group> <status> <status*treatment> <pooled trial centers>;
by <week>;
lsmeans <treatment group>/ diff cl alpha=0.05;
ods output lsmean=xxx diffs=xxx;
run;
```

7.6.1.3 Multiplicity

To strictly control overall type I error rate for the analysis of the primary and the secondary endpoints to a two-sided 5% level, the pairwise comparisons of each active dose versus placebo (E4 15 mg vs Placebo and E4 20 mg vs Placebo) will be performed using Dunnett's test.

There is no adjustment of the type I error needed for the four assessed co-primary endpoints,

since it is only in case they are all significant that a given dose can be claimed as statistically superior to placebo.

7.6.1.4 Pooling of Centers for Analysis

If required, centers will be pooled by states depending on the distribution of randomized subjects.

Small centers with less than 10 subjects (based on the ITT set) will be pooled by state. If the number of ITT subjects per state again is too small to get meaningful results, consider further pooling with small sites from adjacent state(s). If the number of subjects is again small, center will be pooled with others.

The decision, whether and exactly how to pool centers or whether to use state instead in the statistical models will be made before unblinding. This decision will be applied on all statistical models planned with center as fixed effect.

7.6.2 Analysis of Primary Efficacy Endpoints

The primary efficacy comparison will test the following hypotheses for each treatment arm:

- H_0: The average change from baseline in the primary efficacy endpoint in the treatment arm is greater (less pronounced) or equal to the average change from baseline in the placebo arm
- H_1: The average change from baseline in the primary efficacy endpoint in the treatment arm is lower (more pronounced) than the average change from baseline in the placebo arm.

The two dose groups of E4 will be compared to placebo regarding the four co-primary endpoints. The null hypothesis will be rejected if there is statistical evidence of significantly lower average change from baseline in the primary efficacy endpoint in dose groups of E4 compared to placebo. This would be checked based on estimates of difference between treatment effects (difference to placebo) and corresponding 95 % CIs based on Dunnett's test.

7.6.2.1 Primary Analysis

The primary efficacy analyses will be performed on the ITT Set. No imputation for missing weekly data will be used for the descriptive statistics and listings. Missing data will not be imputed in the primary analysis of the primary efficacy endpoints.

The weekly frequency of moderate to severe VMS and change from baseline to week 4 and week 12 will be summarized by using descriptive statistics including number of subjects (n), mean, SD, median, min, max and 95% CIs.

Analysis visits will be derived as specified in Appendix 1.

The mean (weekly) severity score of moderate to severe VMS and change from baseline to week 4 and week 12 will be summarized by using descriptive statistics including number of subjects (n), mean, SD, median, min, max and 95% CIs.

In addition, an MMRM will be used to analyze change from baseline of weekly frequency of moderate to severe VMS on week 4 and week 12 and change from baseline of severity score

of moderate to severe VMS on week 4 and week 12.

Mean change in weekly frequency of moderate to severe VMS from baseline to week 4/ week 12

A comparison between treatment arms of the change from baseline to week 4/week 12 in weekly frequency of moderate to severe VMS will be made by using MMRM as described in Section 7.6.1.1. Treatment effects will be assessed at week 4 and 12.

LS Mean estimates and 95% CIs for the respective treatment effects (difference to placebo) will be presented at week 4 and 12 along with the corresponding p-value.

Line plot for the change from baseline of the weekly frequency of moderate to severe VMS on week 4 and week 12 with the result of the MMRM will be provided.

Mean change in mean (weekly) severity of moderate to severe VMS from baseline to week 4/ week 12

A comparison between treatment arms of the change from baseline to week 4/ week 12 in mean (weekly) severity of moderate to severe VMS will be made by using MMRM as described in Section 7.6.1.1. Treatment effects will be assessed at week 4 and 12.

LS Mean estimates and 95% CIs for the respective treatment effects (difference to placebo) will be presented at week 4/week 12 along with the corresponding p-value.

Line plot for the change from baseline of the weekly mean severity of moderate to severe VMS on week 4 and week 12 with the result of the MMRM will be provided.

A listing will be provided for VMS data for all randomized subjects.

7.6.2.2 Sensitivity Analyses of Primary Efficacy Endpoints

- Additional sensitivity analyses will be implemented in order to assess the robustness of the results. Analysis visits will be derived as specified in Appendix 1.
- The weekly frequency of moderate to severe VMS and change from baseline to week 4 and week 12 will be summarized by using descriptive statistics including number of subjects (n), mean, SD, median, min, max and 95% CIs based on mITT and PP Sets. Missing data will not be imputed.
- A comparison between treatment arms of the change from baseline to week 4/week 12 in weekly frequency of moderate to severe VMS will be made by using MMRM as described in Section 7.6.1.1. The analysis will be done on the mITT and PP Sets.
- A comparison between treatment arms of the change from baseline to week 4/week 12 in weekly frequency of moderate to severe VMS will be made by using MMRM as described in Section 7.6.1.1 on ITT, mITT and PP Sets. Line plot for the change from baseline of the weekly frequency of moderate to severe VMS on week 4 and week 12 with the result of the MMRM will be provided.
- The weekly frequency of moderate to severe VMS and change from baseline to week 4 and week 12 will be summarized by using descriptive statistics including number of subjects (n), mean, SD, median, min, max and 95% CIs based on ITT,

mITT and PP Sets. Missing data will be imputed using LOCF method as described in Section 6.2.2.

- A comparison between treatment arms of the change from baseline to week 4/week 12 in weekly frequency of moderate to severe VMS will be made by using an ANCOVA model as described in Section 7.6.1.2. This analysis will be done based on ITT, mITT and PP Sets. Missing data will be imputed using LOCF method as described in Section 6.2.2. Line plot for the change from baseline of the weekly frequency of moderate to severe VMS on week 4 and week 12 with the result of the ANCOVA will be provided.
- A comparison between treatment arms of the change from baseline at week 4/week 12 in weekly frequency of moderate to severe VMS will be made by using an ANCOVA model as described in Section 7.6.1.2. This analysis will be done based on ITT, mITT and PP Sets. Missing data will be imputed using LOCF method as described in Section 6.2.2. Line plot for the change from baseline of the weekly frequency of moderate to severe VMS on week 4 and week 12 with the result of the ANCOVA will be provided.

In addition, all the above-mentioned sensitivity analyses will be done to compare the change from baseline in mean severity score of moderate to severe VMS on week 4 and 12 between different treatment arms.

The subgroup analyses for primary analysis will be performed for the following groups and based on the ITT Set for change from baseline in the weekly frequency of moderate to severe VMS on week 4 and 12 and change from baseline in the mean severity score of moderate to severe VMS on week 4 and week 12. The subgroup analysis will only include subgroups that represents at least 5% of the ITT Set.

- Ethnicity (Hispanic or Latino/ Not Hispanic or Latino)
- Age (<55 years/ ⩾55 years)
- Race (American Indian or Alaska Native/ Asian/ Black or African American/ Native Hawaiian or Other Pacific Islander/ White/ Other)
- BMI (<25 kg/m^2/ 25 to <30 kg/m^2/ ⩾30 kg/m^2)
- Smoking (Yes/ No)

7.6.3 Analysis of Secondary Endpoints

7.6.3.1 Vasomotor Symptoms Related Secondary Efficacy Variables Analysis

For efficacy endpoints for secondary objective #1:

- Change from baseline to week 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 in the weekly frequency and severity of moderate to severe VMS.
- Change from baseline to weeks 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 in the weekly frequency and severity of mild to severe VMS.

The above endpoints will be analyzed separately for the below variables.

- Weekly frequency of moderate to severe VMS
- Weekly frequency of mild to severe VMS
- Mean weekly severity score of VMS by FDA method
- Mean weekly severity score of VMS by EMA method
- Mean weekly severity score of VMS by internal method
- Weekly weighted score by EMA method
- Weekly weighted score by FDA method
- Weekly weighted score by internal method

The above variables will be summarized separately using descriptive statistics based on ITT Set for each treatment group by visit. Analysis visits will be derived as specified in Appendix 1. The MMRM model as described in Section 7.6.1.1 will be used to analyze the above variables.

For the efficacy endpoints for secondary objective #1

- Percentage of subjects with 50% and 75% reduction from baseline in the weekly frequency of moderate to severe VMS at weeks 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12
- Percentage of subjects with 50% and 75% reduction from baseline in the weekly frequency of mild, moderate and severe VMS at weeks 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12

The proportion of subjects with 50% and 75% reduction from baseline in the weekly frequency of moderate to severe VMS at weeks 1 to 12 and the proportion of subjects with 50% and 75% reduction from baseline in the weekly frequency of mild to severe VMS at weeks 1 to 12 will be provided by number and percentage for each treatment group in ITT Set. Missing data will not be imputed. Treatment differences will be analyzed with the Chi-square test. The two-sided 95% CIs between (E4 15 mg versus Placebo, E4 20 mg versus Placebo) along with corresponding p-value will be calculated.

Line plot for the proportion of subjects with 50% and 75% reduction from baseline in the weekly frequency of moderate to severe VMS at weeks 1 to 12 will be provided.

7.6.3.2 Clinical Global Impression Questionnaire

Efficacy endpoints for secondary objectives #2 and #4 are based on the CGI questionnaire.

Trial participants will answer the question 'Rate the total improvement, whether or not in your judgement it is due entirely to drug treatment. Compared to your condition at administration to the study, how much has it changed using the following categorical scale?

- Very much improved
- Much improved
- Minimally improved
- No change

- Minimally worse
- Much worse
- Very much worse'

For the efficacy endpoints for secondary objective #2

- Percentage of subjects with a CID compared to baseline in the weekly frequency of moderate to severe VMS at weeks 4 and 12 using the CGI questionnaire. Analysis Visits will be derived as specified in Appendix 1.

To relate the clinical meaningfulness of treatment to changes in VMS severity that are clinically significant to subjects, the CGI responses/categories listed above are grouped into three subgroups: clinically meaningful (much or very much improved) corresponding to CIDs, minimally improved corresponding to MCIDs and no change or worse (no change to very much worse).

The percentage of subjects with CID, MCID and worsen/no change compared to baseline in the weekly frequency of moderate to severe VMS at weeks 4 and 12 will be provided by number (n) and percentage for each treatment group in ITT, mITT and PP Sets. Treatment differences will be analyzed with the Chi-square test. The two-sided 95% CI between (E4 15 mg versus Placebo, E4 20 mg versus Placebo) along with corresponding p-value will be calculated.

Vertical bar chart will be displayed for the proportion of subjects with CID, MCID and worsen/no change compared to baseline in the weekly frequency of moderate to severe VMS and mild to severe VMS at weeks 4 and 12.

For the efficacy endpoints for secondary objective #4

- Total score in TS after 4, 12 and 52 weeks of treatment using the CGI questionnaire

Total score improvement after 4, 12 and 52 weeks of treatment using the CGI questionnaire will be summarized using number (n) and percentage in each category on the ITT, mITT and PP Sets. Treatment differences will be analyzed with the Chi-square test. The two-sided 95% CIs between (E4 15 mg versus Placebo, E4 20 mg versus Placebo) along with corresponding p-value will be calculated on the ITT, mITT and PP Sets.

Mean change from baseline in the frequency of moderate to severe VMS will also be summarized within each CGI subgroup (CID, MCID and Worsen/No Change) and category at Weeks 4 and 12, on the ITT, mITT and PP Sets.

The frequency of moderate to severe VMS will also be analyzed at Week 4 and Week 12 by CGI subgroup and category (as defined for the respective week) using an ANCOVA on change from baseline as described in Section 7.6.1.2. Missing data will be imputed using LOCF method as described in Section 6.2.2. The subgroup/category analysis will only include subgroup/categories that represent at least 5% of the ITT, mITT or PP Sets, respectively. The model will be derived as described in Section 7.6.1.2.

Vertical bar chart will for the change from baseline on week 4 and 12 will be provided.

A listing will be provided for CGI questionnaire data for the Randomized Set.

7.6.3.3 Laboratory Assessments for Efficacy Analysis

Samples for lipid and glucose metabolism will be collected at Visit 1 (baseline), Visit 4 (week 12), and Visit 7 (EoT, week 52) in the Efficacy Study part.

- Lipid metabolism: triglycerides, HDL-cholesterol, LDL-cholesterol, total cholesterol, lipoprotein (a) and total cholesterol/HDL-cholesterol ratio.
- Glucose metabolism: fasting glycemia, insulin, glycated hemoglobin and HOMA-IR.

Samples for hemostasis and bone markers will be collected at Visit 1 (baseline), Visit 4 (week 12), and Visit 7 (EoT, week 52) for subjects in the Efficacy Study part:

- Hemostasis parameters: prothrombin fragment 1 + 2, ETP-based APCsr, aPTT based APCr, anti-thrombin III, Protein-C, free Protein-S, Factor VIII, angiotensinogen and SHBG.
- For subjects who gave consent for the DNA sampling only: Factor V Leiden mutation and prothrombin G20210A mutations.
- Bone markers: PINP, CTX-1, calcium and vitamin D.

The observed value and change from baseline to week 12 and week 52 for all parameters will be summarized by using descriptive statistics including number of subjects (n), mean, SD, median, min, max and 95% CIs for the ITT, mITT and PP Sets without imputation of data. In case of multiple measurement at the same visit, the last observed value will be summarized. This last observed value will include all last visits regardless of scheduled or unscheduled. Analysis Visits will be derived as specified in Appendix 1.

The MMRM as described in Section 7.6.1.1 will be used to analyze change from baseline for all parameters on week 12 and 52. The LS means using maximum likelihood estimate and its 95% CI will be provided for the ITT, mITT and PP Sets.

Line plot for the change from baseline on week 12 and week 52 with the result of the MMRM will be provided.

Listings will be provided for lipid and glucose metabolism, hemostasis and bone markers for the Randomized Set.

7.6.3.4 Menopause Specific Quality of Life Questionnaire

7.6.3.4.1 Scoring Rule of Menopause Specific Quality of Life

The MENQOL is a 29 item (Q1-Q29), assessment of quality of life designed to capture self-reported information on the presence and bother of symptoms, feelings and experiences in the domains of vasomotor, psychosocial, physical and sexual functioning, among midlife women

in the immediate post-menopause period. Table 3 provides the list of the relevant numbered items for each domain.

Table 3: Items numbers for each domain

Domains	Items numbers
Vasomotor	1 to 3
Psychosocial	4 to 10
Physical	11 to 26
Sexual functioning	27 to 29

At baseline, women will be asked to complete the modified MENQOL questionnaire, responding for symptoms over the previous 3 days. During the study, women will be asked to complete the MENQOL questionnaire, answering about symptoms over the previous 1-month period.

For each item, women will be asked to report if they experience that symptom or feeling and if they do, to rate how much it bothered them on a scale of 0–6 corresponding to 'not at all bothered' to 'extremely bothered'. Non-endorsement of an item will be scored a '1' and endorsement a '2' plus the number of the particular rating, so that the possible score on any item ranges from 1 (not experiencing symptoms or feeling) to 8 (extremely bothered), as described in Table 4.

Table 4: Scoring of each Items

Experienced that symptom or feeling	No	Yes						
Rating		0	1	2	3	4	5	6
Score	1	2	3	4	5	6	7	8

Each domain score is the mean of the non-missing item scores in that domain (higher scores indicated poorer quality of life), if at least 50% of the items within a domain function have been rated, as described in Section 6.2.3.

Overall MENQOL total score is the mean of the non-missing domain scores, if at least 50% of the domains have a score, as described in Section 6.2.3.

7.6.3.4.2 Analysis of Menopause Specific Quality of Life

The observed value and change from baseline to week 12 and week 52 for MENQOL total score and scores for each domain (vasomotor, psychosocial, physical and sexual functioning) will be summarized by using descriptive statistics including number of subjects (n), mean, SD, median, min, max and 95% CIs for the ITT, mITT and PP Sets. Analysis Visits will be derived as specified in Appendix 1.

The MMRM as described in Section 7.6.1.1 will be used to analyze change from baseline on week 12 and 52. The LS means using maximum likelihood estimate and its 95% CI will be

provided for the ITT, mITT and PP Sets.

Line plot for the change from baseline on week 12 and week 52 with the result of the MMRM will be provided.

A listing will be provided for MENQOL questionnaire data for Randomized Set.

7.6.4 Analysis of Exploratory Endpoint

The subgroup analyses will be performed by women with and without inherited thrombophilia based on the ITT, mITT and PP Sets for change from baseline to week 12 and week 52 of hemostasis parameters.

The observed value and change from baseline to each post-baseline time point for hemostasis parameters will be summarized by using descriptive statistics including number of subjects (n), mean, SD, median, min, max and 95% CIs without imputation of data. In case of multiple measurement at the same visit, the last observed value will be summarized. This last observed value will include all last visits regardless of scheduled or unscheduled. Analysis Visits will be derived as specified in Appendix 1.

The MMRM as described in Section 7.6.1.1 will be used to analyze change from baseline on week 12 and 52.

Line plot for the change from baseline on week 12 and 52 with the result of the MMRM will be provided.

In addition, change in breast density from digitized mammography readings from baseline to EoT in subjects who had a paired digitized mammography will be summarized by using descriptive statistics including number of subjects (n), mean, SD, median, min, max and 95% CIs.

A comparison between treatment arms of the change from baseline to EoT will be made by using an ANCOVA model as described in Section 7.6.1.2.

Line plot for the change from baseline with the result of the ANCOVA will be provided.

7.7 Safety Analyses

The safety evaluations will include analyses of AEs, physical and gynecological examination, breast examination, TVUS, vaginal bleeding, vital signs, ECG, routine clinical laboratory tests and mammography. For all safety endpoints, analysis visits will be derived as specified in Table 8 and Table 9 of Appendix 1.

The safety analyses involving changes from baseline to a specific time point in safety variables (e.g., laboratory parameters, vital signs, and ECG) will only include subjects from the Safety Analysis Set who have data available for both the baseline and the time point under consideration unless otherwise specified.

All safety evaluations will be summarized using data from the Safety Analysis Set if not indicated otherwise. No formal statistical test will be performed. Continuous variables will be summarized using descriptive statistics (n, mean, SD, median, min and max) and categorical variables will be presented in frequency tables using counts (n) and percentages.

7.7.1 Adverse Events

All AE verbatim descriptions will be coded using the MedDRA, version 25.0 or higher.

All AEs will be listed. TEAEs will be summarized as definition of TEAE are described in Section 6.1. Details for imputing missing or partial start dates of AEs are described in Section 6.2.4.

In summaries by SOC and PT, AEs will be sorted by decreasing frequency within each SOC and PT as per the Overall column. In summaries by PT, AEs will be sorted by decreasing frequency as per the Overall column.

If a subject experiences the same TEAE more than once with different intensities, then the event with the highest intensity will be tabulated in 'by intensity' tables. TEAEs with a missing intensity will be presented in the summary table as an intensity category of 'Missing'.

If a subject experiences multiple TEAEs under the same PT (SOC), then the subject will be counted only once for that PT (SOC).

Relationship to study drug, as indicated by the Investigator, is classified as 'Highly Probable', 'Probable', 'Possible', 'Unlikely' and 'Not Related'. If a subject experiences multiple TEAE within the same PT (SOC), the TEAE with the strongest relationship to study drug will be counted for that PT (SOC). Relationship will be summarized as 'Related' and 'Not Related'. AEs with investigator indicated drug-event relationship of 'Unlikely' and 'Not Related' will be classified as 'Not Related' to study drug, while AEs with drug-event relationship documented as 'Highly Probable', 'Probable' and 'Possible' will be considered as 'Related' to study drug. TEAEs having a missing relationship to study drug will also be considered as 'Related' to study drug.

Adverse Events of Special Interest (AESI) are a subset of AEs that include the events listed below.

- Vaginal bleeding events grade 2 as assessed using the vaginal bleeding events scale
- Events resulting from endometrial biopsy reading limited to:
  - Disordered proliferative endometrium;
  - Simple hyperplasia without atypia;
  - Complex hyperplasia without atypia;
  - Simple hyperplasia with atypia;
  - Complex hyperplasia with atypia;
  - Carcinoma

Collecting and recording AESIs will begin on Treatment Start Date and will end at the follow-up visit. AESI collection will start at each site from the date of the Ministry of Health and/or Institutional Review Board (IRB)/ Ethics Committee (EC) approval date of Protocol Version 6.0 (Amendment 5). There is no need for retrospective reporting.

An overview of TEAEs will be presented for the Safety Analysis Set with the frequency counts (n) and percentages of subjects with at least one:

- TEAE
- Severe TEAE
- Drug related TEAE
- Serious TEAE
- Drug related Serious TEAE
- TEAE leading to dose interruption
- TEAE leading to study drug discontinuation
- TEAE leading to death
- Drug related TEAE leading to death.
- AESI

The number of TEAEs in each category will be also included in this overview table.

The overview table will also be presented by hysterectomy status (hysterectomized and non-hysterectomized).

All TEAEs (including AESI and SAE), SAE (including serious AESI) and AESI (including both serious and non-serious AESI) will also be summarized for the Safety Analysis Set as below:

- TEAEs by SOC, PT
- TEAEs of reproductive system and abdominal pain by SOC, PT by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 5
- TEAEs by SOC, PT and maximum intensity
- TEAEs of reproductive system and abdominal pain by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 5
- Drug related TEAEs by SOC, PT and maximum intensity
- Drug related TEAEs of reproductive system and abdominal pain by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 5
- TEAEs leading to study drug interruption by SOC, PT and maximum intensity
- TEAEs of reproductive system and abdominal pain leading to study drug interruption by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 5
- TEAEs leading to study drug discontinuation by SOC, PT and maximum intensity
- TEAEs of reproductive system and abdominal pain leading to study drug discontinuation by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 5

- Serious TEAEs by SOC and PT
- Serious TEAEs of reproductive system and abdominal pain by SOC and PT by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 5
- Serious TEAEs by SOC, PT and maximum intensity
- Serious TEAEs of reproductive system and abdominal pain by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 5
- Drug related serious TEAEs by SOC, PT and maximum intensity
- Drug related serious TEAEs of reproductive system and abdominal pain by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 5
- TEAEs leading to deaths by SOC, PT and maximum intensity
- TEAEs of reproductive system and abdominal pain leading to deaths by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 5
- Drug related TEAEs leading to deaths by SOC, PT and maximum intensity
- Drug related TEAEs of reproductive system and abdominal pain leading to deaths by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 5
- AESI by SOC, PT and hysterectomy status (hysterectomized and non-hysterectomized).
- AESI by SOC, PT, maximum intensity and hysterectomy status (hysterectomized and non-hysterectomized).
- Venous thromboembolism events (VTE) incidence and cardiovascular events incidence, with additional information as defined below.
- Number of subjects with at least one event in each of the following categories:
  - VTE including:
    - pulmonary embolism and pulmonary embolus.
    - deep venous thrombosis.
  - Cardiovascular events including:
    - myocardial infarction, acute myocardial infarction and myocardial ischaemia.
    - cerebrovascular accident, transient ischaemic attack, ischaemic stroke and subarachnoid haemorrhage.
    - peripheral occlusive disease.
    - angina unstable and angina pectoris terms
  - Other thrombotic events

Number of VTE and cardiovascular events, n per 1000 patient years (PY) of exposure will be presented, which is defined as the summation of number of years for all the subjects at risk, calculated as:

$$Events\ per\ 1000\ PY = n/\ PY*1000$$

where,

n = number of treated subjects in each group.

PY = (date of last dose – date of first dose + 1)/365.25.

Table 5: Reproductive System and Abdominal Pain Referred Terms

	AEBODSYS	AEDECOD
Reproductive system	Reproductive system and breast disorders	
Abdominal pain	Gastrointestinal disorders	Abdominal pain Abdominal pain lower Abdominal pain lower

Listings to be provided for the Randomized Set are (including flags for TEAEs):

- AEs
- AEs leading to deaths
- All SAEs (including serious AESI)
- AEs leading to dose interruption and discontinuation.
- All AESI (including serious and non-serious AESI)
- VTE in the cardiovascular and cerebrovascular system

7.7.2 Clinical Laboratory Evaluations

Routine clinical laboratory tests (hematology and chemistry) include:

- Hematology: hematocrit, hemoglobin, erythrocytes, platelets, leukocytes, neutrophils, lymphocytes, basophils, eosinophils and monocytes
- Chemistry: urea, creatinine, total bilirubin, alkaline phosphatase, alanine transaminase, aspartate transaminase, sodium, potassium, chloride, bicarbonate, calcium, albumin, total protein, lactate dehydrogenase (LDH), LDH isoenzymes (I, II, III, IV and V), troponin and blood glucose

All laboratory data will be summarized in International System of Units (SI). Quantitative laboratory measurements reported as '< X', i.e., BLQ or '> X', i.e., above the upper limit of quantification (ULQ), will be converted to X for the purpose of quantitative summaries, but will be presented as recorded, i.e., as '< X' or '> X' in the listings. The summaries will include all laboratory assessments collected no later than 30 days after the last drug/exposure date.

Descriptive statistics of observed value and change from baseline (n, mean, SD, median, min

and max) for continuous assessments will be provided by visits and treatment group for the Safety Analysis Set. In case of multiple measurement at the same visit, the last observed value will be summarized. This last observed value will include all last visits regardless of scheduled or unscheduled.

Values outside the normal range will be categorized as high (above the normal range) or low (below the normal range) based on the reference range of the safety laboratories.

Shift tables will be used to summarize the change in frequency counts (n) and percentages of subjects with normal, high or low from baseline to post-dose scheduled visits for the Safety Analysis Set.

Listings will be provided for hematology and chemistry laboratory data for the Randomized Set.

A listing of urinary pregnancy tests performed will be provided for the Randomized Set.

7.7.3 Vital Signs

Vital signs will include height, body weight, BMI, sitting systolic and diastolic blood pressures (BPs).

Descriptive statistics (n, mean, SD, median, min and max) for each vital sign parameter willbe used to summarize the observed results and the changes from baseline value by scheduled visit and treatment group for the Safety Analysis Set.

A listing will be provided for vital signs data for the Randomized Set.

7.7.4 Physical Examinations

Physical examination results will be summarized by frequency counts (n) and percentages for each body system at baseline and post-dose scheduled visits for the Safety Analysis Set.

A shift table will be produced to summarize change in the frequency counts (n) and percentages of subjects with normal, abnormal not clinically significant and abnormal clinically significant from baseline to post-dose scheduled visits for the Safety Analysis Set.

A listing will be provided for physical examination data for the Randomized Set.

7.7.5 Electrocardiograms

Only ECG status (normal, abnormal not clinically significant and abnormal clinically significant) will be reported.

A shift table will be produced to summarize the change in the frequency counts (n) and percentages of subjects with normal, abnormal not clinically significant and abnormal clinically significant from baseline to post-dose scheduled visits for the Safety Analysis Set.

A listing will be provided for ECG data for the Randomized Set.

7.7.6 Other Safety Assessments

7.7.6.1 Gynecological Examinations

Gynecological examination results will be summarized by frequency counts (n) and

percentages for each body system at baseline and post-dose scheduled visits for the Safety Analysis Set.

A shift table will be produced to summarize change in frequency counts (n) and percentages of subjects with normal, abnormal not clinically significant and abnormal clinically significant from baseline to post-dose scheduled visits for the Safety Analysis Set.

A listing will be provided for gynecological examination data for the Randomized Set.

7.7.6.2 Breast Examinations

Breast examination results will be summarized by frequency counts (n) and percentages for each body system at baseline and post-dose scheduled visits for the Safety Analysis Set.

A shift table will be produced to summarize change in frequency counts (n) and percentages of subjects with normal, abnormal not clinically significant and abnormal clinically significant from baseline to post-dose scheduled visits for the Safety Analysis Set.

A listing will be provided for breast examination data for the Randomized Set.

7.7.6.3 Mammography

Subjects must have a BI-RADS score of 1 or 2 to enroll in the study, as assess by a digitalized mammogram performed during the Screening period. Baseline BI-RADS categories (1-6) will be summarized along with the baseline characteristic.

Breast density will be assessed by a digitalized mammogram at baseline and EoT. Descriptive statistics (n, mean, SD, median, min and max) will be used to summarize the breast density of left and right breasts at baseline and EoT, and change from baseline, by treatment group for the Safety Analysis Set. Breast density will be summarized categorically in the same table using frequency counts (n) and percentages for subjects in categories A-D (A = 0-3.5%, B = >3.5 – 7.5%, C = >7.5 – 15.5%, D = >15.5%) at baseline and EoT for the Safety Analysis Set.

The change from baseline in breast density, will be analyzed separately for left and right breasts, in subjects who had a paired digitized mammography using an ANCOVA as described in Section 7.6.1.2.

A listing will be provided for mammography data for the Randomized Set.

7.7.6.4 Transvaginal Ultrasound

For subjects with uterus and/or ovaries for whom TVUS have been performed, interpretation of TVUS image results will be summarized by frequency counts (n) and percentages on baseline and post-dose scheduled visits (in NH subjects only) for the Safety Analysis Set.

A shift table will be produced to summarize change in frequency counts (n) and percentages of subjects with normal, abnormal not clinically significant and abnormal clinically significant TVUS image interpretation from baseline to post-dose scheduled visits (in NH subjects only) for the Safety Analysis Set.

A listing will be provided for TVUS data for the Randomized Set.

7.7.6.5 Endometrial Thickness Measured by Ultrasound

TVUS will be performed to monitor the endometrium thickness during the treatment at Visit 4, 5, 6, 7 and 8 for NH subjects.

Descriptive statistics of observed value and change from baseline (n, mean, SD, median, min, max, 95% CI) for endometrial thickness measured by ultrasound will be provided by visits and treatment group for the Safety Analysis Set.

A line plot for the change from baseline over time for endometrial thickness measured by ultrasound will be provided.

A listing will be provided for endometrial thickness by ultrasound for the Randomized Set.

7.7.6.6 Endometrial Biopsy

The frequency of endometrial events will be computed on the evaluable biopsies of the endometrium (Initial and Final diagnosis) or adequate specimens (Safety) as well as non-evaluable and inadequate (respectively) biopsies. In addition, counts (n) and percentages will be provided for the sub-categories of the evaluable/adequate biopsies. The incidence of spotting/bleeding will be analyzed by a 28-day period. The rate of endometrial events will be computed on the evaluable endometrium. The 95% two-sided CIs will be computed on the observed frequency of hyperplasia/carcinoma of the endometrium.

7.7.6.6.1 Evaluation of Endometrial Tissue

Standardized criteria as provided in Blaustein's pathology text (Pathology of the Female Genital Tract 8, see Appendix 5) will be used for the characterization of the endometrial tissue. Endometrial polyps will be fully characterized as to the glandular proliferation and atypia (see Appendix 5 for additional histologic characteristics of the specimen).

The endometrial tissue obtained by endometrial biopsy at screening, during the conduct of the study and at the end-of-study will all be processed in the same manner by a central laboratory.

Screening Biopsies/Initial Diagnosis:

Biopsies obtained at screening will be initially read by one safety pathologist (initial pathology report). The Investigator will decide on inclusion of the subject into the study or screen failure based on the initial pathology report at screening.

All inadequate tissue records will be disregarded in the derivation of the Initial diagnosis; the diagnosis will be based on only adequate tissue samples read by the one safety pathologist.

Screening biopsies will also be read by two other expert pathologists after enrollment. If a subject is enrolled incorrectly (e.g. based on an inadequate sample), she will be discontinued.

On Study Biopsies:

Biopsy samples collected during the study and at EoT/ED (Visit 6) will be read by three expert pathologists for the Safety and Final/Consensus diagnosis. These three expert pathologists will be independent and belong to different institutions. They will be blinded to each other's readings. The Safety diagnosis will be based on readings from pathologists 1, 2 and 4, whereas the Final/Consensus diagnosis will be based on readings from pathologists 2, 4 and 5. During the study, the investigator will be informed if subsequent readings impact the initial decision.

Note that endometrial biopsy will be performed at EoT visit if the subject was treated for at least 12 weeks.

Using Blaustein's classification, Benign Endometrium has been defined as any of the following readings: no tissue, insufficient tissue, atrophic endometrium, inactive endometrium, proliferative endometrium (incl. weakly proliferative, active proliferative, disordered proliferative), secretory endometrium (incl. cyclic type, progestational type) and menstrual type endometrium. For the detail of other category, please find Category Indicators for Final Diagnosis in Appendix 5.

Safety Diagnosis:

The investigator will receive final pathology report (or Safety diagnosis) defined as the worst reading of adequate specimens from the 3 pathologists involved in safety reading, where disordered proliferative is the best and carcinoma the worst. In case none of the adequate specimen readings are disordered proliferative or worse, the first reading with adequate specimen will be kept as Safety diagnosis. If the 3 readers concur on 'no tissue or tissue insufficient' then no Safety diagnosis is derived.

All inadequate tissue records will be disregarded in the derivation of the Safety diagnosis; the diagnosis will be based on all available adequate tissue samples, from any number of readers.

A reading of disordered proliferative endometrium, hyperplasia or carcinoma from any of these pathologists will prompt the exclusion or discontinuation of the subject from the study.

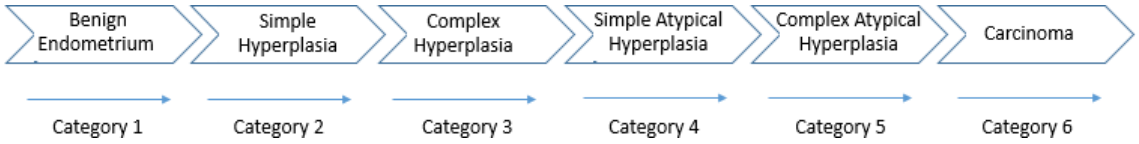
Final/Consensus Diagnosis:

The Final/Consensus diagnosis is the concurrence of at least 2 adequate specimens from the 3 pathologists. If there is no agreement among at least 2 pathologists, the most severe pathologic diagnosis will be used as the Final/Consensus diagnosis. If the 2 or 3 readers concur on 'no tissue or tissue insufficient' then no Final/Consensus diagnosis will be derived.

The Final/Consensus diagnosis will only be derived for evaluable biopsies. All inadequate tissue records will be disregarded; evaluable biopsies are defined as at least 2 readers assessing adequate tissue samples.

The Final/Consensus diagnosis will be used for the analysis of the endometrial safety endpoint. The categorization for Final/Consensus diagnosis is defined on Figure 2.

Figure 2 : Category indicators for Final diagnosis (from less to most severe)



Discontinuation:

If a biopsy shows disordered proliferative endometrium, hyperplasia or carcinoma, the study drug will be stopped, the treatment with 200 mg P4 for 14 days will be given and the subject will return to the site for the last visit. If the event is not resolved, the subject will be followed up until resolution according to local practice/guidelines.

In case of bleeding events grade 2, disordered proliferative endometrium or worse, the event must be reported as an AESI and the Sponsor will be notified.

7.7.6.6.2 Analysis of Endometrial Tissue

All the analyses of endometrial tissues will be performed only for NH subjects based on:

- Initial diagnosis (from first safety reader, based on adequate samples only)
  - By visit analysis: The Initial diagnosis will be summarized at each visit.
- Safety diagnosis
  - By visit analysis: The Safety diagnosis will be summarized at each visit.
- Final /Consensus diagnosis
  - By visit analysis: The Final/Consensus diagnosis will be summarized at each visit performed on-treatment.
  - Worst-case analysis across post-baseline visits: The worst case derived from post-baseline visits, including unscheduled visits, will be summarized for Final/Consensus diagnosis.

Visits will be clustered as specified in Table 8 of Appendix 1.

Frequency counts (n), percentages and the 95% CIs of the histologic characteristics of the endometrium including no tissue, tissue insufficient for diagnosis, atrophic, inactive, proliferative - weakly proliferative, proliferative - active proliferative, proliferative - disordered proliferative, secretory - cyclic type, secretory - progestational type (including stromal decidualization), menstrual type, simple hyperplasia without atypia, simple hyperplasia with atypia, complex hyperplasia without atypia, complex hyperplasia with atypia, carcinoma for Initial and Safety diagnosis will be summarized separately by visits and treatment group for the Safety Analysis Set.

Frequency counts (n), percentages and the 95% CIs of the additional pathological results

including polyps, stromal tissue, metaplasia and cervical tissue for Safety diagnosis will be summarized separately by visits and treatment group for the Safety Analysis Set.

Frequency counts (n), percentages and the 95% CIs of benign endometrium (including subcategory 1a with disordered proliferative as consensus diagnosis), simple hyperplasia without atypia, complex hyperplasia without atypia, simple hyperplasia with atypia, complex hyperplasia with atypia (including total for all hyperplasia), carcinoma will be summarized separately by visits and treatment group for the Safety Analysis Set for Final/ Consensus diagnosis. Worst-case analysis across post-baseline visits will also be summarized on the Safety Analysis Set for Final/Consensus diagnosis.

A listing will be provided for all subjects with all biopsy results from each pathologist, and the Initial, Safety and Final/Consensus diagnosis for the Randomized Set.

For the exclusionary subjects due to disordered proliferative endometrium and worse, a listing of subjects with Abnormal Endometrial Biopsy will be provided with all biopsy results per pathologist, initial pathology report, safety diagnosis and final diagnosis for the Randomized Set.

To analyze inter-reader variability Kappa statistics (coefficient of agreement between the reviewing pathologists) will be calculated. The simple (Cohen) kappa coefficient measure of interrater agreement will be used to estimate agreement among all the pair-wise comparisons of readers. For the comparison of consensus diagnosis (between readers 2, 4 and 5), safety diagnosis (between readers 1, 2, 4) and for the overall comparison (between all readers), the subject level comparison/repeated nature of the biopsies is not of interest, therefore the SAS MAGREE macro will be used to calculate Fleiss's kappa statistics for nominal responses.

Quality control

The slide set distributed to each of the three expert pathologists incorporates control slides representing a randomly selected sample of screening slides; 10% of normal subjects and subjects excluded for the diagnosis of hyperplasia or cancer to ensure quality control. The analysis of results of quality control is outside of the scope of this SAP.

7.7.6.7 Vaginal Bleeding and/or Spotting

Absence or occurrence of vaginal bleeding/ spotting on a daily basis will be assessed using the scale below.

0 = Absence of vaginal bleeding or spotting.

1 = Spotting: evidence of minimal blood loss requiring none or at most one pad, tampon or panty liner per day.

2 = Bleeding: evidence of blood loss requiring more than one pad, tampon or panty liner per day.

Frequency counts (n) and percentages of subjects with absence of any bleeding and/or spotting during the cycle, with vaginal bleeding and/or spotting during the cycle, with bleeding only

during the cycle, with spotting only during the cycle will be provided for each 28- day cycle of treatment based on data in the subject diary (in NH subjects) for the Safety Analysis Set.

The number of days with bleeding/spotting, as reported on subject diaries, will be summarized by 28-day cycle of treatment and by treatment group.

A listing will be provided for vaginal bleeding and/or spotting/amenorrhea for the Randomized Set.

The following data will be excluded from analysis of endpoints for bleeding (to be analysed separately).

- bleeding within 6 days after biopsy and
- bleeding reported after discontinuation of study drug (for example withdrawal bleeding, on P4 treatment)

7.7.6.8 Amenorrhea

Subjects with amenorrhea (absence of any bleeding or spotting) during each 28-day cycle of treatment with E4 (after exclusion of the bleeding or spotting within 6 days after biopsy or after discontinuation of study drug) will be summarized by using frequency counts (n) and percentages based on subject diary (in NH subjects) for the Safety Analysis Set.

Cumulative rates of amenorrhea will be defined as the percentage of women who reported consecutive cycles of amenorrhea for a given cycle of time. For example, if a subject had no bleeding or spotting from Day 1 to Day 364, then this subject has cumulative amenorrhea from the 1st to 13th cycle. The number and percentage of subjects with amenorrhea for each cumulative period will be summarized separately for the 1st to 13th cycle, 2nd cycle to 13th cycle, …., and the 13th cycle.

Cumulative 'no bleeding' will be defined as the percentage of women who reported consecutive cycles without bleeding for a given cycle of time. For example, if a subject had no bleeding from Day 1 to Day 364, then this subject has cumulative no bleeding from the 1st to 13th cycle. The number and percentage of subjects with 'no bleeding' for each cumulative period will be summarized separately for the 1st to 13th cycle, 2nd cycle to 13th cycle, …., and the 13th cycle.

7.8 Subgroup Analysis

Subgroup analysis will be performed based on the following and detailed in the respective analysis sections:

- Hysterectomy status (hysterectomized and non-hysterectomized) for all analyses
- Subgroups sensitivity VMS analysis (Section 7.6.2.2)
- According to inherited thrombophilia for hemostasis parameters (Section 7.6.4)

7.9 Interim Analysis

The final analysis of the primary efficacy objective and the secondary efficacy objectives 1 and 2 of the Efficacy Study part is to be triggered when all subjects in the Efficacy Study part have completed Visit 4 and will be conducted by an independent statistician. A firewall group composed of Estetra SRL representatives independent of the E4 project will be appointed. This group will have access to the unblinded data of final and interim analyses of the Efficacy Study part. The clinical team, the subjects and the site and sponsor personnel involved in the clinical evaluation of the subjects will remain blinded of the treatment assignment in the Efficacy Study part until the end of the study. Exact firewall process will be described in an independent unblinding plan.

Only the overall outcome of the results will be provided to the clinical team. Blinding for the data at subject and treatment arm levels will be kept through a firewall process that will be put in place. No individual listings with the subject numbers' information or tables per treatment arm will be disseminated at the time of final or interim triggered analysis of the Efficacy Study part.

8 CHANGES TO PLANNED ANALYSIS FROM STUDY PROTOCOL

- For the primary efficacy analysis, the MMRM was updated to include week, week*treatment, status, status* treatment in the model, where status is the hysterectomized status.
- Week and week-by-treatment are necessary for the MMRM approach but were an editorial issue in the protocol.
- As a stratification factor in the study design, the hysterectomized status (and its possible impact on the treatment effect) should preferably be considered in the primary analysis in line with the ICH E9 recommendation.
- The primary MMRM will include treatment, week, week*treatment, status, status*treatment and pooled trial centers as fixed effects and baseline as a covariate. The reduced MMRM without center and status effects and interactions will be applied as a sensitivity analysis.
- The Randomized Set which was not included in the protocol is added in Section 5.2 as 'The Randomized Set will include all subjects randomly allocated to one of the three treatment arms in Efficacy Study part (through IWRS system).'
- In Section 6.1, the definitions of TEAE, MCID, CID and TS were added.
- In Section 7.7.6.8, the definitions of 'Cumulative rates of amenorrhea' and 'Cumulative no bleeding' were added.
- In Sections 7.7.1, 7.7.6.5 and 7.7.6.6.2, analyses by NH subgroups were added.
- In Section 5, The Endometrial Safety Analysis Set was removed from the analysis sets because the endometrial analysis can be performed based on the Safety Analysis Set.
- In Section 7.6.3.2, the statistical ANCOVA model for the analysis of the frequency VMS data according to the CGI category was clarified.7.6.3.1a.7.6.3.2

9 REFERENCES

1. ICH Topic E3: Structure and Content of Clinical Study Reports (CPMP/ICH/137/95-adopted December 1995).
2. ICH Topic E9: Statistical Principles for Clinical Trials (CPMP/ICH/363/96 – adopted March 1998).
3. Gerlinger C, Gude K, Hiemeyer F, Schmelter T, Schäfers M. An empirically validated responder definition for the reduction of moderate to severe hot flushes in postmenopausal women. Menopause. 2012 Jul;19(7):799-803.
4. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-415

10 APPENDICES

Appendix 1 Visit Schedule and Permitted Time Window

Table 6: Visit Window for Efficacy Analysis of Efficacy Study part

Visit number	Week number	Time window
Washout visit	Week -12 to week -4[1]	n.a.
Screening period	Week -4 to week -1[1]	n.a.
Visit 1	Week 1	n.a.
Visit 2	Week 5	Day 29 - 35
Visit 3	Week 9	Day 57 - 63
Visit 4	Week 13	Day 85 - 91
Visit 5	Week 29	Day 197 - 203
Visit 6	Week 41	Day 281 - 287
Visit 7	Week 53	Day 365 - 371
Visit 8 NH subjects only NH subjects only	Week 55/56	Day 379 - 392

n.a.: not applicable

[1] The week numbers for the Washout visit and Screening period may deviate. The washout visit is generally up to 8 weeks prior to the screening visit. The Screening period is up to 4 weeks, but may be extended to a maximum of 8 weeks

Table 7: Analysis Visit Window for VMS Related Endpoints

Time window (in ADY)	Analysis visit (AVISIT)
-7 to -1	Baseline (if last non- missing value, including results from repeated and unscheduled measurements, recorded prior to the first dose of trial drug)
1 to 7	Week 1
8 to 14	Week 2
15 to 21	Week 3
22 to 28	Week 4
29 to 35	Week 5
36 to 42	Week 6
43 to 49	Week 7
50 to 56	Week 8
57 to 63	Week 9
64 to 70	Week 10
71 to 77	Week 11

78 to 84	Week 12

Table 8: Visit Window Microscopic Findings and for Laboratory Endpoints

Assessment Visit	Time window (in ADY)	Analysis visit (AVISIT)
Efficacy Visit 2		Week 5
Efficacy Visit 3		Week 9
Efficacy Visit 4		Week 13
Efficacy Visit 5		Week 29
Efficacy Visit 6		Week 41
Efficacy Visit 7/EOT		Week 53 (if discontinuation reason is missing in DS domain)
		Early Discontinuation (if discontinuation reason is non-missing in DS domain)
Efficacy Visit 8 Follow up		Week 55/56 Follow up
Unscheduled Visits	<0, 2 to 18	Unscheduled
	1	Visit 1 Baseline
	19 to 46	Week 5
	47 to 74	Week 9
	75 to 144	Week 13
	145 to 242	Week 29
	243 to 326	Week 41
	327 to 375*	Week 53
	>=327#	Week 53
	>=376*	Week 55/56 Follow up

Note: *only for non-hysterectomized subjects; #: only for hysterectomized subjects

Table 9: Visit Window ECG, Procedures, Questionnaire, Vital Signs, Imaging and Findings

Assessment Visit	Time window (in ADY)	Analysis visit (AVISIT)
Efficacy Visit 2		Week 5
Efficacy Visit 3		Week 9
Efficacy Visit 4		Week 13
Efficacy Visit 5		Week 29

Efficacy Visit 6		Week 41
Efficacy Visit 7/EOT		Week 53 (if discontinuation reason is missing in DS domain)
		Early Discontinuation (if discontinuation reason is non-missing in DS domain)
Efficacy Visit 8 Follow up		Week 55/56 Follow up

Appendix 2 Schedule of Trial Procedure

	Washout period [a]	Screening period	Baseline (random allocation to treatment)	Treatment visits					EoT /early discontinuation visit [b]	Follow up Visit NH subjects
	Washout visit	Screening visit	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8
Procedures and assessments	Generally within 8 weeks before screening	Within 4 weeks before baseline [m]	Day 1	Week 5 [c]	Week 9 [c]	Week 13 [c]	Week 29 [c]	Week 41 [c]	Week 53 [c]	Week 55/56 [d]
Informed consent	X									
I/E criteria	X [c]	X [c]	X							
Start washout [a]	X									
Med. & gyn. history	X									
Demographic data	X									
Physical examination		X							X	
Gyn. examination		X							X	
Breast examination		X					X		X	
Vital signs [f]		X	X	X	X	X	X	X	X	
ECG		X							X	
PAP test [g]		X								
Endometrial biopsy [h]		X							X [i]	
TVUS [h, j]		X				X	X	X	X	X
Mammography		X [k]							X [l]	
Modified MENQOL questionnaire		X								
MENQOL questionnaire			X			X			X	
CGI questionnaire				X		X			X	
Fasted blood sampling for: [m]										
Hematol./chem.		X [v]	X						X	
Lipid/glucose parameters for inclusion	X [u, v]	(X) [v]								
Lipid metabolism			X			X			X	
Glucose metabolism			X			X			X	

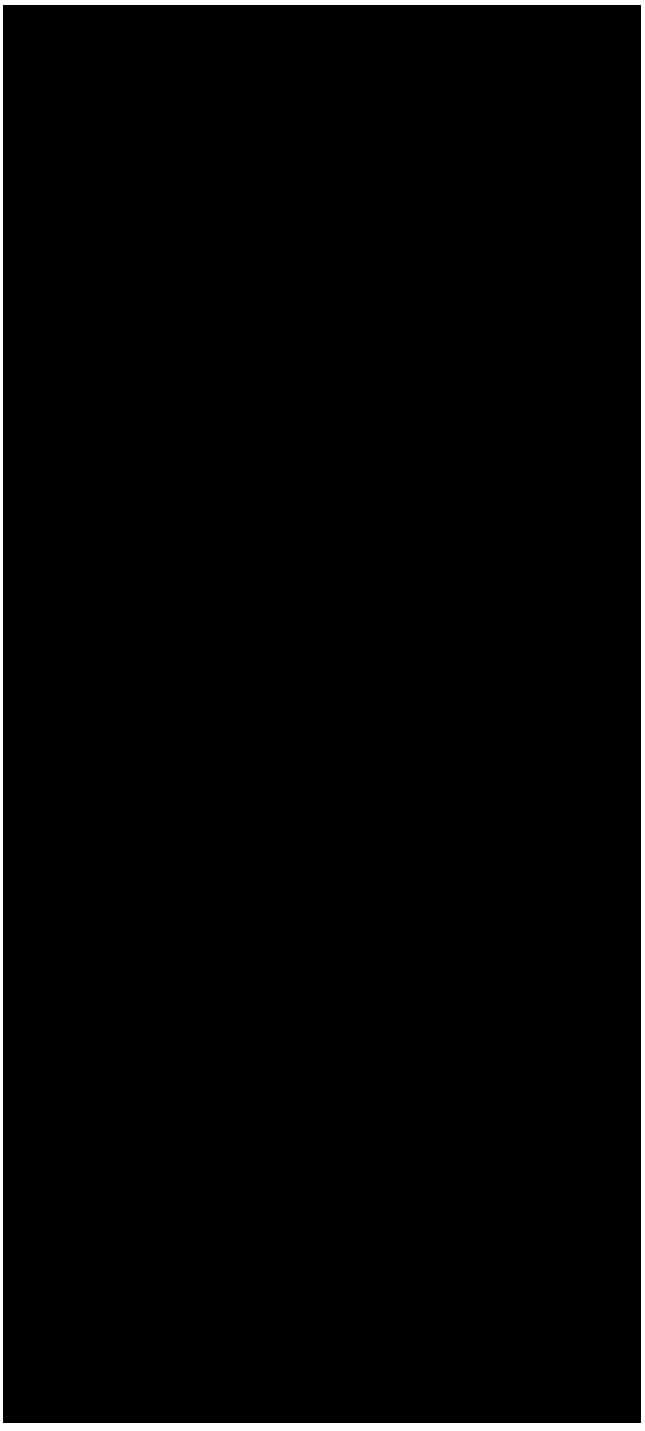
	Washout period [a]	Screening period	Baseline (random allocation to treatment)	Treatment visits					EoT/early discontinuation visit [b]	Follow up visit NH subjects
	Washout visit	Screening visit	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8
	Generally within 8 weeks before screening	Within 4 weeks before baseline [w]	Day 1	Week 5 [c]	Week 9 [c]	Week 13 [c]	Week 29 [c]	Week 41 [c]	Week 53 [c]	Week 55/56 [d]
Fasting blood sampling for: [m]										
Hemostasis			X			X			X	
Bone turnover			X			X			X	
Optional additional analysis [n]			X						X	
Blood sampling for										
FSH, E2 and TSH for inclusion [v]		X								
E2			X	X	X	X				
E4				X	X	X				
Urinary pregnancy test [o]		X	X							
Daily completion of diaries by subjects for VMS [p]		←	←	—	—	→				
for medication intake all subjects			←	—	—	→				
for medication intake [h]						←	—	—	—	→
for bleeding events [h]			←	—	—	—	—	—	—	→
Weekly reminder [q]		←	—	—	—	→				
Dispense paper diary [r]		X	X							
Return of paper diary			X	X	X	X	X [h]	X [h]	X [h]	X [h]
Review of paper diary [s]			X	X	X	X	X [h]	X [h]	X [h]	X [h]
Randomization			X							
Dispense trial medication			X	X	X	X	X	X	X [t]	
Daily intake of trial medication			←	—	—	—	—	—	→	
Return trial medication				X	X	X	X	X	X	X [h]
Drug accountability				X	X	X	X	X	X	X [h]
Prior/concomitant medication	X	X	X	X	X	X	X	X	X	X [h]
Adverse Events (AEs)	X	X	X	X	X	X	X	X	X	X [h]

CGI=Clinical Global Impression; discont=discontinuation; E2=estradiol; E4=estetrol; ECG=electrocardiogram; EoT=end of treatment; FSH= follicle stimulating hormone; gyn=gynecological; Hematol/chem=hematology/chemistry; I/E= in and exclusion criteria; med=medical; MENQOL=Menopause-specific Quality of Life; PAP= Papanicolaou; TSH= Thyroid stimulating hormone; TVUS=transvaginal ultrasound; VMS= vasomotor symptoms.

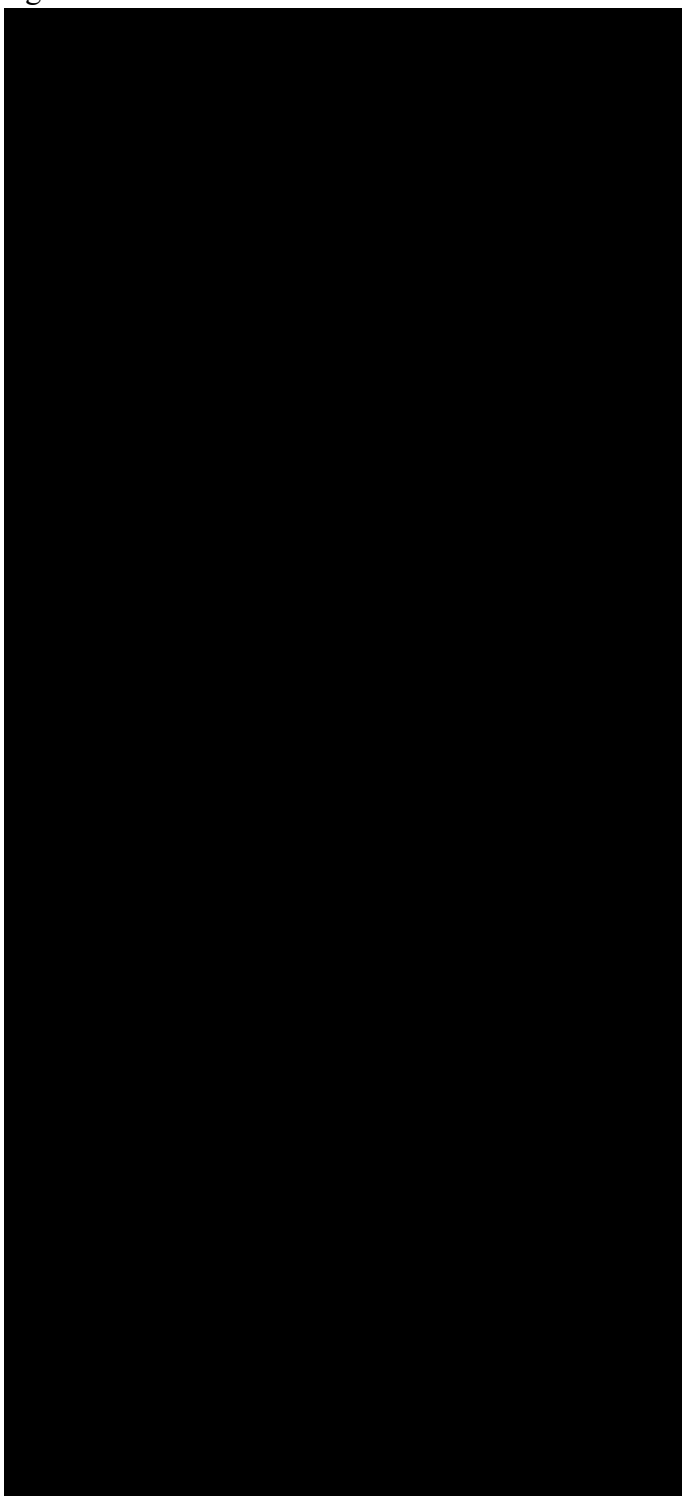
a) In case the subject uses medication that is not allowed during the duration of the trial, she is asked to stop this medication. The subject will return for the screening after a variable washout period that depends on the type of prior medication. Subjects not using medication that needs to be stopped and washed out can continue with the screening procedures.
Washout periods are: 1 week for vaginal hormonal products (rings, creams, gels); 4 weeks for transdermal estrogen or estrogen/progestin products; 8 weeks for oral estrogen and/or progestin and/or selective estrogen receptor modulator therapy; 8 weeks for intrauterine progestin therapy; 3 months for progestin implants or estrogen alone injectable drug therapy; 6 months for estrogen pellet therapy or progestin injectable drug therapy; 8 weeks for oral, topical, vaginal, patch of androgen/DHEA; 6 months for implantable or injectable androgen therapy; 2 weeks for phytoestrogens or black cohosh, 3 months for venlafaxine and desvenlafaxine if used for the treatment of VMS and 4 weeks all other non-hormonal prescription or over-the-counter treatments used for VMS.
b) Every effort needs to be made to perform the procedures of the end of trial visit for those subjects that discontinue their participation.
c) Visits will be scheduled in one of the seven days of week 5, week 9, week 13, week 29, week 41 and week 53, thus allowing a visit window of 7 days. Unscheduled visits may be needed during the treatment period for additional ultrasounds, biopsies or any other unforeseen safety assessments.
d) Visit 8 for non-hysterectomized subjects only at approximately 10 days after completion of the progestin treatment.
e) At washout visit and during the Screening Period only available I/E criteria at that time point will be checked.
f) Blood pressure, heart rate, body height (during screening only) and body weight (during screening and EoT only).
g) To be performed in subjects with uterus and/or cervix, unless a written normal result is available within 18 months before screening start. The PAP test may be repeated once in case of inadequate or insufficient sample.
h) Non-hysterectomized subjects only.
i) If the subject was treated for at least 12 weeks
j) TVUS at screening and EoT will be taken for all subjects who still have a uterus and/or ovaries. At all other time points a TVUS will be performed to monitor the endometrium thickness only for subjects who still have a uterus.
k) Unless a written normal result is available within 9 months before screening start.
l) Mammography at EoT visit may only be performed if the subject was treated for at least 12 weeks and the last mammography was taken at least 9 months earlier)
m) If the subject is not fasting, a new appointment needs to be made as soon as possible. Laboratory tests may be repeated once during the screening period, at the discretion of the investigator.
n) Additional blood samples including one DNA sample at Visit 1 only, will be collected for optional additional analysis of in the context of this program or any further research on E4. o) A pregnancy test may be performed at the discretion of the Investigator.
p) VMS will be recorded daily in the paper diary during the Screening period and during the treatment period up to Visit 4.
q) During VMS recording subjects will be reminded weekly (phone, e-mail, SMS etc.) to complete the diary. The first reminder will be sent 5 ± 2 days after start of VMS recording. r) Subjects are instructed during screening on how to fill in the diary and if needed can be reinstructed.
s) At each visit the diary will be checked. Completed pages will be collected and will remain at the study site.
t) Progestin treatment for non-hysterectomized subjects only.
u) If lipid and/or glucose parameters could be a reason of concern in the judgement of the Investigator, fasted blood sampling for lipid and glucose parameters assessment may be done during the Washout Visit. Otherwise, this assessment will be done during the screening period.
v) Laboratory tests may be repeated once during the screening period, at the discretion of the investigator. FSH measurement is not required for subjects with bilateral oophorectomy.
w) In the event screening requirements cannot be met within 4 weeks, the screening period may be extended with prior Medical Monitor approval, but screening period may not exceed 8 weeks.
a)

Appendix 3 Menopause Specific Quality of Life (MENQOL) Questionnaire

Modified MENQOL for Screening

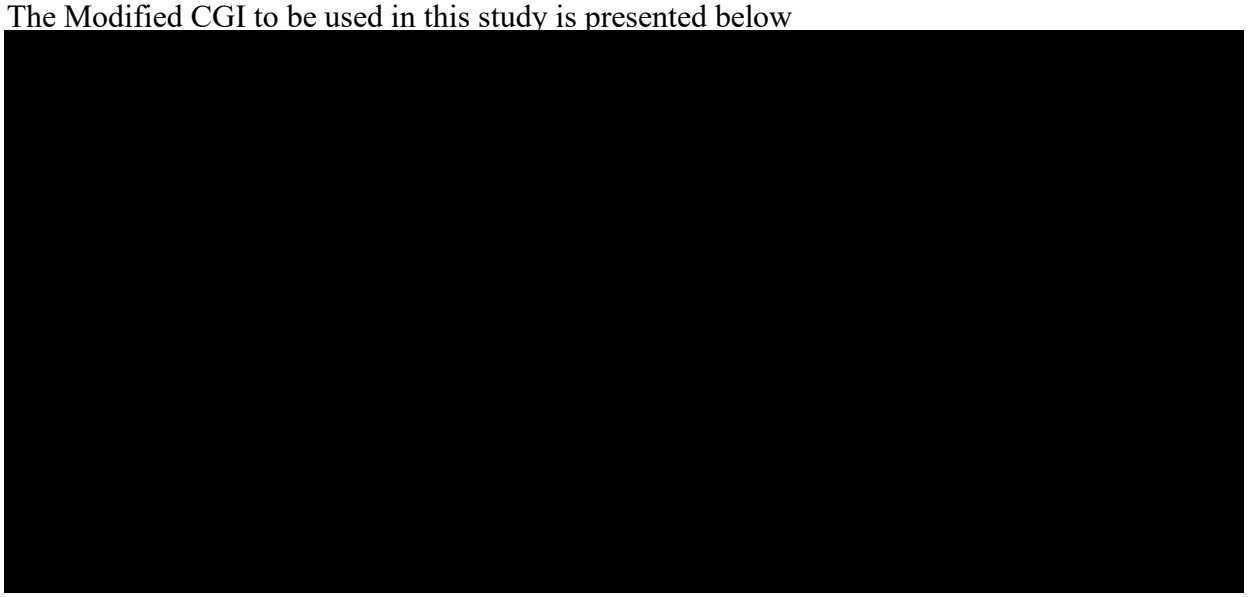


MENQOL for During Treatment Assessments



Appendix 4 Modified Clinical Global Impression

The Modified CGI to be used in this study is presented below



Appendix 5 Histologic Descriptions Recommended for use when Reading Endometrial Biopsy Slides

Histologic Characteristics of the Endometrium

0. No tissue
1. Tissue insufficient for diagnosis
2. Atrophic
3. Inactive
4. Proliferative
    a. Weakly proliferative
    b. Active proliferative
    c. Disordered proliferative
5. Secretory
    a. Cyclic type
    b. Progestational type (including stromal decidualization)
6. Menstrual type
7. Simple hyperplasia without atypia
8. Simple hyperplasia with atypia
9. Complex hyperplasia without atypia
10. Complex hyperplasia with atypia
11. Carcinoma (specify type)

Additional Histologic Characteristics:

If there are any polyps, please specify the type or types.

- Functional
- Atrophic
- Hyperplastic without atypia
- Hyperplastic with atypia
- Carcinomatous

If there is any stromal tissue, please specify the type or types.

- Smooth muscle tissue, normal
- Features suggestive of adenomyoma
- Features suggestive of stromal nodule
- Sarcoma (specify type)

If there is any metaplasia, please specify the type or types.

- Squamous
- Papillary
- Eosinophilic
- Ciliated
- Mucinous
- Syncytial
- Other type (specify type)

If there is any cervical tissue, please specify the type or types.

- Fragments of negative cervical epithelium
- Endocervical polyp
- Atypical endocervical glandular epithelium
- Atypical squamous metaplasia
- Squamous dysplasia
- Cervical carcinoma

Category Indicators for Final Diagnosis

Category	Histologic Characteristics of the Endometrium
Category 1	0,1,2,3,4,5,6
Category 2	7
Category 3	9
Category 4	8
Category 5	10
Category 6	11