Shionogi Study Title:	A Phase 3, Randomized, Double-blind, Placebo controlled Multicenter Study to Evaluate the Efficacy and Safety of Ospemifene in Patients with Moderate to Severe Vaginal Dryness, a Symptom of Vulvo-vaginal Atrophy (VVA) due to Menopause
Shionogi Study Number:	1517I0231
ClinicalTrials.gov registration No.	NCT02638337
Study Document	Statistical Analysis Plan 17 August 2017

STATISTICAL ANALYSIS PLAN: 1517I0231

Study Title:	A Phase 3, Randomized, Double-blind, Placebo-controlled Multicenter Study to Evaluate the Efficacy and Safety of Ospemifene in Patients with Moderate to Severe Vaginal Dryness, a Symptom of Vulvo-vaginal Atrophy (VVA) due to Menopause
Study Number:	1517I0231
Study Phase:	3
Product Name:	Ospemifene
Sponsor:	Shionogi Inc.
Version Number:	Version 1.0
Issue Date:	17 Aug 2017

SIGNATURE PAGE

Sponsor's Authors:

PPO

18AUG2017
Date

18 Aug 2017
Date

RECORDS ON REVISIONS

Version	Date for Creation or Modification	Person in Charge	Remarks
Version 1.0	17 Aug 2017	PPO	Originally prepared

TABLE OF CONTENTS

SIGNATURE PAGE 2
RECORDS ON REVISIONS 3
TABLE OF CONTENTS 4
LIST OF IN-TEXT TABLES 6
1 INTRODUCTION 7
2 STUDY OVERVIEW 7
3 STUDY OBJECTIVES 7
3.1 Primary Objectives 7
3.2 Secondary Objectives 7
4 STUDY DESIGN 8
4.1 Schedule 8
4.2 Study Blinding and Randomization 8
4.3 Sample Size 8
5 ANALYSIS POPULATIONS 9
6 STATISTICAL CONSIDERATIONS FOR DATA ANALYSES AND HANDLING 9
6.1 Statistical Reporting 9
6.2 Statistical Testing 10
6.3 Study Day and Analysis Windows 10
6.4 Missing Data 11
6.5 Pooling of Small Centers 11
7 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS 12
7.1 Subject Disposition and Discontinuation 12
7.2 Demographic and Baseline Characteristics 12
8 STUDY CONDUCT 12
8.1 Protocol deviation 12
8.2 Treatment Exposure and Compliance 13
8.3 Prior and Concomitant Medications 13
9 EFFICACY 14
9.1 Primary Efficacy Analysis 14
9.1.1 Primary Efficacy Endpoint 14
9.1.2 Statistical Analysis Methods for Primary Efficacy Endpoint 15
9.2 Secondary Endpoints That Will Be Analyzed Inferentially 16
9.2.1 Endpoints that correspond to the primary efficacy endpoints, but done at Week 4 and Week 8 rather than at Week 12 16
9.2.2 Severity of VVA symptoms other than vaginal dryness 16

9.2.3 Change from baseline in Maturation Value to Week 4, Week 8, and Week 12 16
9.2.4 Percentage of subjects who are responders at Week 4, Week 8, and Week 12 17
9.2.5 Change from baseline in Vaginal Health Index to Weeks 4, 8, and 12 17
9.2.6 Change from baseline in Vulvar Health Index to Weeks 4, 8, and 12 17
9.2.7 Change from baseline in Vulvo-Vaginal Imaging to Week 12 17
9.2.8 Change from baseline in total score and each domain score of the Female Sexual Function Index questionnaire to Weeks 4, 8, and 12 18
9.2.9 Change from baseline in urinary symptoms to Weeks 4, 8, and 12 18
9.2.10 Change from baseline in markers of bone metabolism to Week 12 19
9.2.11 Frequency of lubricant application and sexual activity 19
9.2.12 Overall satisfaction with treatment 20
9.3 Secondary Endpoints That Will Be Summarized Descriptively Only, With No Inferential Analysis 20
9.3.1 Serum hormone levels 20
9.4 Summary of Statistical Analyses for Efficacy Endpoints 20
10 SAFETY 23
10.1 Adverse Events 23
10.2 Vital Signs 25
10.3 Laboratory Evaluations 25
10.4 Endometrial Safety 26
10.5 Other Safety Variables 26
11 INTERIM ANALYSES 26
12 PROGRAMMING CONVENTIONS 26
13 CHANGES FROM PROTOCOL 28
13.1 Safety Analysis 28
14 REFERENCES 29
Appendix 1 Schedule of Time and Events 30
Appendix 2 Category Score of Patient's Response to Female Sexual Function Index (FSFI) 33
Appendix 3 Pre-defined Limits of Change for Laboratory Test 37
Appendix 4. Pooled Site 38

LIST OF IN-TEXT TABLES

Table 1 Definition of Visit Window 10
Table 2 Definition of Diary Data Window 10
Table 3 Demographic and Baseline Characteristics 12
Table 4 Female Sexual Function Index Scoring System 18
Table 5 Summary of Inferential Efficacy Analyses 21
Table 6 Order of Priority for Counting Number of Subjects 25
Table 7 Safety Laboratory Tests 25
Table 8 Display Digit of Statistics for Efficacy and Safety 27
Table 9 Subject Characteristics and Items of Therapeutic Process 27

1 INTRODUCTION

This "Statistical Analysis Plan" (SAP) presents details, including methods of analyses, of the efficacy and safety analyses described in Section 9 "Planned Statistical Methods" of the protocol [Protocol No. 1517I0231: 02 Nov, 2016] for the clinical trial entitled "A Phase 3, Randomized, Double-blind, Placebo-controlled Multicenter Study to Evaluate the Efficacy and Safety of Ospemifene in Patients with Moderate to Severe Vaginal Dryness, a Symptom of Vulvo-vaginal Atrophy (VVA) due to Menopause".

Details of outputs (or mock-up of tables, listings, and figures) from analyses in this SAP will be included in a separate document ("Mock-up of Tables, Listings, and Figures").

2 STUDY OVERVIEW

This is a multicenter, randomized, double-blind, placebo-controlled study to assess the efficacy and safety of ospemifene 60 mg once daily (QD) in postmenopausal women with vaginal dryness as the most bothersome symptom (MBS) of VVA due to menopause.

3 STUDY OBJECTIVES

The objective of this study is to evaluate the efficacy and safety of ospemifene 60 mg QD compared with placebo in treatment of VVA due to menopause in women with moderate to severe vaginal dryness as the MBS of VVA.

3.1 Primary Objectives

Objective 1: To assess the efficacy of ospemifene 60 mg daily at 12 weeks of treatment, relative to placebo, in the treatment of vaginal dryness, a symptom of vulvar and vaginal atrophy, due to menopause.

Hypothesis: After 12 weeks of treatment, ospemifene 60 mg QD, relative to placebo, significantly improves all 4 co-primary efficacy endpoints:

- Percentage of parabasal cells in the maturation index of the vaginal smear
- Percentage of superficial cells in the maturation index of the vaginal smear
- Vaginal pH
- Intensity of self-reported MBS of vaginal dryness

Objective 2: To assess the safety and tolerability of ospemifene 60 mg QD during 12 weeks of treatment, relative to placebo.

3.2 Secondary Objectives

Objective 1: To assess the effect of ospemifene 60 mg relative to placebo in change from baseline to Week 12 on:

- Severity of other VVA symptoms different from vaginal dryness (i.e., dyspareunia, vulvar/vaginal itching/irritation, dysuria [difficult/painful urination], vaginal bleeding associated with intercourse)
- Markers of bone metabolism
- Vaginal Health Index (VHI)
- Vulvar Health Index (VuHI)

Objective 2: To assess the effect of ospemifene 60 mg relative to placebo in change from baseline to Week 12 on:

- Female Sexual Function Index (FSFI)
- Urinary Distress Inventory (UDI-6)

Objective 3: To assess the effect of ospemifene 60 mg QD, relative to placebo, over 12 weeks on endometrial histology

4 STUDY DESIGN

4.1 Schedule

Schedule of time and events is presented in Appendix 1.

4.2 Study Blinding and Randomization

After screening, subjects who have ≤5% superficial cells in the maturation index of the vaginal smear, a vaginal pH >5.0 and vaginal dryness as MBS of VVA based on self-assessment, who also meet all other eligibility criteria, will be randomized in a 1:1 ratio to receive either ospemifene 60 mg QD or matching placebo for 12 weeks. Randomization will be stratified by severity of MBS of vaginal dryness at Visit 2 and by the presence or absence of the uterus. Randomization of subjects without a uterus will be limited to 60% of the study population.

This study will be conducted in a double-blind manner using a placebo matching the test drug in appearance, labeling, and packaging. All subjects, the investigator, all study personnel, and the sponsor will be blinded to the treatment assigned at randomization until database lock.

4.3 Sample Size

Approximately 600 subjects, about 300 subjects per arm (1:1 randomization ratio), will be randomized in the study. A sample size of 600 subjects provides ≥ 90% power for comparison between ospemifene 60mg and placebo in the 4 co-primary efficacy endpoints (changes from baseline to Week 12 in the percentage of parabasal cells, the percentage of superficial cells, the vaginal pH, and the intensity of MBS of vaginal dryness) with two sided significance level of 0.05.

5 ANALYSIS POPULATIONS

The following populations will be used for analyses.

- **Intent-to-Treat (ITT) Population** is defined as any subject that is randomized and receives at least one dose of study medication. All efficacy analyses will be based on this population. This population will be analyzed according to the treatment to which the subject was randomized. This population will be considered primary.
- **Modified ITT (mITT)** Population will include ITT subjects who had also met the following inclusion criteria:
  1) The subject had 5% or fewer superficial cells in the maturation index of the vaginal smear as baseline values. (Note that subjects with no baseline value for superficial cells will not be included in the mITT Population).
  2) The subject had vaginal pH > 5.0 as baseline values.
  3) The subject had moderate or severe vaginal dryness as baseline values and selected vaginal dryness as the most bothersome VVA symptom as baseline values.

  This population will be used to analyze the primary efficacy endpoints and will be considered supportive.
- **Per-Protocol Population (PP)** will consist of all subjects in the mITT population who meet all of the following:
  1) Complete at least 10 weeks of treatment
  2) Take at least 85% of the study medication within 12 weeks
  3) Do not have any other major protocol deviation. Major protocol deviation will be reviewed to determine eligibility for the PP Population and finalized before breaking the study blind.
  4) Do not have a vaginal infection or any other medical condition that would confound the primary efficacy assessment.

  This population will be used to analyze the primary efficacy endpoints and will be considered supportive.
- **Safety Population** includes all randomized subjects who receive at least one dose of the study medication. This population will be analyzed according to the treatment the subjects actually received rather than the treatment to which the subjects were randomized. If a subject takes both treatments, the subject will belong to the ospemifene group.

6 STATISTICAL CONSIDERATIONS FOR DATA ANALYSES AND HANDLING

6.1 Statistical Reporting

Unless otherwise noted, continuous variables will be summarized by using the number of non-missing observations, arithmetic mean, standard deviation (SD), median, minimum,

and maximum values as descriptive statistics; categorical variables will be summarized by using the frequency count and the percentage of subjects in each category as descriptive statistics.

In general, all tables will be presented by treatment group. Individual subject data and any derived data will be presented by treatment and subject. All analyses and tabulations will be performed by using SAS® Version 9.2 or higher.

6.2 Statistical Testing

All statistical tests will be performed at the 0.05 significance level using two-sided tests, except where otherwise noted.

6.3 Study Day and Analysis Windows

Day 1 is the initial dose date for subjects. Previous day to Day 1 is expressed as Day -1, and days before Day -1 in Screening Period are expressed as Day -2, Day -3, … , etc.

Allowance window for each of the scheduled visits is shown in Table 1.

Allowance window for diary data is shown in Table 2.

If more than one visit occurs during a visit window, the one closest to the target day will be assigned to the visit. If two visits are equidistant from the target day, the latter visit will be assigned to the visit for post-baseline values.

If a subject discontinues from the study, the data obtained at early termination visit will be treated as one visit data based on the allowance windows.

Table 1 Definition of Visit Window

Scheduled Visit	Target Day	Window (Days)
Visit 1	Day -28	≤ Day -1
Visit 2 (Day 1)	Day 1(Randomization)	Day 1
Visit 3 (Week 4)	Day 29	Day 16 to Day 43
Visit 4 (Week 8)	Day 57	Day 44 to Day 71
Visit 5 (Week 12)	Day 85	Day 72 to Day 99
Early termination	Last dosing date	+ 14

Table 2 Definition of Diary Data Window

Week	Daily Data	Weekly Data (Target Day)
Week 1	Day 1 to Day 7	Day 4 to Day 10 (Day 7)
Week 2	Day 8 to Day 14	Day 11 to Day 17 (Day 14)
Week 3	Day 15 to Day 21	Day 18 to Day 24 (Day 21)
Week 4	Day 22 to Day 28	Day 25 to Day 31 (Day 28)
Week 5	Day 29 to Day 35	Day 32 to Day 38 (Day 35)

Week	Daily Data	Weekly Data (Target Day)
Week 6	Day 36 to Day 42	Day 39 to Day 45 (Day 42)
Week 7	Day 43 to Day 49	Day 46 to Day 52 (Day 49)
Week 8	Day 50 to Day 56	Day 53 to Day 59 (Day 56)
Week 9	Day 57 to Day 63	Day 60 to Day 66 (Day 63)
Week 10	Day 64 to Day 70	Day 67 to Day 73 (Day 70)
Week 11	Day 71 to Day 77	Day 74 to Day 80 (Day 77)
Week 12	Day 78 to Day 84	Day 81 to Day 87 (Day 84)

Assessments that occur more than 14 days after the last dose of study medication will be excluded from the above visit windows.

The baseline is defined as the last value that is observed prior to the first dose of study medication. For a variable that is planned to be observed at screening and randomization, the data at randomization will be used as baseline. For such variable, if the data at randomization is missing for a subject, the data observed at screening will be used as baseline.

6.4 Missing Data

If numerical values are missing but characteristic values are not missing because some signs, namely < and >, are included in the characteristic values, the numerical values will be imputed following the below rules:

- If '<' is included in the characteristic value, a smaller value between the value without < or the lower range value will be imputed,
- If '>' is included in the characteristic value, a greater value between the value without > or the upper range will be imputed.

If the information of year or month is missing for duration of VVA, 0 will be imputed for calculating summary statistics but kept missing for listings.

Other missing data will not be imputed. All analyses will be based on observed data.

6.5 Pooling of Small Centers

The study will be conducted under the same protocol across all the study centers. It is anticipated that some centers will not enroll a sufficient number of subjects. To ensure estimable results, small centers will be pooled. Centers with less than 40 subjects will be combined with other centers prior to breaking the blind. Centers with 40 or more subjects may also be pooled if necessary, so that the size of the pooled centers is consistent with the size of the larger centers. The pooling will be done by geographical location if possible. These pooled centers will be used in the analyses model, not the individual centers.

7 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

7.1 Subject Disposition and Discontinuation

The number of subjects who failed at Screening Period will be summarized along with the reasons for failure.

The following analyses will be performed for randomized subjects. Treatment group will be defined as randomized.

Among subjects randomized to each treatment group, the number and percentage of subjects who complete Visit 5, prematurely discontinue before Visit 5 and continue taking medications after Visit 5 will be summarized. In addition, reasons leading to study discontinuation before Visit 5 and after Visit 5 will be summarized for each treatment group. We confirmed all subjects, whose discontinuation reason is entered as 'other' after Visit 5, discontinued the study because of the protocol amendment.

The number and percentage of randomized subjects included in each analysis population and the reasons for excluded subjects will also be presented.

7.2 Demographic and Baseline Characteristics

The distribution of demographics and baseline characteristics shown in Table 3 will be described for the ITT Population by treatment group. Summary statistics will be calculated for quantitative scale items, and frequency and proportion of subjects in each category will be obtained for ordered and qualitative (i.e., binary and nominal) scale items. Quantitative scale items for age and body mass index (BMI) will be categorized and tabulated by treatment group.

For the ITT Population, medical history will be summarized by treatment group. The medical histories will be classified by System Organ Class (SOC) and preferred term (PT), using the Medical Dictionary for Regulatory Activities (MedDRA). MedDRA Version 18.0 will be used.

Table 3 Demographic and Baseline Characteristics

Quantitative scale items	Age, height, body weight, BMI, duration of VVA
Binary or nominal scale items	Ethnicity, race , uterus, current hot flashes, previous hormone treatment, severity of vaginal dryness

BMI = body mass index;
See Section 12 for category.

8 STUDY CONDUCT

8.1 Protocol deviation

Subjects with major protocol deviations will be listed.

8.2 Treatment Exposure and Compliance

Summary statistics of number of days of treatment exposure and treatment compliance rate will be summarized by treatment group based on electronic data capture (EDC) data.

Number of days of treatment exposure is defined as 'actual days of correctly dosed.' The dosing period during which a subject takes medication as follows:

{(End date of dosing) – (initial dose date) + 1} [days]
for subjects who withdraw the study before Visit 5.

{(Date of Visit 5) – (initial dose date) + 1} [days]
for subjects who complete the study or take medication beyond Visit 5.

Treatment compliance rate is defined as:

100{(actual days of correctly dosed)
/ (duration of treatment exposure)} [%].

Data until Visit 5 in ITT and Safety populations will be analyzed.

The same analysis for after Visit 5 will also be performed for Safety Population. For this analysis, subjects who continue to Part B in the previous protocol and continue taking study medication after Visit 5 will be included.

Furthermore, categorized duration of treatment exposure and treatment compliance rate will be summarized by treatment group (see Section 12 for category).

8.3 Prior and Concomitant Medications

Prior and concomitant medications will be coded using the World Health Organization (WHO) Drug Dictionary, March 2015. Prior medication is defined as all medications taken before Day 1 of study medication. If the medication was provided more than 6 months prior to screening and the start date is missing, the drug will also be classified as a prior medication. Concomitant medication is defined as all medications taken at Day 1 or later. If the medication was still ongoing at the end of the study and the end date is missing, the drug will be classified as a concomitant medication. A medication can be classified as both prior and concomitant.

Concomitant medication will be summarized as follows.

- if a subject takes study medication for more than 12 weeks (beyond Visit 5), concomitant medications that will be taken until Visit 5 will be summarized;
- if a subject takes study medication for 12 weeks or less than 12 weeks, concomitant medications taken up to 14 days from the last dose will be summarized .

For ITT Population, the number and percentage of subjects who take prior and concomitant medications will be summarized by preferred name for each treatment group.

All prior and concomitant medications will be listed.

If a subject has more than 1 medication that are coded as the same preferred name, the subject will be counted only once for that preferred name.

If the medication cannot be classified into concomitant medication or prior medication based on the above definitions due to a partial missing date, the rules below will be applied for the classification of the medication.

For start date:

- If the year and month are observed but the day is missing, the first day of the month will be used.
- If the year is observed but the month is missing, the first day of the year, 1st January, will be used.

And, for end date:

- If the year is missing and it is not ongoing, the concomitant medication should be treated as an ongoing treatment.
- If the year and month are observed but the day is missing, the last day of the month will be used.
- If the year is observed but the month is missing, the last day of the year, 31st December, will be used.

The imputed dates will not be displayed in the listings.

9 EFFICACY

All of efficacy endpoints described below will be summarized using descriptive statistics. The subject who doesn't have a baseline value or any post baseline values for a specific parameter will be excluded from the analysis.

9.1 Primary Efficacy Analysis

9.1.1 Primary Efficacy Endpoint

The primary efficacy endpoints will be the difference between ospemifene 60 mg QD and placebo in treatment of VVA due to menopause in women with moderate to severe vaginal dryness as the MBS of VVA at 12 weeks based upon:

- Change from baseline in percentage of parabasal cells in the maturation index to Week 12

- Change from baseline in percentage of superficial cells in the maturation index to Week 12
- Change from baseline in vaginal pH to Week 12
- Change from baseline in severity of MBS of vaginal dryness to Week 12

The definition of baseline is described in the Section 6.3.

9.1.2 Statistical Analysis Methods for Primary Efficacy Endpoint

The following analyses will be performed for both of ITT and PP Populations.
For each of percentage of parabasal cells, percentage of superficial cells and vaginal pH, mixed-effects model for repeated measures (MMRM) approach will be used. In the model, repeated measurements of the change from baseline at Week 4, Week 8, and Week 12 will be the response variable. Treatment, week, treatment by week interaction and study center will be fixed effects. Baseline value will be modeled as covariates. Significance test to compare the ospemifene group with the placebo group will be based on the mean difference between the two groups at Week 12.

SAS code to the above model is as follows:

```
proc mixed data=DRYDATA;
    class USUBJID TREAT CLUSTER WEEK;
    model CHG = TREAT BASE CLUSTER WEEK TREAT*WEEK/ ddfm=kenwardroger;
    repeated WEEK/subject=USUBJID type=un;
    estimate 'Treatment Group Difference at Week12'
             TREAT 1 -1 TREAT*WEEK 0 0 1 0 0 -1;
run;
```

CLUSTER: Pooled center
CHG: Change from baseline to each week (4w,8w,12w)

For the MBS of vaginal dryness, generalized estimating equations (GEE) model will be used to fit a marginal proportional odds model to the longitudinal ordered categorical data [1]. In the model, repeated measurements of the change from baseline in MBS of vaginal dryness at Week 4, Week 8, and Week 12 will be the response variable. Treatment, week, treatment by week interaction and study center will be fixed effects. Baseline severity of dryness will be modeled as covariates. Significance test to compare ospemifene group with the placebo group will be based on the mean of cumulative log odds ratio (over ordered category responses) at Week 12.

SAS code to the above model is as follows:

```
proc genmod data=DRYDATA;
    class USUBJID TREAT CLUSTER WEEK;
    model CHG = TREAT BASE CLUSTER WEEK TREAT*WEEK/
    dist=multinomial link=clogit type3 aggregate=TREAT;
    repeated subject=USUBJID/type=ind;
    estimate 'Treatment Group Difference in Mean of Cumulative Log Odds Ratio at Week12'
             TREAT 1 -1 TREAT*WEEK 0 0 1 0 0 -1 / exp;
run;
```

CLUSTER: Pooled center
CHG: Change in intensity of dryness from baseline to each week (4w,8w,12w)

In order to evaluate the robustness of the GEE analysis for the MBS of vaginal dryness, a sensitivity analysis using MMRM method described above will be performed. In this analysis, ordered categorical data will be treated as continuous data.

9.2 Secondary Endpoints That Will Be Analyzed Inferentially

9.2.1 Endpoints That Correspond to the Primary Efficacy Endpoints, but Done at Week 4 and Week 8 Rather Than at Week 12

- Change from baseline in percentage of parabasal cells to Week 4 and to Week 8
- Change from baseline in percentage of superficial cells to Week 4 and to Week 8
- Change from baseline in vaginal pH to Week 4 and to Week 8
- Change from baseline in severity of MBS of vaginal dryness to Week 4 and to Week 8

For percentage of parabasal cells, percentage of superficial cells and vaginal pH, same MMRM model as in 9.1.2 will be used. Significance test to compare the ospemifene group with the placebo group will be based on the mean difference between the two groups at Week 4 and at Week 8.

For MBS of vaginal dryness, same GEE model as in 9.1.2 will be used. Significance test to compare ospemifene group with the placebo group will be based on the mean of cumulative log odds ratio (over ordered category responses) at Week 4 and at Week 8.

9.2.2 Severity of VVA Symptoms Other Than Vaginal Dryness

Change from baseline in the following VVA symptoms to Week 4, Week 8, and Week 12.

- Vaginal and/or vulvar irritation or itching
- Difficult or painful urination
- Vaginal pain associated with sexual activity
- Vaginal bleeding associated with sexual activity

Same GEE model as in 9.1.2 will be used. The analysis population will be subjects whose baseline values are moderate or severe in ITT for this analysis.

9.2.3 Change From Baseline in Maturation Value to Week 4, Week 8, and Week 12

Maturation value (MV) is defined as follows:

$MV = (S \times 1) + (I \times 0.5) + (P \times 0)$, where

- MV = Maturation value
- S = % of superficial cells
- I = % of intermediate cells

- P = % of parabasal cells

Analysis of Covariance (ANCOVA) will be used for observed case data for Week 4, Week 8, and Week 12, respectively. Baseline MV will be included as covariate.

9.2.4 Percentage of Subjects Who are Responders at Week 4, Week 8, and Week 12

A subject is defined as a responder if all the following conditions are met:

- MV increase of at least 10 from baseline
- Vaginal pH decrease of at least 0.5 from baseline
- Improvement (decrease in severity) of at least 1 point in the most bothersome symptom of dryness from baseline

Fisher's exact test will be used for observed case data for Week 4, Week 8, and Week 12, respectively.

9.2.5 Change from Baseline in Vaginal Health Index to Weeks 4, 8, and 12

Vaginal Health Index (VHI) consists of 5 items: (1) overall elasticity; (2) fluid secretion; (3) pH; (4) epithelial mucosa; and (5) moisture. Each item consists of 5 categories with scores 1 to 5, the better the each item, the greater the score. Add scores of 5 items to obtain the total score. If any item score is missing, the total score will be missing.

Change from baseline in the total score will be analyzed using ANCOVA for Week 4, Week 8, and Week 12, respectively. Baseline score will be included as covariate.

9.2.6 Change from Baseline in Vulvar Health Index to Weeks 4, 8, and 12

Vulvar Health Index (VuHI) consists of 7 items: (1) labia majora; (2) labia minora; (3) clitoris, (4) introitus and elasticity; (5) color; (6) discomfort and pain; and (7) other findings (petechiae, excoriation, ulceration, etc.). Each item will be assessed on a 4-point scale (0 = normal, 1 = mild, 2 = moderate, and 3 = severe). Add scores of 7 items to obtain the total score. If any item score is missing, the total score will be missing. For (6) discomfort and pain, the severity data will be derived from VVA symptoms "vaginal pain associated with sexual activity", and apply with the 4-point scale score.

Change from baseline in the total score will be analyzed using ANCOVA for Week 4, Week 8, and Week 12, respectively. Baseline score will be included as covariate.

9.2.7 Change from Baseline in Vulvo-vaginal Imaging to Week 12

From vaginal imaging, 9 parameters {Labia Majora, Labia Minora, Clitoris, Urethra, Introitus and Elasticity, Color, Erythema, Moisture, Other Findings (Petechiae, Excoriation, Ulceration, etc.)} will be assessed. These items are evaluated as 4 categories with scores 0 to 3 by 3 independent reviewers. The median value evaluated from 3

reviewers will be used for analyses. Add scores of 9 items to obtain the total score. If any item score is missing, the total score will be missing.

Change from baseline (Randomization) to Week 12 in the vaginal imaging total score will be analyzed using ANCOVA. Baseline score will be included as covariate.

The visit window for valid image is ± 14 days from targeted visit dates.

The data will be collected as optional parameters.

9.2.8 Change from Baseline in Total Score and Each Domain Score of the Female Sexual Function Index Questionnaire to Weeks 4, 8, and 12

Category score of patient's response to each of 19 questionnaires is presented in Appendix 2. The individual domain scores and total score of the Female Sexual Function Index (FSFI) can be derived from the computational formula outlined in the table below. For the domain scores, add the scores of the individual items that comprise the domain and multiply the sum by the domain factor. Add six domain scores to obtain the total score.

Table 4 Female Sexual Function Index Scoring System

Domain	Questions	Score range	Factor	Minimum score	Maximum score
Desire	1,2	1-5	0.6	1.2	6
Arousal	3,4,5,6	0-5	0.3	0	6
Lubrication	7,8,9,10	0-5	0.3	0	6
Orgasm	11,12,13	0-5	0.4	0	6
Satisfaction	14,15,16	0 (or 1) - 5	0.4	0.8	6
Pain	17,18,19	0-5	0.4	0	6
Total Score				2.0	36

If the response to any question within an individual domain is missing, that domain score will be missing. If any domain score is missing, the total score will be missing.

Change from baseline in the total score and the domain scores will be analyzed using ANCOVA for Week 4, Week 8, and Week 12, respectively. Baseline score will be included as covariate.

9.2.9 Change from Baseline in Urinary Symptoms to Weeks 4, 8, and 12

Urinary symptoms will be assessed as absent or present using the Urinary Distress Inventory (UDI-6). The symptoms include: (1) frequent urination; (2) urine leakage related to the feeling of urgency; (3) urine leakage related to physical activity, coughing, or sneezing; (4) small amounts of urine leakage; (5) difficulty emptying bladder; and (6) pain or discomfort in the lower abdominal or genital area. When a symptom is present,

the degree a subject is bothered will be rated on a 4-point scale. In the analysis, the following category score will be used.

0 : the symptom is absent
1 : the symptom is present, however not bothers her at all
2 : the symptom is present and bothers her slightly
3 : the symptom is present and bothers her moderately
4 : the symptom is present and bothers her greatly

Add scores of 6 items to obtain the total score. If any item score is missing, the total score will be missing. If the data is missing for a symptom is present or not but the degree is entered, the individual score and the total score will be calculated.

Change from baseline in the total score will be analyzed using ANCOVA for Week 4, Week 8, and Week 12, respectively. Baseline score will be included as covariate.

9.2.10 Change From Baseline in Markers of Bone Metabolism to Week 12

On bone resorption, the following markers will be measured:

- Serum N-telopeptide of type 1 collagen (NTX)
- Serum C-terminal telopeptide of type 1 collagen (CTX)
- Serum bone sialoprotein (BSP)
- Urinary deoxypyridinoline (DPD)
- Serum tartrate-resistant acid phosphatase 5b (TRACP-5b)

On bone formation, the following markers will be measured:

- Serum total alkaline phosphatase
- Serum bone-specific alkaline phosphatase (BSAP)
- Serum amino-terminal propeptide of type 1 collagen (P1NP)
- Serum osteocalcin

Change from baseline to Week 12 and to Week 12/ET in each marker will be analyzed using ANCOVA. Baseline value will be included as covariate.

9.2.11 Frequency of Lubricant Application and Sexual Activity

These will be recorded daily in an electronic diary (eDiary). The number of days using lubricant and having intercourses will be summarized by using the number of non-missing observations, arithmetic mean, SD, and median by week for each treatment. They will be compared among each treatment group using Welch's t-test. And the mean

values derived from data in each week will be compared using the Welch's t-test. In the summarization, subjects with no record over half days during each week will be excluded.

9.2.12 Overall Satisfaction With Treatment

This will be recorded in an eDiary and summarized for each treatment group by week. The percentage of each category will be compared among each treatment group using Wilcoxon rank-sum test.

9.3 Secondary Endpoints That Will Be Summarized Descriptively Only, With No Inferential Analysis

9.3.1 Serum Hormone Levels

Estradiol (E2), luteinizing hormone (LH), sex hormone-binding globulin (SHBG), free testosterone, and total testosterone values will be summarized at baseline, Week 12 for each treatment group. Change from baseline to Week 12 will also be summarized. Basically, observation at randomization will be used as baseline.

9.4 Summary of Statistical Analyses for Efficacy Endpoints

Refer to Table 5 for a summary of all inferential efficacy analyses that will be performed, including the statistical method, model, population, timepoint, and interpretation.

Table 5 Summary of Inferential Efficacy Analyses

Endpoint	Timepoint	Population	Statistical Method	Model	Interpretation
Change from baseline (Randomization) in % parabasal cells, % superficial cells, vaginal pH	Week 4	ITT	MMRM	Treatment, Week, Treatment*week, Center, Baseline	Secondary analysis
	Week 8	ITT			
	Week12	ITT			Primary analysis
		PPS			Supportive analysis
		mITT			
Change from baseline (Randomization) in severity of MBS of dryness	Week4	ITT	GEE	Treatment, Week, Treatment*week, Center, Baseline	Secondary analysis
	Week 8	ITT			
	Week 12	ITT			Primary analysis
		PPS			Supportive analysis
		mITT			
	Week 4	ITT	MMRM	Treatment, Week, Treatment*week, Center, Baseline	Sensitivity analysis
	Week 8				
	Week 12				
Change from baseline (Randomization) in severity of VVA symptoms other than dryness	Week 4	ITT	MMRM	Treatment, Week, Treatment*week, Center, Baseline	Secondary analysis
	Week 8				
	Week 12				
Change from baseline (Randomization) in MV	Week 4	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
	Week 8	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
	Week 12	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
Proportion of responders	Week 4	ITT	Fisher's exact test	NA	Secondary analysis
	Week 8	ITT	Fisher's exact test	NA	Secondary analysis
	Week 12	ITT	Fisher's exact test	NA	Secondary analysis
Change from baseline (Randomization) in VHI (total score and each item score)	Week 4	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
	Week 8	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
	Week 12	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
Change from baseline (Randomization) in VuHI (total score and each item score)	Week 4	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
	Week 8	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
	Week 12	ITT	ANCOVA	Treatment, Baseline	Secondary analysis

Endpoint	Timepoint	Population	Statistical Method	Model	Interpretation
Change from baseline (Randomization) in vulvo-vaginal imaging	Week 12	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
Change form baseline (Randomization) in FSFI (total score and each domain score)	Week 4	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
	Week 8	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
	Week 12	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
Change from baseline (Randomization) in UDI-6	Week 4	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
	Week 8	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
	Week 12	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
Change from baseline (Randomization) in bone marker	Week 12	ITT	ANCOVA	Treatment, Baseline	Secondary analysis
Frequency of lubricant application and sexual activity	Weekly	ITT	Welch's t-test	---	Secondary analysis
	Overall	ITT	MMRM	Treatment, Week, Treatment*week	Secondary analysis
Overall satisfaction with treatment	Weekly	ITT	Wilcoxon rank-sum test	---	Secondary analysis

10 SAFETY

Safety analyses will be conducted on the Safety Population. For all safety analyses, emphasis will be on characterizing the changes from baseline and flagging clinically relevant abnormal findings, if feasible. If a test or assessment is measured more than once prior to the first dose of study medication, the baseline measurement is the last non-missing value that is observed prior to the first dose of study medication.

To calculate 95% confidence intervals for differences in incidence proportions, FREQ procedure and RISKDIFF option will be used.

10.1 Adverse Events

All adverse events (AEs) will be collected from the time of informed consent through 14 days after the last dose of study medication. All AEs will be classified by SOC and PT using MedDRA Version 18.0 and presented in subject listings.

AEs that occurred from the first dose of study medication up to 14 days after the last dose will be considered treatment-emergent AEs (TEAEs). Only TEAEs will be counted in the summary tables described below. If more than one AEs that map to the same SOC or the same PT are occurred in a subject, the subject will be counted only once at each level of summarization (i.e., SOC level or PT level) using AE with the strongest outcome and/or severity.

If partial dates of onset are missing, the following imputation algorism will be applied;

- If the information of period of onset is 'After first dose of study drug' and the month of onset is same as the first dosing date but day is unknown, the day will be same as the first dosing date.
- If the information of period of onset is 'After first dose of study drug' and the month of onset is different from the first dosing date but day is unknown, the day will be first day of the month.

All AEs will be listed. TEAEs will be summarized as assessments of short term safety and long term safety, separately. For assessment of short term safety, following TEAEs will be included;

- if a subject takes study medication for more than 12 weeks (beyond Visit 5), AEs that occurred until Visit 5 will be summarized;
- if a subject takes study medication for 12 weeks or less than 12 weeks, all TEAEs will be summarized.

For assessment of long term safety, the target safety population will be limited to subjects who continue to Part B in two previous protocols (original protocol:06Oct2012 and protocol amendment 1: 11Mar2016) and take study drugs more than 12 weeks (beyond Visit 5) and TEAEs after 12 weeks (beyond Visit 5) will be summarized.

The following summary tables of TEAEs will be produced:

1) Overview of Adverse Events

For each treatment group, the number of events and subjects who experience

- TEAEs
- Serious TEAEs
- TEAEs leading to withdrawal
- TEAEs with an outcome of deaths

will be counted and the incidence proportion will be calculated for each treatment group. The difference in the incidence proportion and its 95% confidence interval (CI) will also be calculated. Treatment-related AEs will be summarized by the same category as TEAEs, except for death.

2) Summary of TEAEs and Treatment-related AEs by System Organ Class and Preferred Term

For each treatment group, the number and percentage of subjects who experience TEAE (and Treatment-related AEs) will be counted by SOC and PT for assessments of short and long term safety. The difference in the percentage and its 95% CI will also be calculated. For assessment of short term safety, This analysis will be performed for specific TEAEs for assessment of short term safety, namely

- Breast-related TEAE
- Reproductive system-related TEAE
- Lipid-related TEAE
- Coagulation-related TEAE
- Chemistry-related TEAE

TEAEs and Treatment-related AEs at least 2% in any treatment groups will be analyzed in a similar way.

3) Summary of Serious TEAEs and TEAEs Leading to Withdrawal by System Organ Class and Preferred Term

For each treatment group, the number and percentage of subjects who experience serious TEAE and TEAE leading to withdrawal will be counted by SOC and PT for assessment of short term safety. The difference in the percentage and its 95% CI will also be calculated.

4) Summary of TEAEs and Treatment-related AEs by System Organ Class by Maximum Severity and Outcome

For each treatment group, the number and percentage of subjects who experience TEAE (and treatment-related AEs) will be counted by SOC and PT, and by maximum

severity and outcome for assessment of short term safety. The order of severity and outcome are shown in Table 6.

Table 6 Order of Priority for Counting Number of Subjects

Order of Priority	Severity	Outcome
1	Severe	Fatal
2	Moderate	Recovered/Resolved with sequelae
3	Mild	Not recovered/Not resolved
4		Recovering/Resolving
5		Recovered/Resolved
6		Unknown

10.2 Vital Signs

Summary statistics of observations (blood pressure, pulse rate, and oral temperature) and its change from baseline to each scheduled visit (Weeks 4, 8, and 12) will be calculated by treatment group.

10.3 Laboratory Evaluations

Safety laboratory tests shown in Table 7 will be conducted at screening, randomization, Week 12, and early termination before Visit 5 (hereafter ET).

Table 7 Safety Laboratory Tests

Category	Evaluation Items
Hematology Tests	hematocrit, hemoglobin, platelet count, red blood cell (RBC) count, white blood cell (WBC) count with differential (eosinophil count, basophil count, neutrophil count, monocyte count, lymphocyte count), platelet count, mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), RBC distribution width (RDW), mean platelet volume (MPV)
Blood Chemistry Tests	ALP, ALT, AST, total bilirubin, GGT, LDH, BUN, creatinine, uric acid, sodium (Na), potassium (K), chloride (Cl), calcium (Ca), glucose, total protein
Lipids	Total cholesterol, low-density lipoprotein (LDL) cholesterol, very-low-density lipoprotein (VLDL) cholesterol, high-density lipoprotein (HDL) cholesterol, and triglycerides
Coagulation	Thromboplastin time, fibrinogen, antithrombin III, , protein-C, and protein-S
Urinalysis for Quantitative	pH, specific gravity
Urinalysis for Qualitative	Color, bilirubin, glucose, ketones, nitrite, occult blood, protein, microscopic evaluation of sediment, and urobilinogen

ALT = alanine aminotransferase; AST = aspartate aminotransferase; ALP = alkaline phosphatase; GGT = gamma-glutamyl transferase; LDH, lactate dehydrogenase; BUN, blood urea nitrogen

The measurements on safety laboratory evaluations will be handled as follows:

1) For quantitative scale items, summary statistics of measurements taken at baseline, Week 12 will be calculated by treatment group. The change from baseline to Week 12 will be also calculated by treatment group.
2) For qualitative scale items, frequencies for each category at baseline, Week 12 will be counted by treatment group.
3) For all laboratory analytes in Appendix 3, proportion of subjects who meet the predefined limits at any timepoints will be calculated. Difference in the incidence proportion among treatments and its 95% CI will be calculated. This analysis will be performed using data until Visit 5. The analysis will be applied to data taken at the last timepoint before stopping dosing until Visit 5.

The factor V Leiden will not be summarized.

10.4 Endometrial Safety

For subjects with intact uterus, summary tables of endometrial histology and endometrial thickness will be presented at baseline, Week 12 for each treatment group. Change from baseline to Week 12 in endometrial thickness will be summarized for each treatment group.

10.5 Other Safety Variables

Summary table of BMI will be presented at baseline, Weeks 4, 8, 12 for each treatment group. Change from baseline to each timepoint will also be summarized by treatment group.

11 INTERIM ANALYSES

No interim analysis is planned in this study.

12 PROGRAMMING CONVENTIONS

Unless otherwise stated, tables, figures, and listings (TFLs) will be made in accordance with the following specifications on programming:

1) Display digit of various statistics

The display digit of various tests (such as laboratory tests) is as specified for each respective test, in principle. Display digit of statistics for efficacy and safety analyses are defined in Table 8.

Table 8 Display Digit of Statistics for Efficacy and Safety

	Statistics	Display Digit
Efficacy and Safety	The number of subjects	Displayed as integer
	Mean, median, 95% confidence interval	One more decimal place than raw data
	Standard deviation, standard error	One more decimal place than raw data
	Maximum, minimum	Same number of decimal places as raw data
	Percentage (%)	Round off to one decimal place
	p-value	Round off to four decimal places Note: $p < 0.0001$ is displayed as "< .0001"
VVA symptoms	Mean, median, 95% confidence interval, standard deviation, standard error	Round off to two decimal place

2) Handling of outliers

Possible outliers will not be omitted from the analyses.

3) Categories used in the summarization

Categories used in the summarization of items are shown in Table 9. The breakdown of categories may be changed with blind inspection of distribution as needed.

Table 9 Subject Characteristics and Items of Therapeutic Process

Items	Categories
Age (years)	< 45, >= 45 to < 55, >=55 to < 65, >= 65
BMI (kg/m^2)	< 18.5, >=18.5 to < 25, >=25 to < 30, >=30
Ethnicity	Not Hispanic or Latino, Hispanic or Latino, Not reported, Unknown
Race	American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Other
Uterus	Yes, No
Current hot flushes	Yes, No
Previous hormone treatment	None, any hormone treatment, hormone replacement therapy (excluding vaginal),vaginal hormone products
Number of Days of Treatment Exposure within Week 12 (days)	>=1 to <= 7, >=8 to <= 14, >=15 to <= 28, >=29 to <= 42, >=43 to <= 56, >=57 to <= 70, >=71 to <= 84, >= 85
Number of Days of Treatment Exposure after Week 12 (days)	>=1 to <=14, >=15 to <=28, >=29 to <=56, >=57 to <=84, >=85 to <=168, >=169 to <=273, >=274 to <=364, >=365
Medication compliance (%)	< 85, >= 85

13 CHANGES FROM PROTOCOL

13.1 Safety Analysis

- Calculation of the CI of the incidence proportion for each treatment group was planned using the Clopper-Pearson method in the Section 9.11.1. It was replaced with the CIs for the difference between treatments because these are more informative.

14 REFERENCES

1. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. Second edition. John Wiley & Sons; 2011.

Appendix 1 Schedule of Time and Events

Evaluation	Screening	Randomization	Treatment			FU[q]
	V1	V2	V3	V4	V5	TC
	Within 4 Weeks Prior to Day 1 [a]	Day 1	Wk 4	Wk 8	Wk 12/ ET	Wk 14
Informed Consent	X					
Assign Subject ID Number [b]	X					
I/E Criteria	X	X				
Demographics	X					
Medical History [c]	X					
Physical Examination	X	X			X	
Electrocardiogram	X					
Breast Examination	X	X			X	
Gynecological Examination	X[o]	X[p]			X	
Safety Laboratory Tests [d]	X	X[e, p]			X[e]	
Hormone Levels [f]		X			X	
FSH [g]	X					
Markers of Bone Metabolism [h]		X			X	
Urinalysis	X	X			X	
Urine Dipstick [i]	X	X				
Vital Signs (blood pressure, pulse, body temperature)	X	X	X	X	X	
Height	X					
Weight (BMI)	X	X	X	X	X	
Prior Therapy	X	X				
Concomitant Therapy			X	X	X	X
Adverse Events	X[j]	X	X	X	X	X

Evaluation	Screening	Randomization	Treatment			FU[q]
	V1	V2	V3	V4	V5	TC
	Within 4 Weeks Prior to Day 1 [a]	Day 1	Wk 4	Wk 8	Wk 12/ ET	Wk 14
Cervical PAP Smear	X[o]					
Vaginal pH	X[o]	X[p]	X	X	X	
Vaginal Smear/ Maturation Index	X[o]	X[p]	X	X	X	
Vaginal Health Index (VHI)		X	X	X	X	
Vulvar Health Index (VuHI)		X	X	X	X	
Vulvo-vaginal Imaging [k]		X			X	
Endometrial Thickness (TVU)	X				X	
Endometrial Histology (Biopsy)	X				X[n]	
Mammogram	X[l]					
Dispense eDiary/Return eDiary		X			X	
eDiary Compliance			X	X	X	
Assessment of symptoms of VVA	X[o]	X[p]	X	X	X	
Dispense Study Medication		X	X	X		
Dispense Lubricant		X	X	X		
Study Medication Compliance [m]			X	X	X	
Reinforce Compliance with Study Medication		X	X	X		
Study Medication Return			X	X	X	
Female Sexual Function Index (FSFI)		X	X	X	X	
Urinary Distress Inventory (UDI-6)		X	X	X	X	

ET = early termination; FU = follow-up; TC = telephone call

a Enough time between Visit 1 and Visit 2 should be allowed to receive the results of tests conducted as part of Visit 1; therefore, Visit 2 should not take place earlier than 1 week after Visit 1.

b Subject ID number will be assigned through the interactive web or voice response system (IxRS).

c Medical history should be captured for the 2 years prior to screening, except for any history of cancer which should be captured regardless of the time.

d Safety laboratory tests: red blood cell (RBC) count, white blood cell (WBC) count, differential, platelet count, hemoglobin (HgB), hematocrit (HCT), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), RBC distribution width (RDW), mean platelet volume (MPV), albumin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin, creatinine, estimated glomerular filtration rate (eGFR), total protein, glucose, uric acid, blood urea nitrogen (BUN), creatine kinase (CK), thromboplastin time, fibrinogen, antithrombin III, factor V Leiden (screening only), protein-C, protein-S, lipids total cholesterol, LDL cholesterol (LDL-C), HDL cholesterol (HDL-C) and triglycerides.

e Subjects should be fasting for at least 8 hours prior to blood sample collection.

f Hormones: Estradiol (E2), follicle-stimulating hormone (FSH), luteinizing hormone (LH), sex hormone-binding globulin (SHBG), free and total testosterone.

g FSH values may be measured for confirmation of postmenopausal status (inclusion criterion No. 3) in women ≥45 years old who do not remember the date of the last spontaneous menstrual period OR women who have had a hysterectomy WITHOUT oophorectomy.

h Markers of bone metabolism: Markers of bone resorption: serum N-terminal telopeptide of type 1 collagen (NTX), serum C-terminal telopeptide of type 1 collagen (CTX), serum bone sialoprotein (BSP), urinary deoxypyridinoline (DPD), and serum tartrate-resistant acid phosphatase 5b (TRACP-5b). Markers of bone formation: serum total alkaline phosphatase, serum bone-specific alkaline phosphatase (BSAP), serum N-terminal propeptide of type 1 collagen (P1NP), and serum osteocalcin.

i At Visits 1 and 2, in addition to obtaining a sample for urinalysis, a urine dipstick will be used to rule out urinary tract infections prior to randomization. Subjects otherwise eligible, who are found to have a urinary tract infection should receive treatment and may continue after confirmation of cure.

j Adverse events spontaneously reported after signature of the informed consent should be captured.

k Unless subject withdraws consent for vaginal imaging during the course of the study, if the quality of the photographs is not adequate as assessed by Canfield, a re-shoot should be scheduled as soon as possible but no more than seven days after the original shoot day.

l May be omitted in subjects with a normal mammogram within 9 months prior to screening for which the results are available.

m To be assessed based on pill count. When there is a discrepancy between the pill count and the subject-reported compliance, the subject-reported compliance should be used.

n Do not perform if discontinuation (early termination) occurs PRIOR to VISIT 5 (Week 12).

o Repeat after cure of infection at V1.

p Repeat after cure of infection at V2.

q Follow-up call should occur ON Day 14 after last dose. In case subject does not respond, three follow-up attempts should be made.

Appendix 2 Category Score of Patient's Response to Female Sexual Function Index (FSFI)

Desire

1. Over the past 4 weeks, how **often** did you feel sexual desire or interest?
   - **5** Almost always or always
   - **4** Most times (more than half the time)
   - **3** Sometimes (about half the time)
   - **2** A few times (less than half the time)
   - **1** Almost never or never

2. Over the past 4 weeks, how would you rate your **level** (degree) of sexual desire or interest?
   - **5** Very high
   - **4** High
   - **3** Moderate
   - **2** Low
   - **1** Very low or none at all

Arousal

3. Over the past 4 weeks, how **often** did you feel sexually aroused ("turned on") during sexual activity or intercourse?
   - **0** No sexual activity
   - **5** Almost always or always
   - **4** Most times (more than half the time)
   - **3** Sometimes (about half the time)
   - **2** A few times (less than half the time)
   - **1** Almost never or never

4. Over the past 4 weeks, how would you rate your **level** of sexual arousal ("turn on") during sexual activity or intercourse?
   - **0** No sexual activity
   - **5** Very high
   - **4** High
   - **3** Moderate
   - **2** Low
   - **1** Very low or none at all

5. Over the past 4 weeks, how **confident** were you about becoming sexually aroused during sexual activity or intercourse?
   - **0** No sexual activity
   - **5** Very high confidence
   - **4** High confidence
   - **3** Moderate confidence

2 Low confidence
1 Very low or no confidence

6. Over the past 4 weeks, how **often** have you been satisfied with your arousal (excitement) during sexual activity or intercourse?
   0 No sexual activity
   5 Almost always or always
   4 Most times (more than half the time)
   3 Sometimes (about half the time)
   2 A few times (less than half the time)
   1 Almost never or never

Lubrication

7. Over the past 4 weeks, how **often** did you become lubricated ("wet") during sexual activity or intercourse?
   0 No sexual activity
   5 Almost always or always
   4 Most times (more than half the time)
   3 Sometimes (about half the time)
   2 A few times (less than half the time)
   1 Almost never or never

8. Over the past 4 weeks, how **difficult** was it to become lubricated ("wet") during sexual activity or intercourse?
   0 No sexual activity
   1 Extremely difficult or impossible
   2 Very difficult
   3 Difficult
   4 Slightly difficult
   5 Not difficult

9. Over the past 4 weeks, how often did you **maintain** your lubrication ("wetness") until completion of sexual activity or intercourse?
   0 No sexual activity
   5 Almost always or always
   4 Most times (more than half the time)
   3 Sometimes (about half the time)
   2 A few times (less than half the time)
   1 Almost never or never

10. Over the past 4 weeks, how **difficult** was it to maintain your lubrication ("wetness") until completion of sexual activity or intercourse?
    0 No sexual activity
    1 Extremely difficult or impossible
    2 Very difficult

3 Difficult
4 Slightly difficult
5 Not difficult

Orgasm

11. Over the past 4 weeks, when you had sexual stimulation or intercourse, how **often** did you reach orgasm (climax)?
    0 No sexual activity
    5 Almost always or always
    4 Most times (more than half the time)
    3 Sometimes (about half the time)
    2 A few times (less than half the time)
    1 Almost never or never

12. Over the past 4 weeks, when you had sexual stimulation or intercourse, how **difficult** was it for you to reach orgasm (climax)?
    0 No sexual activity
    1 Extremely difficult or impossible
    2 Very difficult
    3 Difficult
    4 Slightly difficult
    5 Not difficult

13. Over the past 4 weeks, how **satisfied** were you with your ability to reach orgasm (climax) during sexual activity or intercourse?
    0 No sexual activity
    5 Very satisfied
    4 Moderately satisfied
    3 About equally satisfied and dissatisfied
    2 Moderately dissatisfied
    1 Very dissatisfied

Satisfaction

14. Over the past 4 weeks, how **satisfied** have you been with the amount of emotional closeness during sexual activity between you and your partner?
    0 No sexual activity
    5 Very satisfied
    4 Moderately satisfied
    3 About equally satisfied and dissatisfied
    2 Moderately dissatisfied
    1 Very dissatisfied

15. Over the past 4 weeks, how **satisfied** have you been with your sexual relationship with your partner?

5 Very satisfied
4 Moderately satisfied
3 About equally satisfied and dissatisfied
2 Moderately dissatisfied
1 Very dissatisfied

16. Over the past 4 weeks, how **satisfied** have you been with your overall sexual life?
5 Very satisfied
4 Moderately satisfied
3 About equally satisfied and dissatisfied
2 Moderately dissatisfied
1 Very dissatisfied

Pain

17. Over the past 4 weeks, how **often** did you experience discomfort or pain during vaginal penetration?
0 Did not attempt intercourse
1 Almost always or always
2 Most times (more than half the time)
3 Sometimes (about half the time)
4 A few times (less than half the time)
5 Almost never or never

18. Over the past 4 weeks, how **often** did you experience discomfort or pain following vaginal penetration?
0 Did not attempt intercourse
1 Almost always or always
2 Most times (more than half the time)
3 Sometimes (about half the time)
4 A few times (less than half the time)
5 Almost never or never

19. Over the past 4 weeks, how would you rate your **level** (degree) of discomfort or pain during or following vaginal penetration?
0 Did not attempt intercourse
1 Very high
2 High
3 Moderate
4 Low
5 Very low or none at all

Appendix 3 Pre-defined Limits of Change for Laboratory Test

Test Name (Unit)	Criteria
Hemoglobin (g/dL)	Dcrease from baseline ≥ 1.5 g/dL
WBC Count (10^3/μL)	Decrease from baseline ≥ 50% and value < LLN
	Increase from baseline ≥ 20% and value > ULN
Neutrophil Count (10^3/μL)	Decrease from baseline ≥ 20% and value < LLN
	Increase from baseline ≥ 20% and value > ULN
Lymphocyte Count (10^3/μL)	Decrease from baseline ≥ 20% and value < LLN
	Increase from baseline ≥ 20% and value > ULN
Platelet Count (10^3/μL)	Decrease from baseline ≥ 25% and value < LLN
	Increase from baseline ≥ 100% and value > ULN
BUN (mg/dL)	Increase from baseline ≥ 50% and value > ULN
Creatinine (mg/dL)	Increase from baseline ≥ 0.3 mg/dL
Total Bilirubin (mg/dL)	Increase from baseline ≥ 50% and value > ULN
AST (U/L)	Value > 3 x ULN
	Value > 5 x ULN
	Value > 10 x ULN
	Value > 20 x ULN
ALT (U/L)	Value > 3 x ULN
	Value > 5 x ULN
	Value > 10 x ULN
	Value > 20 x ULN
AST (U/L) or ALT (U/L)	Value > 3 x ULN
	Value > 5 x ULN
	Value > 10 x ULN
	Value > 20 x ULN
AST (U/L) or ALT (U/L) + Total Bilirubin (mg/dL)	(AST > 3 x ULN or ALT > 3 x ULN) + Total Bilirubin > 2 x ULN simultaneously
Alkaline Phosphatase (U/L)	Increase from baseline ≥ 50% and value > ULN
Creatine Kinase (CK) (U/L)	Value > 5 x ULN
Potassium (mmol/L)	Value > ULN and > Baseline
	Value < LLN and < Baseline
Sodium (mmol/L)	Value > ULN and > Baseline
	Value < LLN and < Baseline
Calcium (mg/dL)	Value > ULN and > Baseline
	Value < LLN and < Baseline

ALT = alanine aminotransferase; AST = aspartate aminotransferase; ALP = alkaline phosphatase; BUN, blood urea nitrogen; GGT = gamma-glutamyl transferase; LDH = lactate dehydrogenase; WBC = white blood cell

Appendix 4 Pooled Site

Pooled Site Number	Site ID
1	5EX, 5AA, 5FA, 5AE, 5BB, 5BS, 5CM
2	5CJ, 5FB, 5CG, 5BM, 5BQ, 5EG, 5EN, 5EZ, 5EK
3	5EY, 5FT, 5AP, 5BG, 5BC
4	5AY, 5FJ, 5AH, 5CK
5	5FX, 5EC, 5FZ, 5FY, 5FC, 5AQ
6	5FP, 5BF, 5CY
7	5BJ, 5FG, 5BK, 5FW, 5AM, 5AC, 5AL
8	5FS, 5CW, 5BR, 5BL, 5EH, 5ER, 5BA, 5EJ, 5FH, 5CZ, 5ET
9	5CT, 5AB, 5FK, 5CR, 5BE, 5EF, 5CS, 5ES
10	5EQ, 5AK, 5CA, 5CE, 5BX, 5AG, 5CB, 5CC