Safety 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 (E4 Comfort Study II)
Protocol Number:	MIT-Do001-C302
Protocol Version, Date	Final Version 7.1 (Amendment 6.1), 30 July 2021
ICON ID:	3803/0007
Document Version, Date:	Final 3.0, 07 July 2022

Prepared by:

ICON Clinical Research Services

On behalf of:

Estetra SRL

Confidentiality statement:

- The information provided in this document is strictly confidential.
- The recipients of the SAP must not disclose the confidential information contained within this document or any related information to other persons without the permission of the sponsor.
- In addition, the recipients of the SAP must keep this confidential document in a controlled environment which prevents unauthorized access to the document.

SIGNATURE PAGE

Prepared at ICON Clinical Research by:

[redacted]

[redacted]

Date
(DD MMM YYYY)

Reviewed at ICON Clinical Research by:

[redacted]

[redacted]

Date
(DD MMM YYYY)

[redacted]

[redacted]

Date
(DD MMM YYYY)

Approved at Estetra SRL by:

[redacted]

[redacted]

Date
(DD MMM YYYY)

[redacted]

[redacted]

Date
(DD MMM YYYY)

[redacted]

[redacted]

Date
(DD MMM YYYY)

REVISION HISTORY

Version/Date	Version name	Section	Changes implemented
Final Version 1.0/ 24JAN2020	Initial approved version	N/A	N/A
Final Version 1.1/ 31Mar2020	Updated version	Section 1, Section 2.2/2.4, Section 3.1, Section 3.3, Section 4.1, Section 4.2, Section 8.2, Section 8.4.2, Section 8.7.6.5, Section 8.7.6.6, Section 8.7.6.7, Section 8.7.6.8, Section 10, Appendix 1 and 2, Appendix 5, Section 8.7.1, Section 8.1	Whole document: Update according to Protocol Version 5.0 dated 03Feb2020 (Amendment 4). Section 1: Reference to Protocol Version 5.0 dated 03Feb2020 (Amendment 4). Section 2.2/2.4: Addition of two secondary objectives for the non-hysterectomized subjects according to Protocol Version 5.0 dated 03Feb2020 (Amendment 4). Section 3.1: Updated according to Protocol Version 5.0 dated 03Feb2020 (Amendment 4). Section 3.3: Replacement of the text 'randomization visit' by 'treatment allocation visit' (wording more adapted to Safety Study part). Section 4.1: Addition of the four safety endpoints corresponding to the two secondary objectives for the non-hysterectomized subjects according to Protocol Version 5.0 dated 03Feb2020 (Amendment 4). Section 4.2: Updated according to Protocol Version 5.0 dated 03Feb2020 (Amendment 4). Section 8.2: For subjects who complete the treatment percentages will be computed on the number of subjects treated (and not on the Included Set as planned in the previous version). Section 8.4.2:

			hysterectomy status (hysterectomized and non-hysterectomized) added to the baseline characteristics to be summarized. New sections added corresponding to the four new safety endpoints according to Protocol Version 5.0 dated 03Feb2020 (Amendment 4): Section 8.7.6.5 Section 8.7.6.6 Section 8.7.6.7 Section 8.7.6.8 Section 10: Reference to Protocol Version 5.0 dated 03Feb2020 (Amendment 4). Appendix 1 and 2: Updated according to Protocol Version 5.0 dated 03Feb2020 (Amendment 4). Appendix 5 added according to Protocol Version 5.0 dated 03Feb2020 (Amendment 4). Section 8.7.1: Addition of the table of Serious TEAEs by SOC, PT. Section 8.1: Addition of the missing values count in the descriptive statistics if there is missing values.
Final Version 1.2/ 23Sep2020	Updated version		Updated per Protocol Version 6.0 dated 23Sep2020 (Amendment 5)
Final v2.0/5Nov2020	Updated version	Section 7.3 Section 8.7.6.6	Updated per Sponsor's request
Final 2.1/09Mar2022	Updated version		Updated per Protocol Version 7.1 dated 30Jul2021 (Amendment 6.1)
Final 2.2/27Apr2022	Updated version		Updated as per sponsor's request.
Final 3.0/7July2022	Updated version	Section 7.7.6.6	Updated derivations of Initial, Safety and Final/Consensus diagnoses based on adequate and evaluable biopsies.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse Event
AESI	Adverse Event of Special Interest
ATC	anatomical therapeutic chemical
BMI	body mass index
BI-RADS	Breast Imaging-Reporting And Data System
BLQ	below the lower limit of quantification
BP	blood pressure
CGI	clinical global impression
CI	confidence interval
CRF	Case Report Form
CSR	Clinical Study Report
CTMS	clinical trial management system
E2	Estradiol
E4	Estetrol
EC	Ethics Committee
ED	early discontinuation
ECG	electrocardiogram
EoT	end of treatment
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
LDL	low-density lipoprotein
LDH	lactate dehydrogenase
Max	maximum

MedDRA	Medical Dictionary for Regulatory Activities®
MENQOL	menopause-specific quality of life
Min	minimum
mITT	modified Intent-To-Treat
NH	non-hysterectomized
P4	progesterone
PAP	Papanicolaou
PP	Per-Protocol
PT	preferred term
PY	patient year
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	standard deviation
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 Safety Study part of study based on:

Protocol MIT Do001-C302 version 7.1 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 (E4 Comfort Study II), dated 30Jul2021

Case Report Form (CRF) MIT-Do001-C302_aCRF_version 2.7 for Clinical Study Report (CSR) analysis, dated 08Jan2021.

MIT-Do001-C302 Efficacy Study Part SAP version 3.0, dated 05Apr2022.

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 Endpoint

To evaluate the general safety of treatment with Estetrol (E4) 20 mg.

Safety endpoints for primary objective #1

- Frequency of Treatment Emergent Adverse Event (TEAE) (including Serious Adverse Events [SAEs])
- Frequency of changes in results in physical and gynecological examinations, 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

2.2 Secondary Objectives and Endpoints

1) To evaluate the effect of treatment with E4 20 mg on lipid and glucose metabolism.

Efficacy endpoints for secondary objective #1

- 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)

2) To evaluate the effect of treatment with E4 20 mg on health related quality of life (HRQoL) and treatment satisfaction (TS).

Efficacy endpoints for secondary objective #2

- 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 weeks, 12 weeks, and 52 weeks of treatment using the clinical global impression (CGI) questionnaire

3) To evaluate the effect of treatment with E4 20 mg on the endometrium in non-hysterectomized subjects.

Safety endpoints for secondary objective #3

- 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 (see Appendix 16.3 of protocol)

4) To evaluate the effect of treatment with E4 20 mg on vaginal bleeding in non-hysterectomized subjects.

Safety endpoints for secondary objective #4

- Frequency of women with vaginal bleeding and/or spotting during each 28-day cycle of treatment with E4 based on data in the subject diary
- Number of days with bleeding and/or spotting during each 28-day cycle of treatment based on data 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 data 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

Not Applicable

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 vasomotor symptoms (VMS) associated with menopause.

The Safety Study part will be conducted in approximately 200 hysterectomized and approximately 200 non-hysterectomized postmenopausal women, who are seeking treatment for the relief of VMS (condition clinically indicated by the investigator) associated with menopause and have at least 1 moderate to severe VMS symptom per week.

The Safety Study part is an open label study, to evaluate the effect of E4 20 mg (Arm 4) on the primary safety and secondary efficacy endpoints (lipid and glucose metabolism, HRQoL and TS) over a 52 week treatment period. For endometrial protection, all non-hysterectomized subjects will receive treatment with 200 mg progesterone (P4) once daily for 14 consecutive days, after completion of the E4 treatment. The target number of subjects to be included in the Safety Study part is approximately 400. Enrollment in the Safety Study part will stop as soon as the minimum required number of subjects in each study group and study part is enrolled.

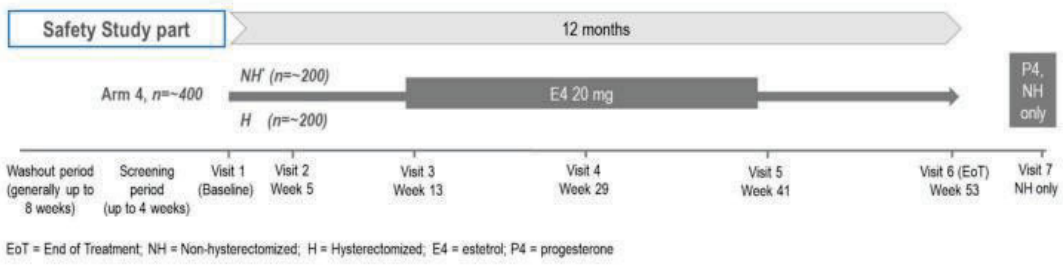
At least 8 visits (up to 15 months) are planned in the Safety Study part, this includes:

- 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 1 to Visit 6), include:
  - Baseline (Visit 1),
  - Five On-treatment visits (Visits 2, 3, 4, 5 and 6), the last treatment visit (Visit 6) is also the end of treatment (EoT) visit, which will also occur in case of early discontinuation.

For non-hysterectomized women an additional endometrial safety follow-up visit (visit 7) is planned approximately 10 days after completion of the 14 days of P4 treatment, which will also occur in case of early discontinuation

The trial schedule for Safety Study part is summarized in Figure 1.

Figure 1: Trial Schedule



EoT = End of Treatment; NH = Non-hysterectomized; H = Hysterectomized; E4 = estetrol; P4 = progesterone

*For non-hysterectomized (NH) subjects endometrial thickness will be measured after 3, 6, 9 and 12 months (Visits 4, 5, 6 and 7 of the Efficacy Study part and Visits 3, 4, 5 and 6 of the Safety Study part respectively) of treatment by TVUS and subjects will record daily the vaginal bleeding events. If after the first 4 weeks of treatment a subject presents with a bi-layer endometrial thickness >10 mm as assessed by TVUS or presents with persistent and/or recurrent bleeding, confirmed by the investigator, 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 treatment.

Before Protocol Version 7.1 (Amendment 6.1) disordered proliferative endometrium was considered as a reason for the exclusion or discontinuation of the subject from the study.

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. Subjects who do not fulfill the VMS requirements for participation in the Efficacy Study part, may participate in the Safety Study part, providing that they meet all other inclusion and exclusion criteria.

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 completed during the same visit.

At baseline (Visit 1), secondary efficacy assessments (lipid and glucose metabolism) will be performed and the MENQOL questionnaire will be completed. Subjects will visit the trial center after 4 weeks (Visit 2), 12 weeks (Visit 3), 6 months (Visit 4) and 9 months (Visit 5) of treatment for safety assessments, and for completion of the CGI questionnaire (Visit 2 only). At Visit 3, secondary efficacy assessments will also be performed and the MENQOL and CGI questionnaires will be completed. For non-hysterectomized subjects endometrial thickness will be assessed on visits 3, 4 and 5. During treatment non-hysterectomized subjects will record daily study drug intake and vaginal bleeding events in a paper diary. After 12 months of treatment subjects will visit the trial center for the EoT visit (Visit 6) during which secondary efficacy assessments and safety assessments will be performed, questionnaires will be completed and a mammography will be made. For non-hysterectomized subjects endometrial thickness will be assessed and, if they have completed at least 12 weeks of E4, an endometrial biopsy will be performed.

After the EoT procedures, all non-hysterectomized 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 7).

3.2 Randomization and Blinding

At each study center, all subjects included in the Safety Study part (open label study part) will receive E4 20 mg (oral, once daily). Blinding is not applicable for this part of the trial.

3.3 Study Treatments and Assessments

After inclusion in the safety study part, the trial medication will be dispensed (see Section 9.4 in protocol) and subjects will take trial medication (20 mg E4, oral, once daily) on the day of treatment allocation visit (day 1 of week 1) and during the treatment period (Visit 1 - Visit 6). Subjects will be instructed to take the trial medication at approximately the same time each day. If a subject misses a dose, she will 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 will be instructed to omit dosing for that day and continue with the next dose as usual.

An overview of 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 target number of subjects to be included in the Safety Study part is approximately 400.

Enrollment will stop when the minimum required number of subjects in each study group and study part have been enrolled. The number of subjects has been selected to fulfill the requirement stated in ICH E1 guideline (CPMP/ICH/375/95) regarding number of subjects exposed to the drug for safety assessments.

5 ANALYSIS POPULATIONS

5.1 Enrolled Set

Enrolled Set will include all subjects who sign the ICF.

5.2 Included Set

The Included Set will include all subjects allocated to the Safety Study part treatment arm. This will be the main analysis set for the disposition table and all the listings for the Safety study part.

5.3 Safety Analysis Set

The Safety Analysis Set will include all subjects who receive at least one dose of study drug. The Safety Analysis Set will be used for all analyses of efficacy, safety, tolerability, and background characteristics.

5.4 Per-Protocol Set

The Per-Protocol (PP) Set will include all subjects who receive at least one dose of study drug and 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. The PP Set will be used for efficacy sensitivity analyses.

5.5 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 1) 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 1: Derived Variables

<table>
<tr><th>Variables</th><th>Formula</th></tr>
<tr><td colspan="2">Demographic and Baseline Characteristics</td></tr>
<tr><td>Body mass index (BMI) (kg/m^2)</td><td>weight (kg)/[height (m)]^2</td></tr>
<tr><td colspan="2">Study Treatment</td></tr>
<tr><td>Treatment Start Date</td><td>The date of first dose of study drug.</td></tr>
<tr><td>Treatment Stop Date</td><td>The date of last dose of study drug.</td></tr>
<tr><td>Duration of exposure (Days)</td><td>(Treatment Stop Date – Treatment Start Date) + 1</td></tr>
<tr><td>Treatment compliance (%)</td><td>100 x [(total number of tablets dispensed) – (total number of tablets returned)]/(duration of exposure in days)</td></tr>
<tr><td>Study Day</td><td>Prior to Treatment Start Date:
Date of interest – Treatment Start Date

On or after Treatment Start Date:
Date of interest – Treatment Start Date +1</td></tr>
<tr><td>Study Week</td><td>Study Week is defined as follows for the first 7 weeks
<table>
<tr><th>Week</th><th>Start (Study Day)</th><th>End (Study Day)</th></tr>
<tr><td>1</td><td>1</td><td>7</td></tr>
<tr><td>2</td><td>8</td><td>14</td></tr>
<tr><td>3</td><td>15</td><td>21</td></tr>
<tr><td>4</td><td>22</td><td>28</td></tr>
<tr><td>5</td><td>29</td><td>35</td></tr>
<tr><td>6</td><td>36</td><td>42</td></tr>
<tr><td>7</td><td>43</td><td>49</td></tr>
</table></td></tr>
</table>

Study Cycle	28-day period starting from Treatment Start Day
Adverse Events	
Adverse Event (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 (Days)	(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	
Baseline	Unless otherwise noted, Baseline is defined as the last non-missing value, including results from repeated and unscheduled measurements, prior to Treatment Start Date.
Change from Baseline	Date of interest – Baseline
Other	
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 group.

6.2 Handling of Missing Data and Outliers

6.2.1.1 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.

6.2.1.2 Handling of Missing or Incomplete Dates

Imputation rules for missing or partial 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 will be done.

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. 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 relevant in the future, post-hoc analysis may be performed.

7 STATISTICAL METHODS

7.1 General Statistical Conventions

All statistical procedures will be completed using SAS version 9.4 or higher, there is only one treatment group, thus no formal statistical analysis will be performed.

Confidence intervals (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 %', then 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. Standard deviations 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 early discontinuation (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 5 of Appendix 1. Similarly, the unscheduled visit results will be clustered with scheduled visits based on the visit windows defined in Table 6 (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 7 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. 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 for MIT-Do001-C302 study will be described in the SAP of the Efficacy Study part. This summary will also be produced for the Safety Study part based on the Included Set.

Data will be also summarized by hysterectomy status (hysterectomized and non-hysterectomized), including:

- Subjects in the Safety Study part (Included 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 Safety Study part (Included Set). The denominator for subjects who completed or discontinued treatment will be the number of subjects treated in the Safety Study part (Safety Analysis Set).

A table summarizing the analysis sets will be presented by hysterectomy status (hysterectomized and non-hysterectomized) and overall for the Included Set, including:

- Subjects included in the Included Set
- Subjects in the Safety Analysis Set and reason for exclusion
- Subjects included in the PP Set and reasons for exclusion

In addition, a 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 Included Set, with the extent of participation in the study and the reason for discontinuation. In addition, hysterectomy status will be flagged in the listing.

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

7.3 Protocol Deviations

A summary of major protocol deviations will be tabulated by type of deviation for the Safety Analysis Set.

All protocol deviations will be listed for the Included Set.

7.4 Demographics and Baseline Characteristics

7.4.1.1 Demographics

Demographic data will be summarized using the Safety Analysis Set.

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

Listings of demographics will be produced for the Included Set.

7.4.1.2 Baseline Characteristics

Subject baseline characteristics including hysterectomy status (hysterectomized and non-hysterectomized), smoking habits, gynecological history, PAP test, modified MENQOL questionnaire, Follicle Stimulating Hormone (FSH), Estradiol (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 Set. The summary will include descriptive statistics for continuous measures and frequency counts and percentages for categorical measures.

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

Listings of baseline characteristics will be based on the Included Set.

7.4.1.3 Medical History

A summary of medical history will be presented, by system organ class (SOC) and preferred term (PT) using Medical Dictionary for Regulatory Activities® (MedDRA) Version 24.1 or higher, for the Safety Analysis Set.

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

7.4.1.4 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 2019 March version, or higher).

Prior medications: A medication with a start date prior to Treatment Start Date.

Concomitant medications: A medication with a start date on or after Treatment Start Date.

If a medication start date is before the Treatment Start Date and is ongoing at Treatment Start Date, 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 Set. A listing of medication data will be produced for the Included Set.

Partial medication start and stop dates will be imputed as detailed in Section 6.2.2.

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 Set. Also compliance will be summarized categorically as the number of subjects with compliance < 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 Safety Analysis Set.

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

Included Set.

7.6 Efficacy Analyses

Descriptive statistics will be presented for efficacy endpoints. No formal statistical analyses are planned for evaluation of efficacy endpoints. The Safety Analysis Set will be the primary analysis set for the efficacy endpoints. The PP Set will be the sensitive analysis set for the efficacy endpoints.

7.6.1.1 Analysis Methods

Not applicable

7.6.1.2 Multiplicity

Not applicable

7.6.1.3 Treatment by Center Interaction Analysis (multi-center study)

Not applicable

7.6.1.4 Analysis of Efficacy Endpoints

7.6.1.5 Lipid and Glucose Metabolism

Samples for lipid and glucose metabolism will be collected at Visit 1 (baseline), Visit 3 (week 12) and Visit 6 (EoT, week 52), including:

- 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.

The observed value and change from baseline to week 12 and week 52 for all parameters will be summarized using descriptive statistics including the number of subjects (n), mean, SD, median, min, and max and 95% CIs for the Safety Analysis Set and PP Set without imputation of missing data. In case of multiple measurement at the same visit, the last observed value will be used for the summary, regardless of scheduled or unscheduled. Visits will be clustered based on the windows specified in Table 5 of Appendix 1.

Listings will be provided for lipid and glucose metabolism data for the Included Set.

7.6.1.6 Menopause-Specific Quality of Life Questionnaire

7.6.1.6.1 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 immediate post-menopause period. Table 2 provides the list of the relevant numbered items for each domain.

Table 2: 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 complete the modified MENQOL questionnaire, responding for symptoms over the previous 3 days. During the study women 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 3.

Table 3: Scoring of each Items

Experienced that symptom or feeling?	No	Yes						
Endorsement	1	2						
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 6.2.1.

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 6.2.1.

7.6.1.6.2 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, and max and 95% CIs for the Safety Analysis Set and PP Analysis Set. Visits

will be clustered as specified in Table 5 of Appendix 1.

A listing will be provided for MENQOL questionnaire data for the Included Set.

7.6.1.7 Clinical Global Impression Questionnaire

Subjects will answer the question 'Rate 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

Total score improvement after 4, 12 and 52 weeks of treatment using the CGI questionnaire will be summarized using number (n), percentage and percentage 95% CIs in each category for the Safety Analysis Set and PP Set.

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

7.6.1.8 Analysis of Exploratory Endpoint

Not applicable

7.7 Safety Analyses

The safety evaluations will include analyses of TEAEs, physical and gynecological examination, breast examination, vital signs, ECG, routine clinical laboratory tests and mammography. For all safety endpoints, visits will be clustered as specified in Table 6 and Table 7 of Appendix 1.

The safety analyses involving changes from baseline to a specific time point for 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, max), categorical variables will be presented in frequency tables using counts and percentages.

7.7.1.1 Adverse Events

All AE verbatim descriptions will be coded using the MedDRA, version 24.1 or the most updated version.

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 6.2.2.

In summaries by SOC and PT, TEAEs will be sorted by decreasing frequency within each SOC and PT. In summaries by PT, TEAEs will be sorted by decreasing frequency.

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 within 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'. Adverse Events with investigator indicated drug-event relationship of 'Unlikely' and 'Not Related' will be classified as 'Not Related' to study drug, while Adverse Events with drug-event relationship documented as 'Highly probable', 'Probable', 'Possible' will be considered as 'Related' to study drug. Where TEAEs have 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 TEAEs 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 will not be any retrospective reporting.

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

- TEAE;
- Severe TEAEs;
- Drug related TEAEs;

- Serious TEAE;
- Drug related serious TEAEs;
- TEAEs leading to dose interruption;
- TEAEs leading to study drug discontinuation;
- TEAEs leading to death;
- Drug related TEAEs leading to death;
- AESI

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 4;
- 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 4;
- 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 4;
- TEAEs leading to study drug interruption by SOC, PT, and maximum intensity;
- TEAEs of reproductive system and abdominal pain that leading to study drug interruption by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 4;
- TEAEs leading to study drug discontinuation by SOC, PT, and maximum intensity;
- TEAEs of reproductive system and abdominal pain that leading to study drug discontinuation by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non- hysterectomized)), as defined in Table 4;
- Serious TEAEs by SOC, PT;
- Serious TEAEs of reproductive system and abdominal pain by SOC, PT by hysterectomy status (hysterectomized and non- hysterectomized), as defined in Table 4;
- 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 4;
- 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 4;

- TEAEs leading to deaths, by SOC, PT and maximum intensity;
- TEAEs of reproductive system and abdominal pain that leading to deaths, by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non- hysterectomized), as defined in Table 4;
- Drug related TEAEs leading to deaths, by SOC, PT and maximum intensity;
- Drug related TEAEs of reproductive system and abdominal pain that leading to deaths, by SOC, PT and maximum intensity by hysterectomy status (hysterectomized and non-hysterectomized), as defined in Table 4;
- AESI by SOC, PT and hysterectomy status (hysterectomized and non- hysterectomized);
- AESI by SOC, PT and 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 CV events, n per 1000 patient-years (n/1000 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 Patient-Years = n/ PY*1000

  where,

  n = number of events in each group.

  PY = (End of Treatment Date – First treatment date + 1)/365.25 or
  (cut-off date – First treatment date + 1)/365.25 for ongoing subjects.

Table 4: Reproductive system and abdominal pain preferred terms

	AEBODSYS	AEDECOD
Reproductive system	Reproductive system and breast disorders	

Abdominal pain	Gastrointestinal disorders	Abdominal pain Abdominal pain lower Abdominal pain upper

Listings will be provided for the Included Set (including flagged for TEAEs) include:

- 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

7.7.1.2 Clinical Laboratory Examinations

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. below the lower limit of quantification (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 values and change from baseline (n, mean, SD, median, min, and max) for continuous assessments will be provided by visit 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 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 Included Set.

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

7.7.1.3 Vital Signs

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

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

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

7.7.1.4 Physical Examinations

Physical examination results will be summarized by frequency counts (n) and percentages for each body system at baseline and post-dose scheduled visit 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 without clinical significance, and abnormal with clinical significance from baseline to post-dose scheduled visits for the Safety Analysis Set.

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

7.7.1.5 Electrocardiograms

Only ECG status (normal, abnormal non clinically significant, abnormal clinically significant) will be collected in the CRF.

A shift table will be produced to summarize the change in the frequency counts (n) and percentages of subjects with normal, abnormal without clinical significance, and abnormal with clinical significance from baseline to post-dose scheduled visits for the Safety Analysis Set.

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

7.7.1.6 Other Safety Assessments

7.7.1.7 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 without clinical significance, and abnormal with clinical significance from baseline to post-dose scheduled visits for the Safety Analysis Set.

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

7.7.1.8 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 without clinical significance, and abnormal with clinical significance from baseline to post-dose scheduled visits for the Safety Analysis Set.

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

7.7.1.9 Mammography

Subjects must have a Breast Imaging-Reporting and Data System (BI-RADS) score of 1 or 2 to enroll in the study.

A shift table will be produced to summarize change in frequency counts (n) and percentages of subjects BI-RADS score from baseline to EoT visit for the Safety Analysis Set.

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

7.7.1.10 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 and percentages at baseline and post-dose scheduled visits (in non-hysterectomized subjects) 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 without clinical significance, and abnormal with clinical significance TVUS image interpretation from baseline to post-dose scheduled visits (in non-hysterectomized subjects) for the Safety Analysis Set.

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

7.7.1.11 Endometrial Thickness Measured by Ultrasound

TVUS will be performed to monitor the endometrium thickness during the treatment at Visit 3, 4, 5 and 6 for non-hysterectomized subjects.

Descriptive statistics of observed values and change from baseline (n, mean, SD, median, min, and max) for endometrial thickness measured by ultrasound will be provided by visit for the Safety Analysis Set.

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

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

7.7.1.12 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, count and percentage 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.

7.7.1.12.1 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 the 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 the end of treatment/early discontinuation (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, endometrial biopsy will be performed at end of treatment 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.

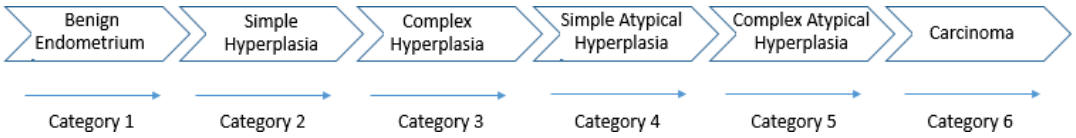
Final/Consensus Diagnosis:

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.

Starting from Protocol Version 7.1 (Amendment 6.1) disordered proliferative endometrium is not considered as a reason for the exclusion or discontinuation of the subject from the study.

The subject with disordered proliferative endometrium will continue in the trial with regular follow-up according to the Protocol.

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.

7.7.1.12.2 Analysis of Endometrial Tissue

All the analyses of endometrial tissues will be performed only for non-hysterectomized subjects.

- 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 all visits: The worst case derived from all on-treatment visits, including unscheduled visits, will be summarized for Final/Consensus diagnosis.

Visits will be clustered as specified in Table 6 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 diagnosis and Safety diagnosis will be summarized separately by visits 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 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 for the Safety Analysis Set for Final/Consensus diagnosis. Worst-case analysis across all on-treatment visits will also be summarized for the Safety Analysis Set for the Final/Consensus diagnosis.

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

For the exclusionary subjects due to disordered proliferative endometrium and worse, a listing of Patients with Abnormal Endometrial Biopsy will be provided with all biopsy results per pathologist, initial pathology report, safety and final diagnosis for the Included 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 quality control is outside of the scope of this SAP.

7.7.1.13 Vaginal Bleeding and/or Spotting

Absence or occurrence of vaginal bleeding and/or 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 percentage 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 non-hysterectomized subjects) for the Safety Analysis Set.

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

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

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

- bleeding for the day on which an endometrial biopsy was performed and 6 days after biopsy and
- bleeding reported after discontinuation of study drug (for example withdrawal bleeding, on P4 treatment)

7.7.1.14 Amenorrhea

Subjects with amenorrhea (absence of any bleeding or spotting) during each 28-day cycle of treatment with E4 (after exclusion the bleeding reported on the day of biopsy and 6 days after biopsy, or after discontinuation of study drug) will be summarized by using frequency count and percentages based on subject diary (in non-hysterectomized 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 Study Day 1 to Study 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

Summary tables will also be presented by hysterectomy status (hysterectomized and non-hysterectomized) as detailed in the respective statistical methods sections.

7.9 Interim Analysis

Not applicable

8 CHANGES TO PLANNED ANALYSIS FROM STUDY PROTOCOL

- The Endometrial Safety Analysis Set, defined in protocol section 12.1, has been removed from SAP section 5; the endometrial analysis will be performed based on the Safety Analysis Set.
- The Intent-To-Treat (ITT) Set, defined in protocol section 12.1, has been removed from SAP section 5. The definition of the ITT and Safety Analysis Sets is same as this is an open label study and only one treatment group. The Safety Analysis Set will be the primary analysis population for the efficacy analyses.
- The modified Intent-To-Treat (mITT), defined in protocol section 12.1, has been removed from SAP section 5. The mITT was intended to be used in a sensitivity analysis of the primary efficacy endpoint, which is not relevant for the Safety Part of the study.

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. 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), Final Version 5.0, February 3, 2020.
4. 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), Final Version 6.0, 23Sep2020.
5. 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), Final Version 7.1, 30July2021.

10 APPENDICES

Appendix 1 Visit Schedule and Permitted Time Window:

Table 5: Visit Window for Efficacy Analysis of Safety Study part

Assessment Visit	Time window	Visit
Washout visit	n.a.	Week - 12 to week -4[1]
Screening period	n.a.	Week - 4 to week - 1[1]
Visit 1	n.a.	Week 1
Visit 2	Day 29 - 35	Week 4
Visit 3	Day 85 - 91	Week 12
Visit 4	Day 197 - 203	Week 28
Visit 5	Day 281 - 287	Week 40
Visit 6	Day 365 - 371	Week 52
Visit 7 NH subjects only	Day 379 - 392	Week 55/56

[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 6 weeks.

n.a.: not applicable

Table 6: Visit Window Microscopic Findings and for Laboratory Endpoints

Assessment Visit	Time window	Visit
Safety Visit 2		Week 5
Safety Visit 3		Week 13
Safety Visit 4		Week 29
Safety Visit 5		Week 41
Safety Visit 6/EoT		Week 53 (if discontinuation reason is missing in DS domain)
		Early Discontinuation (if discontinuation reason is non-missing in DS domain)
Safety Visit 7 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 325	Week 41
	326 to 375*	Week 53
	>=327#	Week 53
	>=376*	Week 55/56 Follow-up

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

Table 7: Visit Window ECG, Procedures, Questionnaire, Vital Signs, Imaging and Findings

Assessment Visit	Time window	Visit
Safety Visit 2		Week 5
Safety Visit 3		Week 13
Safety Visit 4		Week 29
Safety Visit 5		Week 41
Safety Visit 6/EoT		Week 53 (if discontinuation reason is missing in DS domain)
		Early Discontinuation (if discontinuation reason is non-missing in DS domain)
Safety Visit 7 Follow-up		Week 55/56 Follow-up

Appendix 2 Schedule of Trial Procedure

Procedures and assessments	Washout period [a]	Screening period	Baseline (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
	Generally within 8 weeks before screening	Within 4 weeks before baseline [a]	Day 1	Week 5 [c]	Week 13 [c]	Week 29 [c]	Week 41 [c]	Week 53 [c]	Week 55/56 [m]
Informed consent	X								
I/E criteria	X [d]	X [d]	X						
Start washout [a]	X								
Med. & gyn. history	X								
Demographic data	X								
Physical examination		X						X	
Gyn. examination		X						X	
Vital signs [e]		X	X	X	X	X	X	X	
ECG		X						X	
PAP test [f]		X							
Endometrial biopsy [n]		X						X [o]	
TVUS [l]		X			X	X	X	X	X
Breast examination		X				X		X	
Mammography		X [g]						X [h]	
Modified MENQOL questionnaire		X							
MENQOL questionnaire			X		X			X	
CGI questionnaire				X	X			X	
Fasted blood sampling for: [i]									
Hematol./chem.		X [i]	X					X	
Lipid/glucose parameters (for inclusion)	X [a, i]	(X) [i]							
Lipid metabolism			X		X			X	
Glucose metabolism			X		X			X	
Optional additional analysis [j]			X					X	

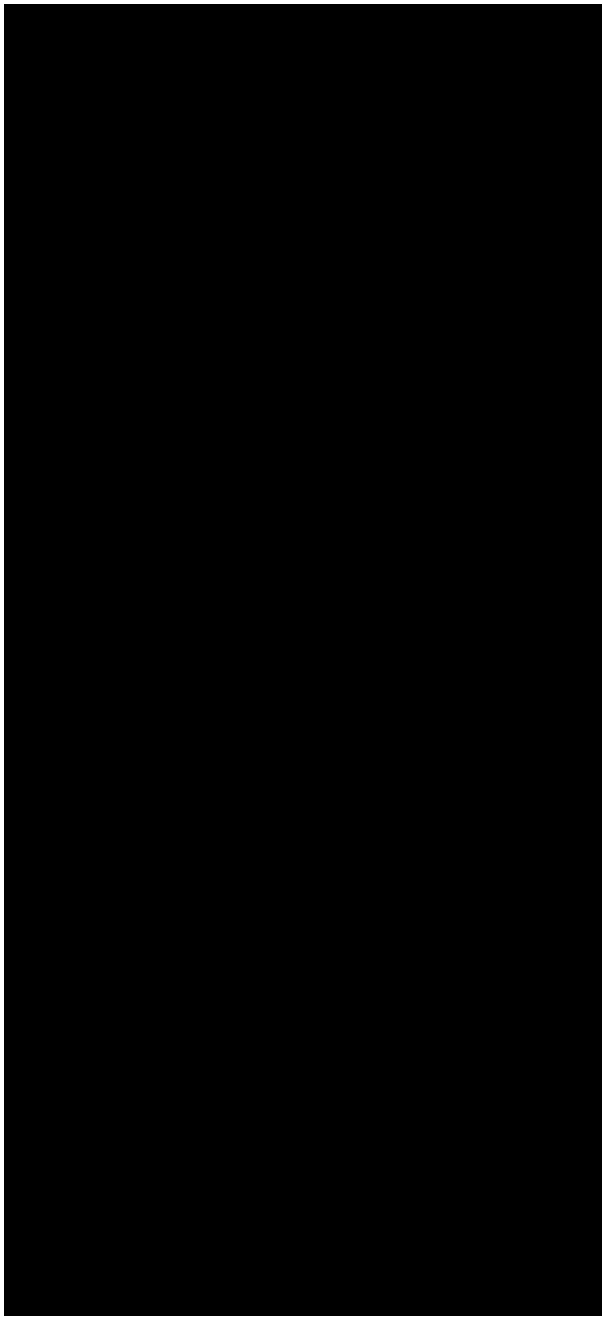
	Washout period [a]	Screening period	Baseline (allocation to treatment)	Treatment visits				EoT/early discontinuation visit [b]	Follow up visit NH subjects
	Washout visit	Screening visit	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7
	Generally within 8 weeks before screening	Within 4 weeks before baseline [a]	Day 1	Week 5 [c]	Week 13 [c]	Week 29 [c]	Week 41 [c]	Week 53 [c]	Week 55/56 [m]
Blood sampling for:									
FSH, E2 and TSH for inclusion [i]		X							
Urinary pregnancy test [k]		X	X						
Daily completion of diaries by subjects [n]									
for medication intake			←	—	—	—	—	—	→
for bleeding events			←	—	—	—	—	—	→
Dispense paper diary [n,p]		X	X						
Return of paper diary [n]			X	X	X	X	X	X	X
Review of paper diary [n,q]			X	X	X	X	X	X	X
Allocation to Treatment			X						
Dispense trial medication			X	X	X	X	X	X [r]	
Daily intake of trial medication			←	—	—	—	—	→	
Return trial medication				X	X	X	X	X	X
Drug accountability				X	X	X	X	X	X
Prior/concomitant medication	X	X	X	X	X	X	X	X	X
Adverse Events (AEs)	X	X	X	X	X	X	X	X	X

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

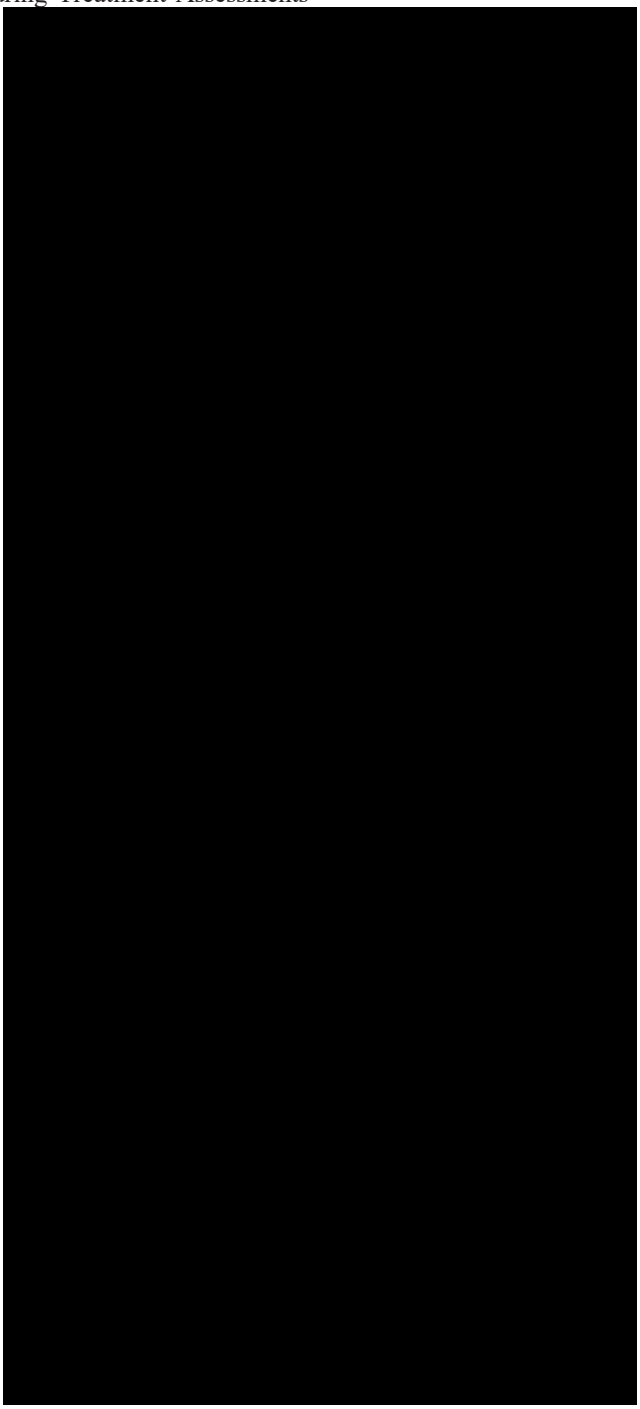
a) In case the subject uses medication that is not allowed during the duration of the trial, she is asked to stop this medication. The subject will return to complete the screening after a variable washout period that depends on the type of prior medication. Subjects not using medication that needs to 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 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 unforeseen safety assessments.
d) At Washout visit and Screening visit only available I/E criteria at that time point will be checked.
e) Blood pressure, heart rate, body height (during screening only) and body weight (during screening and EoT only).
f) 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.
g) Unless a written normal result, including digital imaging, is available within 9 months before screening start.
h) 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
i) 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.
j) 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.
k) A pregnancy test may be performed at the discretion of the Investigator.
l) 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.
m) Visit 7 for non-hysterectomized subjects only at approximately 10 days after completion of the progestin treatment.
n) Non-hysterectomized subjects only.
o) If the subject was treated for at least 12 weeks.
p) Subjects are instructed during screening on how to fill in the diary and if needed can be reinstructed.
q) At each visit the diary will be checked. Completed pages will be collected and will remain at the study site.
r) Progestin treatment for non-hysterectomized subjects only.
s) 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
t) 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.
u) 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.

Appendix 3 Menopause Specific Quality of Life Questionnaire

Modified MENQOL for Screening

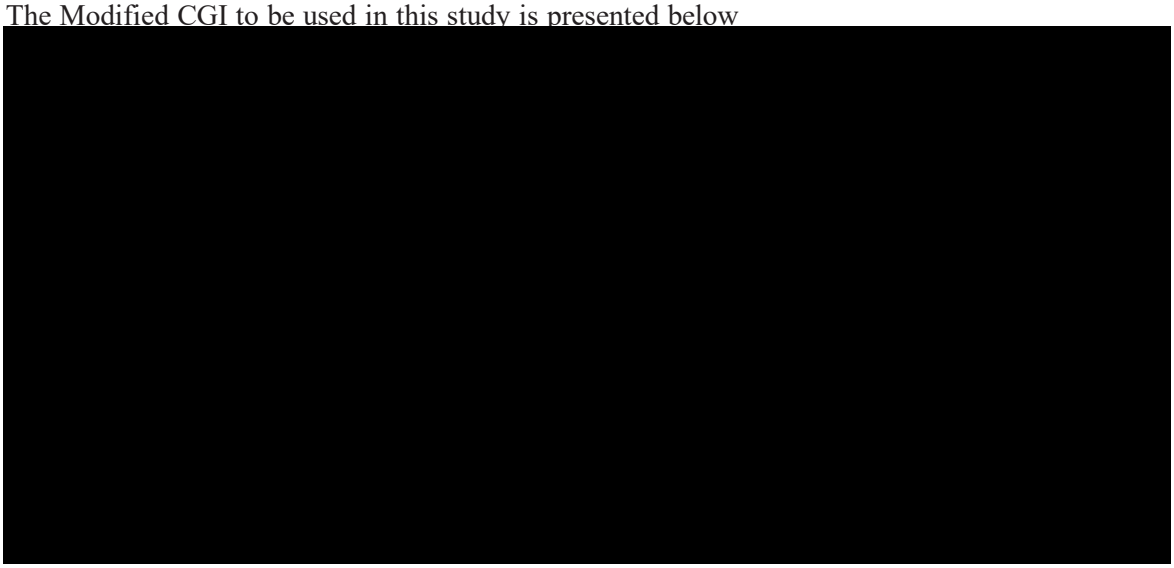


MENQOL for During Treatment Assessments



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)

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

Table 8: 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