Statistical Analysis Plan for Study M14-702

A Phase 3 Study to Evaluate the Safety and Efficacy of Elagolix in Combination with Estradiol/Norethindrone Acetate in Subjects with Moderate to Severe Endometriosis-Associated Pain

Date: 08 May 2020

Version 2.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for Study M14-702 "A Phase 3 Study to Evaluate the Safety and Efficacy of Elagolix in Combination with Estradiol/Norethindrone Acetate in Subjects with Moderate to Severe Endometriosis-Associated Pain."

Study M14-702 examines the efficacy and safety of elagolix 200 mg administered twice daily (BID) in combination with estradiol 1 mg/0.5 mg norethindrone acetate (E2/NETA 1 mg/0.5 mg) in subjects with moderate to severe endometriosis associated pain.

The analyses of pharmacokinetic endpoints, pharmacodynamic biomarker endpoints and exploratory research endpoints will not be covered in this SAP.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

Unless noted otherwise, all analyses will be performed using SAS Version 9.4 (SAS Institute Inc., Cary, NC 27513) or later under the UNIX operating system.

2.0 Study Design, Objectives and Procedures

2.1 Objectives

The objectives of this study are to:

- Assess the safety and efficacy of elagolix 200 mg administered twice daily (BID) in combination with estradiol 1 mg/0.5 mg norethindrone acetate (E2/NETA 1 mg/0.5 mg) QD compared to placebo at 6 months and 12 months;
- Assess the effect of elagolix 200 mg BID in combination with E2/NETA (1 mg/0.5 mg) QD on bone mineral density (BMD) compared to elagolix 200 mg BID alone at 6 months and compared to placebo at 6 months and 12 months;

- Evaluate the continued safety and efficacy of elagolix 200 mg BID plus E2/NETA (1 mg/0.5 mg) QD for up to 48 months in premenopausal women with moderate to severe endometriosis-associated pain.

2.2 Study Design Overview

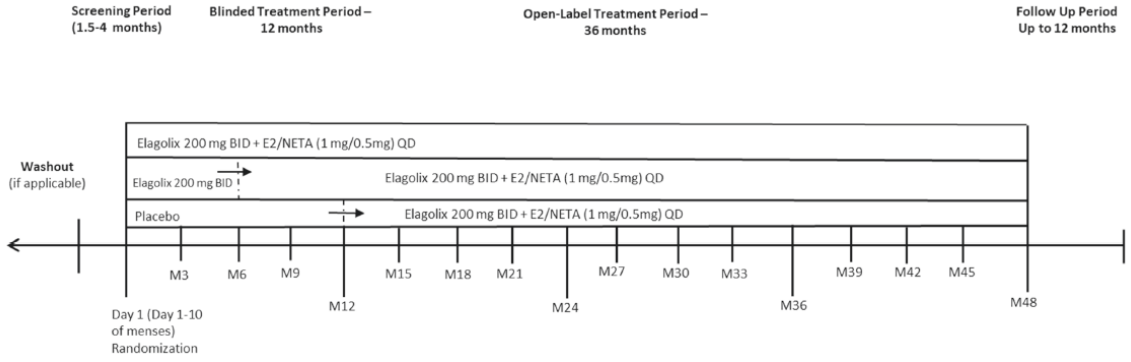
This Phase 3 study includes a 48-month Treatment Period and is designed to evaluate the safety and efficacy of elagolix in combination with concomitant hormonal add-back therapy (E2/NETA 1 mg/0.5 mg) in the management of endometriosis-associated pain in premenopausal women. The first 12 months of the Treatment Period will employ a randomized, double-blind, placebo-controlled design (12 months), with an elagolix 200 mg BID alone arm (the first 6 months followed by elagolix 200 mg BID plus E2/NETA [1 mg/0.5 mg] QD), and an elagolix 200 mg BID plus E2/NETA (1 mg/0.5 mg) QD arm. The last 36 months of the Treatment Period will be open-label, such that all subjects will receive elagolix 200 mg BID plus E2/NETA (1 mg/0.5 mg) QD.

The total duration for a subject's participation in this study is approximately 51 to 74 months, consisting of the following 4 study periods:

- Washout Period – up to 10 months prior to Screening (only applicable if subject is taking exclusionary medication at the time of consent; duration of washout depends on the type of exclusionary medication being taken).
- Screening Period – approximately 1.5 to 4 months prior to first dose of study drug.
- Treatment Period – up to 48-month treatment duration.
- Follow-Up Period – up to 12 months duration following the last dose of study drug. Subjects are expected to enter Follow-Up after completing Treatment Month 48, or if a subject prematurely discontinues from the Treatment Period at the time of or after Treatment Month 6.

The schematic of the study is shown in Figure 1.

Figure 1. Study Schematic



2.3 Treatment Assignment and Blinding

Approximately 700 subjects will be randomly assigned on Study Day 1 in a 4:1:2 ratio as follows:

- Elagolix plus E2/NETA group: elagolix 200 mg BID plus E2/NETA (1 mg/0.5 mg) QD (n = 400)
- Elagolix alone group: elagolix 200 mg BID (n = 100)
- Placebo group: placebo (n = 200)

Unless otherwise specified, data will be summarized by group as listed above.

2.4 Sample Size Determination

This study plans to enroll approximately 700 subjects, with 400 subjects in the elagolix 200 mg BID plus E2/NETA (1 mg/0.5 mg) QD, 100 subjects in the elagolix 200 mg BID, and 200 subjects in placebo, representing a randomization ratio of 4:1:2. These sample sizes were chosen to ensure a sufficient number of subjects completing 12 months of total treatment with elagolix during the placebo-controlled treatment period for the evaluation of safety assessment. Assuming Non-Menstrual Pelvic Pain (NMPP) responder rates of 54% for the elagolix with E2/NETA therapy dose group and 35% for placebo, and

Dysmenorrhea (DYS) responder rates of 59% for the elagolix with E2/NETA therapy dose group and 23% for placebo, these sample sizes provide greater than 90% power to detect a difference in response rate between elagolix with E2/NETA and placebo for both DYS and NMPP, based on a 2-sided test at the significance level of $\alpha = 0.05$.

2.5 End of 12-Month Placebo-Controlled Treatment Period Analysis

An end-of-placebo-controlled treatment period analysis of the primary, secondary and other efficacy variables along with demographic and safety variables will be performed after the last subject completes the 12-Month placebo-controlled Treatment Period. These analyses will include data collected during the Treatment Period (placebo-controlled and open-label Treatment Periods) and the Follow-Up Period up to the data cut-off for the interim lock. The database will be versioned for an interim soft lock and any discrepant data will be clarified before the versioning. Analyses will be performed by AbbVie.

Since this end-of-placebo-controlled treatment period analysis is the only and final analysis of the primary and key secondary endpoints, no additional adjustment of alpha-level is necessary.

2.6 Treatment Month 24 and Month 36 Analyses

At the end of Treatment Month 24 and Month 36, analysis of the efficacy and safety variables may be performed after the last subject completes the Treatment Month 24 and Month 36 visit, respectively. For subjects who prematurely discontinued during the treatment period, their data collected during the Follow-up Period will be included in the analysis, as appropriate. The database will be versioned for an interim lock and any discrepant data will be clarified before the versioning. Analyses will be performed by AbbVie.

3.0 Endpoints

3.1 Primary Endpoint(s)

The co-primary efficacy endpoints (elagolix plus E2/NETA versus placebo) will be the proportion of responders (defined in Section 3.3.1) at Month 6 based upon the mutually-exclusive scales for daily assessment of Dysmenorrhea (DYS) and Non-Menstrual Pelvic Pain (NMPP) measured by the 4-point Endometriosis Daily Pain Impact Diary. Use of protocol specific rescue analgesic medication for endometriosis-associated pain will also be included in the responder definition.

3.2 Secondary Endpoint(s)

The ranked secondary efficacy endpoints (elagolix plus E2/NETA versus placebo) will be tested in the following order:

1. Change from Baseline to Month 12 in DYS
2. Change from Baseline to Month 6 in DYS
3. Change from Baseline to Month 3 in DYS
4. Change from Baseline to Month 12 in NMPP
5. Change from Baseline to Month 6 in NMPP
6. Change from Baseline to Month 3 in NMPP
7. Change from Baseline to Month 6 in PROMIS Fatigue Short Form 6a T-Score
8. Change from Baseline to Month 12 in dyspareunia
9. Change from Baseline to Month 6 in dyspareunia
10. Change from Baseline to Month 3 in dyspareunia
11. Change from Baseline to Month 12 in PROMIS Fatigue Short Form 6a T-Score
12. Change from Baseline to Month 12 in Overall Endometriosis-Associated Pain
13. Change from Baseline to Month 6 in Overall Endometriosis-Associated Pain

14. Change from Baseline to Month 3 in Overall Endometriosis-Associated Pain

3.3 Other Efficacy Endpoint(s)

The primary and secondary efficacy endpoints are listed in Section 3.1 and Section 3.2, respectively. The additional efficacy endpoints (elagolix plus E2/NETA versus placebo) are:

- Proportion of responders as assessed by change from Baseline in average pain score and rescue analgesic use monthly during the 12-month Placebo-Controlled Treatment Period for DYS, NMPP and dyspareunia, respectively.
- Proportion of responders as assessed by change from Baseline in average pain score alone (i.e., not taking into consideration of rescue analgesic use) monthly during the 12-month Placebo-Controlled Treatment Period for DYS, NMPP and dyspareunia, respectively.
- Change from Baseline in the monthly average pain score (DYS, NMPP, and dyspareunia), as assessed using information from the e-Diary during the 12-month Placebo-Controlled Treatment Period.
- Change from Baseline in rescue analgesic use (average pill counts) to treat endometriosis-related pain, as assessed using information from the e-Diary monthly during the first 12 months of the Treatment Period.
- Proportion of rescue analgesic use responders monthly (based only on reduction of rescue analgesics used) during the first 12 months of the Treatment Period.
- Monthly response in the PGIC and overall endometriosis-associated pain questionnaire during the first 12 months of the Treatment Period.
- Change from baseline for each of six domains of EHP-30 questionnaire scores at each scheduled assessment during the Treatment Period.
- Change from baseline for each category of the EuroQoL-5D (EQ-5D-5L) and index values at each scheduled assessment during the Treatment Period.
- Change from baseline for each measure of the WPAI:SHP at each scheduled assessment during the Treatment Period.

- HCRU at each scheduled assessment during the Treatment Period.
- Change from baseline for the PROMIS Fatigue short Form 6A T-Score at each scheduled assessment during the Treatment Period.

3.4 Safety Endpoint(s)

The safety endpoints will be based on the following evaluations:

- Adverse events Monitoring
- Clinical laboratory tests
- Dual energy X-ray absorptiometry (DXA) scan
- Vital signs
- Endometrial biopsy
- Pelvic ultrasound (transabdominal ultrasound [TAU]/transvaginal ultrasound [TVU])
- Columbia Suicide Severity Rating Scale (C-SSRS)
- Pregnancy
- Uterine bleeding measured by e-Diary

4.0 Analysis Populations

The Full Analysis Set (FAS) is composed of all randomized subjects who have received at least 1 dose of study drug in this study. The data from the FAS will be presented by the treatment group assigned at the time of randomization, even if a subject does not receive the correct treatment or does not follow the protocol until completion. The FAS will be used for all efficacy and baseline analyses unless otherwise specified in the SAP. Subjects who do not continue into the open-label Treatment Period will be excluded from the analyses in the open-label Treatment Period.

The following populations will be used for safety analyses:

- The Safety Analysis Set includes all randomized subjects who received at least 1 dose of study drug. The data from the Safety Analysis Set will be analyzed based on the actual treatment received at the time of randomization (elagolix plus E2/NETA, elagolix alone or placebo). Subjects who do not continue into the open-label Treatment Period will be excluded from the analyses in the open-label Treatment Period.
- The All Elagolix plus E2/NETA Analysis Set is defined as subjects who received at least 1 dose of elagolix plus E2/NETA in the study (either in the placebo-controlled or open-label Treatment Periods). Subjects will be summarized based on the actual treatment received at the time of randomization (elagolix plus E2/NETA, elagolix alone or placebo).

5.0 Subject Disposition

A summary of subject accountability will be provided where the number of subjects in each of the following categories will be summarized for all randomized subjectsby treatment group and overall:

- Subjects randomized in the study.
- Subjects who took at least one dose of study drug.
- Subjects who completed the placebo-controlled Treatment Period.
- Subjects who prematurely discontinued during the placebo-controlled Treatment Period.
- Subjects who entered into the open-label Treatment Period.
- Subjects who prematurely discontinued during the open-label Treatment Period.
- Subjects who completed protocol-specified treatment.
- Subjects who prematurely discontinued study drug (all reasons and primary reason).
- Subjects who entered into the Follow-up Period.

- Subjects who Completed the Follow-up Period.

The number and percentage of subjects who discontinued study drug will be summarized by treatment group and overall as follows based on the Full Analysis Set:

- By any reason for discontinuation
- By primary reason for discontinuation

In addition, the number and percentage of subjects who discontinued study will be summarized for the FAS by treatment group and overall as follows:

- By any reason for discontinuation
- By primary reason for discontinuation

6.0 Study Drug Duration

Duration (days) of treatment will be summarized for the Safety Analysis Set using the number of subjects treated, mean, standard deviation, median, minimum and maximum.

Study drug duration will be summarized as follows:

Placebo-controlled Treatment Period (Safety Analysis Set):

Duration of treatment during the placebo-controlled Treatment Period is defined as follow:

- For subjects who did not continue into the open-label Treatment Period:
  - Last dose date during the placebo-controlled Treatment Period – First dose date during the placebo-controlled Treatment Period + 1
- For subjects who continued into the open-label Treatment Period:
  Minimum of
  - Last dose date during the placebo-controlled Treatment Period – First dose date during the placebo-controlled Treatment Period + 1

- First dose date during the open-label Treatment Period – First dose date during the placebo-controlled Treatment Period.

In addition, the number and percentage of subjects in each treatment duration interval (1 – 28 days, > 28 – 56 days, > 56 – 84 days, > 84 – 112 days, > 112 – 140 days, > 140 – 168 days, > 168 days) will be summarized.

Open-label Treatment Period (Safety Analysis Set):

Duration of treatment during the open-label Treatment Period is defined as last dose date during the open-label Treatment Period minus first dose date during the open-label Treatment Period plus 1 day.

In addition, the number and percentage of subjects in each treatment duration interval (1 – 84 days, > 84 – 168 days, > 168 – 252 days, > 252 – 336 days, > 336 – 420 days, > 420 – 504 days, > 504 – 588 days, > 588 – 672 days, > 672 – 756 days, >756 – 840 days, > 840 – 924 days, > 924 – 1008 days, > 1008 days) will be summarized.

All Elagolix plus E2/NETA Analysis Set:

For subjects who received at least 1 dose of elagolix plus E2/NETA (either in the placebo-controlled or open-label Treatment Periods), the duration of treatment is defined as the last dose date of receiving elagolix plus E2/NETA minus the first dose date of receiving elagolix plus E2/NETA plus 1 day.

In addition, the number and percentage of subjects in each treatment duration interval (1 – 84 days, > 84 – 168 days, > 168 – 252 days, > 252 – 336 days, > 336 – 420 days, > 420 – 504 days, > 504 – 588 days, > 588 – 672 days, > 672 – 756 days, > 756 – 840 days, > 840 – 924 days, > 924 – 1008 days, > 1008 – 1092 days, > 1092 – 1176 days, > 1176 – 1260 days, > 1260 – 1344 days, > 1344 days) will be summarized.

7.0 Demographics, Baseline Characteristics, Medical History, and Prior/Concomitant Medications

Demographics, baseline characteristics, medical history, prior medication, concomitant medication, and post-treatment medication will be summarized by treatment group and overall using the Full Analysis Set. Categorical variables will be summarized with the number and percentage of subjects; percentages will be calculated based on the number of non-missing observations. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum and maximum).

7.1 Demographics and Baseline Characteristics

Continuous demographic variables include age, weight, height, and body mass index (BMI). Categorical demographic variables include sex, ethnicity, race, age (18 – 19, 20 – 24, 25 – 29, 30 – 34, 35 – 39, 40 – 44, 45 – 49, ≥ 50 years), BMI (≤ 18.5, > 18.5 – < 25, ≥ 25 – < 30, ≥ 30 – < 35, ≥ 35 – < 40, ≥ 40 kg/m^2), tobacco use (current, former, never, unknown), and alcohol use (current, former, never, unknown).

For endometriosis history, the stage of endometriosis at diagnosis as reported by subjects at screening based on their gynecological history will be summarized as the number and percentage of subjects in each category (Stage 1, 2, 3, 4, or unknown) for each treatment group and overall. The number of months between the first study drug dose date and the subject's surgical diagnosis of endometriosis will be summarized for each treatment group and overall.

7.2 Medical History

Medical history and gynecological medical/surgical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term) will be summarized by treatment group and overall

for the FAS. The system organ class (SOC) will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (SOC or preferred term).

7.3 Prior, Concomitant, and Post-Treatment Medications

Prior, concomitant, and post-treatment medications will be summarized by generic name for the FAS.

Prior medications are those medications with a start date prior to the first study drug administration date. Prior medications will be summarized separately for prior endometriosis medications and other medications.

Concomitant medications are those medications, other than study drug and protocol specified analgesic medications taken for endometriosis-associated pain which will be recorded in e-Diary during Screening and the first 12 months of Treatment, taken during the Treatment Period with an end date after the first dose of study drug or ongoing at the end of the study, and a start date prior to the last dose of study drug. A medication will be considered a concomitant medication where one of the following three cases occur (1) the start date is missing and the end date is either after or on the first study drug dose date; (2) the start date is not missing and the end date is missing; (3) both the start date and the end date are missing. Concomitant medications will be summarized separately for analgesic medications taken for endometriosis-associated pain recorded in eCRF during the Open-Label Treatment Period and other medications.

Post-treatment medications are those medications with an end date after the last dose of study drug or ongoing at the end of study. Post-treatment medications will be summarized separately for analgesic medications taken for endometriosis-associated pain recorded in eCRF during the Follow-up Period and other medications.

Prior, concomitant, and post-treatment medications will be summarized by ATC Classification and preferred terms from the World Health Organization (WHO) Drug

Dictionary with number and percentage by treatment group and overall. A subject who reports two or more uses of the same medication will be counted only once within each WHO preferred term. A subject with medications with more than one generic name will be counted only once in the overall total.

8.0 Efficacy Analyses

8.1 General Considerations

All efficacy analyses will be conducted in the FAS Population.

Unless otherwise specified, during the 12-month placebo-controlled Treatment Period, the analyses of efficacy endpoints will only include data from the placebo group and elagolix plus E2/NETA group; elagolix plus E2/NETA group will be compared against placebo. Statistical tests will be conducted at an alpha level of 0.05 (two-sided). A test will be deemed significant if the P value rounded to three decimal places is less than or equal to 0.05 unless otherwise specified.

In addition, efficacy endpoints collected during each of the placebo-controlled and open-label Treatment Periods will also be summarized as observed cases (by treatment group); no comparisons or statistical tests will be performed.

The primary and secondary analyses will be performed after all ongoing subjects have completed the 12-month placebo-controlled Treatment Period and the database has been locked. This will be the only and final analysis for the co-primary and key secondary efficacy endpoints as well as all other efficacy endpoints collected in the 12-Month Placebo-Controlled Treatment Period.

Unless otherwise specified, categorical data will be summarized by frequencies and percentages; continuous data will be summarized by the mean, standard deviation, median, minimum, and maximum. For the analyses of change from baseline, the within-group changes will be summarized with the mean, standard deviation or standard error,

and 95% confidence intervals (CIs); between-group differences will be summarized with the mean, standard error, P value (as appropriate), and the 95% CIs.

For efficacy variables collected using the daily e-Diary, the baseline will be based on the average of the last 35 calendar days during the Screening Period prior to and including Study Day 1. For other efficacy variables, the baseline during the 12-month placebo-controlled Treatment period (Months 1 – 12) will be the last non-missing value obtained prior to or on Study Day 1. The baseline during the open-label Treatment Period (Months 13 – 48) will be the last non-missing value obtained prior to or on the initiation day of elagolix (either elagolix alone or elagolix pus E2/NETA).

8.2 Handling of Missing Data

Missing data will be imputed using the following methods for the efficacy analyses:

- Last Observation Carried Forward (LOCF): The LOCF analyses will use the completed evaluation from the previous post-baseline evaluation to impute missing data at the current evaluations. Values will only be carried forward if a subject prematurely discontinues and if the analysis month is equal to or after the month in which the subject prematurely discontinued. For Month 1 analysis, only observed cases will be used.
- Mixed-Effect Model Repeat Measurement (MMRM): The repeated measures analysis will be conducted using a mixed model including observed measurements at all visits. The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, and the continuous fixed covariate of baseline measurement. An unstructured variance covariance matrix will be used. Parameter estimation is based on the assumption of data being missing at random and using the method of restrictive maximum likelihood (REML).
- Observed Cases (OC): The OC analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the OC analysis for that visit. OC will exclude values after a subject prematurely discontinues from study drug.

8.3 Primary Efficacy Endpoint(s) and Analyses

8.3.1 Primary Efficacy Endpoint(s)

The co-primary efficacy endpoints will be the proportion of responders at Month 6 based upon the mutually-exclusive scales for daily assessment of Dysmenorrhea (DYS) and Non-Menstrual Pelvic Pain (NMPP) measured by the 4-point Endometriosis Daily Pain Impact Diary. Use of protocol specific rescue analgesic medication for endometriosis-associated pain as collected in the daily e-Diary will also be included in the responder definition.

For each of the co-primary endpoints, a subject is considered as a responder at Month 6 if (1) she has an absolute reduction of X or greater from baseline in average pain scores (DYS or NMPP) at Month 6; and (2) she has stable or decreased rescue analgesic use as defined in Table 1 at Month 6. The pain threshold (i.e., the value of X) will be determined based on a Receiver Operating Characteristics (ROC) analysis using the PGIC assessment at Month 6 as an anchor and change from baseline in average pain score (DYS or NMPP) at Month 6 as the independent variable. All ROC analyses will be conducted prior to blind break and will contain all randomized subjects who receive at least 1 dose of study drug. Thresholds obtained for Month 6 will be used for all monthly visits during the 12-month placebo-controlled Treatment Period, i.e., the threshold obtained for DYS at Month 6 will be used for DYS at Months 1 – 12. The details of the ROC analysis are presented in Appendix A and the pain threshold values (X) are presented in Appendix C.

Table 1. Analgesic Change During the Placebo-Controlled Treatment Period

Use of No Analgesics at Baseline		
Analgesic used at Baseline	**Analgesic dose status at End of Time Window**	**Responder?***
None	None	Responder
	Opioid analgesic and/or NSAID is started	Nonresponder
Use of Only NSAID at Baseline		
Analgesic used at Baseline	**Analgesic dose status at End of Time Window**	**Responder?***
NSAID	Dose stopped, decreases, or is stable**	Responder
	Dose increases by 15% or more	Nonresponder
	Opioid analgesic is substituted or added	Nonresponder
Use of Only Opioid Analgesic at Baseline		
Analgesic used at Baseline	**Analgesic dose status at End of Time Window**	**Responder?***
Opioid analgesic	Dose stopped, decreases, or is stable**	Responder
	Dose stopped and NSAID substituted (any dose)	Responder
	Dose decreases and NSAID added (any dose)	Responder
	Dose stable** and NSAID added (any dose)	Nonresponder
	Dose increases by 15% or more	Nonresponder
Use of NSAID and Opioid Analgesic at Baseline		
Analgesics used at Baseline	**Analgesic dose status at End of Time Window**	**Responder?***
NSAID + opioid analgesic	NSAID dose stops + opioid analgesic use stops, decreases, or is stable**	Responder
	NSAID use stops + opioid analgesic dose increases by more than 15%	Nonresponder
	NSAID dose decreases + opioid analgesic use stops, decreases, or is stable**	Responder

Table 1. Analgesic Change During the Placebo-Controlled Treatment Period (Continued)

Use of NSAID and Opioid Analgesic at Baseline		
Analgesics used at Baseline	**Analgesic dose status at End of Time Window**	**Responder?***
	NSAID dose decreases + opioid analgesic dose increases by more than 15%	Nonresponder
	NSAID dose stable** + opioid analgesic use stops, decreases, or is stable**	Responder
	NSAID dose stable** + opioid analgesic dose increases by more than 15%	Nonresponder
	NSAID dose increases by more than 15% + opioid analgesic use stops	Responder
	NSAID dose increases by more than 15% + opioid analgesic dose decreases	Responder
	NSAID dose increases by more than 15% + opioid analgesic dose is stable**	Nonresponder
	NSAID dose increases by more than 15% + opioid analgesic dose increases by 15% or more	Nonresponder

* Responder = Defined as a subject who meets the protocol-specified pain score criteria for no increase in analgesic use.

** Stable = Dose is the same as the screening dose or increases by less than 15% of the screening dose.

Dysmenorrhea (DYS) and Non-Menstrual Pelvic Pain (NMPP)

Daily instances when subjects respond "Yes" to the question "Did you have your period in the last 24 hours" is counted as a day with reported DYS. Pain reported daily in response to the e-Diary question "On the next screen, choose the item that best describes your pain during the last 24 hours when you had your period" is referred in this document as DYS related pain. Daily instances when subjects respond "No" to the question "Did you have your period in the last 24 hours" is counted as a day with reported NMPP. Pain reported daily in response to the e-Diary question "On the next screen, choose the item that best describes your pain during the last 24 hours without your period" is referred in this document as NMPP related pain.

For both DYS and NMPP, responses of "None," "Mild," "Moderate," and "Severe" will be assigned a score of 0, 1, 2, and 3, respectively.

The monthly average pain score for DYS will be based on averaging the daily pain socre over the number of days when the subject reported DYS related pain within each respective time window. The monthly average pain score for NMPP will be based on averaging the daily pain scores over the number of days when the subject reported NMPP related pain within each respective time window. If the subject's monthly average pain score (DYS or NMPP) is missing because the subject reported only the other kind of pain (mutually exclusive daily pain impact scale) in the corresponding time frame of interest (e.g., Month 6 DYS score is missing because the subject only reported NMPP during Month 6 analysis window), then the monthly average pain score for the relevant parameter (DYS or NMPP) will be set to 0 (zero).

Rescue Analgesic Use

For each class of rescue analgesics (NSAIDs and Opioids), analgesic use for each defined monthly period during the 12-month placebo-controlled Treatment Period will be based on the average pill count as collected in the daily e-Diary. The monthly average pill count for each class of rescue analgesics will be calculated by dividing the total pill count for the corresponding class of rescue analgesic by the length of the analysis window. For each class of rescue analgesic, if no analgesic use is reported for a period of interest, but e-Diary information is available for the corresponding time period, then the average pill count for the corresponding class of rescue analgesic use will be set to 0 (zero).

8.3.2 Handling of Missing Data for the Primary Efficacy Endpoints

For the co-primary endpoints, subjects who prematurely discontinue at or prior to Month 6 will have their average pain scores (DYS and NMPP) and rescue analgesic use carried forward (LOCF) from the last month in which both monthly average pain score (DYS and NMPP) and analgesic use are observed.

8.3.3 Primary Efficacy Analysis

The primary null hypothesis for this study is that there is no difference in the proportion of responders for each co-primary efficacy endpoint at Month 6 between elagolix plus E2/NETA and placebo. The null hypothesis will be tested against the alternate hypothesis that there is a difference between elagolix plus E2/NETA and placebo in the proportion of responders for each co-primary efficacy endpoint at Month 6.

Elagolix plus E2/NETA will be considered statistically more efficacious than placebo for the co-primary endpoint if and only if the two-sided P value for the comparison of the proportion of responders for both DYS and NMPP is less than or equal to 0.05 and in favor of elagolix plus E2/NETA.

The primary analysis of each of the co-primary endpoints will be based on a logistic regression model with the responder/non-responder categorization as the dependent variable, treatment as the main effect, and baseline pain score (DYS or NMPP) as the covariate. Elagolix plus E2/NETA will be compared to placebo for each of the co-primary endpoints. The proportion of responders, difference in proportions between each elagolix group and placebo, and odds ratio with corresponding 95% confidence intervals (CIs) and P values will be presented.

8.3.4 Additional Analyses of the Primary Efficacy Endpoint(s)

The following sensitivity analyses for each of the co-primary endpoints of the proportion of responders at Month 6 in pain scores will be conducted:

1. The difference between the elagolix plus E2/NETA and placebo in response rates will be analyzed using a chi-square test with a corresponding 95% CI for the difference in response rate based on normal approximation to the binomial distribution.
2. All subjects who prematurely discontinue the study drug at or before Month 6 will be considered as non-responders and the primary analysis will be repeated.

3. The primary analysis will be repeated using mixed-imputation. In particular, subjects who discontinue prior to or at Month 6 due to an adverse event (AE) or lack of efficacy will be considered non-responders, while subjects who discontinue prior to or at Month 6 due to other reasons will have their monthly pain score carried forward following the LOCF rule.
4. All subjects who switched NSAIDs use at Baseline, Month 6, or from Baseline to Month 6 will be considered as non-responders and the primary analysis will be repeated.
5. All subjects who switched NSAIDs use at Baseline, Month 6, or from Baseline to Month 6 will be excluded from the primary analysis.
6. The following rules will be used to standardize different NSAIDs use and opioid use. The standardized rescue analgesic use will be re-evaluated based on Table 1 and the primary analysis will be repeated.
    - NSAID: 1 pill Naproxen = 1 pill Ibuprofen = 1 pill Diclofenac = 4 pills Celecoxib = 1 pill of NSAIDs use
    - Opioid: 1 pill hydrocodone + acetaminophen = 1 pills codeine phosphate + acetaminophen = 1 pill of opioid use

8.4 Secondary Efficacy Analyses

8.4.1 Key Secondary Efficacy Analyses

8.4.1.1 Change from Baseline in DYS, NMPP, Dyspareunia, and Overall Endometriosis-Associated Pain

The change from baseline in DYS, NMPP, dyspareunia, and Overall endometriosis-associated pain at Months 3, 6, and 12 will be evaluated with mixed-effects model repeated measures (MMRM) analysis. The MMRM analysis will include the fixed categorical effects of treatment, visit, and treatment-by-visit interaction, and continuous fixed covariate of baseline score. A random subject effect term is included to account for the correlation among the repeated measurements in the changes from baseline. The

following covariance structures will be used to estimate the within subject errors in preliminary analyses for the changes from baseline to each scheduled visit: Spatial Power, Compound Symmetry, and Unstructured. For each efficacy parameter, the covariance structure converging to the best fit, as determined by the smallest value for the Akaike Information Criterion (AIC), will be used as the covariance structure in the MMRM analysis for that efficacy parameter. The LS mean and standard error will be reported for each treatment group; the LS mean of treatment difference and associated 95% CI and P value will be reported comparing elagolix plus E2/NETA group against placebo group.

8.4.1.2 Change from Baseline in PROMIS Fatigue Short Form 6a T-Score

The PROMIS Fatigue Short Form 6a will be summarized as observed. Missing data will not be imputed. The change from baseline in PROMIS Fatigue Short Form 6a T-score at Month 6 and 12 will be evaluated using ANCOVA with treatment as the main effect and baseline score as a covariate. The LS mean and standard error will be reported for each treatment group; the LS mean of treatment difference and associated 95% CI and P value will be reported comparing elagolix plus E2/NTEA group against placebo group.

8.5 Additional Efficacy Analyses

8.5.1 Endpoints related to DYS, NMPP, Dyspareunia and Overall endometriosis-associated pain

The proportion of responders at Months 1 – 12 based on change from baseline in average pain scores (DYS, NMPP, or dyspareunia) and analgesic use criteria (as defined in Table 1) will be analyzed using a logistic regression model with the responder/non-responder categorization as the dependent variable, treatment as the main effect, and baseline pain score (DYS, NMPP or dyspareunia) as a covariate. The proportion of responders, difference in proportions between elagolix plus E2/NETA group and placebo group, and odds ratio with corresponding 95% CI and P values will be presented. The pain threshold for defining responders will be determined using the same ROC analysis

described for the co-primary endpoints. Details of the ROC analysis are presented in Appendix A and the pain threshold values (X) are presented in Appendix C. The same analyses will be repeated for the proportion of responders at Months 1 – 12 based only on change from baseline in average pain scores (DYS, NMPP, or dyspareunia). The responder analysis will be using the LOCF approach.

In addition, the proportion of responders at Months 1 – 12 based on change from baseline in average pain scores (DYS, NMPP, or dyspareunia) and analgesic use criteria and proportion of responders based on change from baseline in average pain scores (DYS, NMPP, or dyspareunia) only will also be summarized by treatment group using observed cases.

The change from baseline to each month in DYS, NMPP, Dyspareunia, and overall endometriosis-associated pain during the 12-month placebo-controlled Treatment Period will be evaluated using MMRM analysis. The MMRM analysis will include the fixed categorical effects of treatment, visit, and treatment-by-visit interaction, and the continuous fixed covariate of baseline pain score. A random subject effect term is included to account for the correlation among the repeated measurements in the changes from baseline. The selection and use of the covariance structure used in this analysis will follow the same process as described previously for the ranked secondary endpoints in Section 8.4.1.1. The LS mean and standard error will be reported for each treatment group; the LS mean of treatment difference and associated 95% CI and P value will be reported comparing elagolix plus E2/NETA group against placebo group.

In addition, the change from baseline to each month in DYS, NMPP, dyspareunia, and overall endometriosis-associated pain during the 12-month placebo-controlled Treatment Period will be summarized by treatment group using observed cases. The change from baseline to each monthly visit in overall endometriosis-associated pain during the open-label Treatment Period will also be summarized by treatment group using observed cases.

In addition, the following plots will be provided using observed cases:

- Cumulative distribution function for reduction from baseline in DYS, NMPP, and dyspareunia to Months 1 – 12.

8.5.2 Rescue Analgesic Use for Endometriosis-Associated Pain

The monthly average rescue medication pill counts as collected in the e-diary will be summarized using observed cases only with descriptive statistics by treatment group for the categories of analgesic use – NSAIDs, opioids, any class of analgesic medications (NSAIDs or opioids), and both classes of analgesic medications (NSAIDs and opioids) during the 12-month placebo-controlled Treatment Period.

For the following categories of analgesic use – NSAIDs, opioids, any class of analgesic medications (NSAIDs or opioids), and both classes of analgesic mediations (NSAIDs and opioids), the change from baseline to each month during the 12-month placebo-controlled Treatment Period will be evaluated using an MMRM analysis. The MMRM analysis will include the fixed categorical effects of treatment, visit, and treatment-by-visit interaction, and the continuous fixed covariate of baseline analgesic use. A random subject effect term is included to account for the correlation among the repeated measurements in the dependent variable. The selection and use of the covariance structure used in this analysis will follow the same process as described previously for the ranked secondary endpoints in Section 8.4.1.1. The LS mean and standard error will be reported for each treatment group; the LS mean of treatment difference and associated 95% CI and P value will be reported comparing elagolix plus E2/NETA group against placebo group. In addition, the change from baseline to each monthly visit during the 12-month placebo-controlled Treatment Period will be summarized by treatment group using observed cases.

Proportion of responders based solely on the change from baseline in analgesic use will be summarized by treatment group at each visit during the 12-month placebo-controlled Treatment Period using the observed cases. Definition of analgesic use responder will be based on Table 1. In addition, the proportion of responders will be analyzed using a chi-square test based on LOCF approach. The difference in proportions between elagolix plus

E2/NETA group and placebo group, and odds ratio with corresponding 95% CI and P values will be presented.

The proportion of responders for each category in Table 1 during the 12-month placebo-controlled Treatment Period will be summarized using LOCF for elagolix plus E2/NETA group and placebo group. In addition, the proportion of responders for each category in Table 1 during the 12-month placebo-controlled Treatment Period will be summarized by treatment group using observed cases.

For both NSAIDs and opioids, the monthly average rescue medication pill counts as collected in the e-diary will be summarized by medication name and treatment group using observed cases during the 12-month placebo-controlled Treatment Period. In addition, number of subjects who switch NSAID and opioid medication at baseline, Month 6, and any time from Baseline to Month 6 during the placebo-controlled Treatment Period will be summarized.

In addition, the standardized monthly average rescue medication pill counts as collected in the e-diary for both NSAIDs and opioids will be summarized by treatment group using observed cases during the 12-month placebo-controlled Treatment Period. The following rule will be used to standardize NSAIDs and opioid:

- NSAIDs: 1 pill Naproxen = 1 pill Ibuprofen = 1 pill Diclofenac = 4 pills Celecoxib = 1 pill of NSAID use;
- Opioids: 1 pill hydrocodone + acetaminophen = 1 pills codeine phosphate + acetaminophen = 1 pill of opioid use.

The monthly total rescue medication pill counts as collected in the e-diary will be summarized using observed cases only with descriptive statistics by treatment group for the NSAIDs and opioids use during the placebo-controlled Treatment Period.

For subjects with at least 1 pill use of NSAIDs at baseline, the number and percentage of subjects who are off NSAIDs use at each month will be summarized using observed cases. In addition, for subjects who didn't use NSAIDs at baseline, the number and percentage of

subjects who take at least 1 pill count of NSAIDs at each month will be summarized using observed cases. Same analyses will be repeated for opioids.

8.5.3 Patient Global Impression of Change (PGIC)

During the 12-month placebo-controlled Treatment Period and open-label Treatment period, PGIC will be summarized by treatment group using the observed cases. In addition, PGIC will be summarized for elagolix plus E2/NETA group and placebo group during the 12-month placebo-controlled Treatment Period using LOCF approach.

Categorical summary of the PGIC results during the Treatment Period will be presented as follows.

- The number and percentage of subjects in each PGIC response category (7-point scale) will be summarized. No statistical tests will be performed.
- The response categories of "Very Much Improved" and "Much Improved" will be combined and labeled as "Much Improved or Better." The remaining five categories of the PGIC scale will be combined and labeled "Otherwise." The number and percentage of subjects with response of a) "Much Improved or Better" and b) "Otherwise" will be summarized. Elagolix plus E2/NETA group will be compared with placebo using a chi-square test at each of Months 1 – 12. No statistical tests will be performed at Months 13 – 48.

8.5.4 Endometriosis Health Profile Questionnaire (EHP-30)

EHP-30 data will be summarized and analyzed using observed cases. Missing data will not be imputed.

EHP-30 data will be summarized for the five dimensions of the core questionnaires (Pain, Control and Powerlessness, Emotional well-being, Social Support, Self-Image) and the dimension of sexual intercourse of the modular questionnaire during the Treatment Period.

During the 12-month placebo-controlled Treatment Period, the change from baseline to each scheduled assessment for each dimension of the EHP-30 will be evaluated using an ANCOVA model with treatment as the main effect and baseline as a covariate. The LS mean and standard error will be reported for each treatment group; the LS mean of treatment difference and associated 95% CI and P value will be reported comparing elagolix plus E2/NETA group with placebo.

In addition, the change from baseline to each scheduled assessment for each dimension will be summarized with descriptive statistics during the Treatment Period. No comparisons or statistical tests will be performed.

8.5.5 EuroQol-5D 5 Level (EQ-5D-5L)

EQ-5D-5L data will be summarized and analyzed using observed cases. Missing data will not be imputed.

The number and percentage of subjects with answers in each level of the EQ-5D-5L dimensions (Mobility, Self-care, Usual activities, Pain/Discomfort, and Anxiety/Depression) will be summarized with descriptive statistics during the Treatment Period. No comparisons or statistical tests will be performed.

Subject's responses to the EQ-5D-5L will be combined into a unique health state, which consists of a 5-digit code with 1 digit from each of the 5 dimensions at each scheduled assessment during the Treatment Period. The EQ-5D-5L states will be converted into a single preference-weighted health utility index score by application of country-specific weights (if available) or US weights if country-specific weights are unavailable.

During the 12-month placebo-controlled Treatment Period, the change from baseline to each scheduled assessment for health utility index score and EQ VAS will be analyzed using ANCOVA with treatment as the main effect and baseline as a covariate. The LS mean and standard error will be reported for each treatment group; the LS mean of treatment difference and associated 95% CI and P value will be reported comparing elagolix plus E2/NETA group with placebo.

In addition, the change from baseline to each scheduled assessment for health utility index score and EQ VAS will be summarized with descriptive statistics during the Treatment Period. No comparisons or statistical tests will be performed.

8.5.6 Work Productivity and Activity Impairment Questionnaire: Specific Health Problem (WPAI:SHP)

Each of the 4 measures of the WPAI:SHP data (work time missed, impairment while working, overall work impairment, and activity impairment) will be summarized and analyzed using observed cases. Missing data will not be imputed.

During the 12-month placebo-controlled Treatment Period, the change from baseline to each scheduled assessment for each of the measures as presented in the scoring guide will be analyzed using ANCOVA with treatment as the main effect and baseline as a covariate. The LS mean and standard error will be reported for each treatment group; the LS mean of treatment difference and associated 95% CI and P value will be reported comparing elagolix plus E2/NETA group with placebo.

In addition, the change from baseline to each scheduled assessment for each of the measures as presented in the scoring guide will be summarized with descriptive statistics during the Treatment Period. No comparisons or statistical tests will be performed.

8.5.7 PROMIS Fatigue Short Form 6a

The PROMIS Fatigue Short Form 6a will be summarized and analyzed using observed cases. Missing data will not be imputed.

During the 12-month placebo-controlled Treatment Period, the change from baseline to each scheduled assessment for PROMIS Fatigue Short Form 6a T-score will be analyzed using ANCOVA with treatment as the main effect and baseline as a covariate. The LS mean and standard error will be reported for each treatment group; the LS mean of treatment difference and associated 95% CI and P value will be reported comparing elagolix plus E2/NETA group with placebo.

In addition, the change from baseline to each scheduled assessment for PROMIS Fatigue Short Form 6a T-score will be summarized with descriptive statistics during the Treatment Period. No comparisons or statistical tests will be performed.

8.5.8 Health Care Resource Utilization Questionnaire (HCRU)

HCRU data will be summarized using observed cases. Missing data will not be imputed.

HCRU data will be summarized with descriptive statistics by Visit during the Treatment Period. The total number and percentage of Non-Study Health Care Practitioner visits and endometriosis-related Non-Study Health Care Practitioner visits overall and by the type of facility will be summarized. In addition, the number and percentage of subjects with Non-Study Health Care Practitioner visits and endometriosis-related Non-Study Health Care Practitioner visits overall and by the type of facility will be summarized.

The number and percentage of subjects seen at office will be presented by the type of Non-Study Health Care Practitioner and endometriosis-related Non-Study Health Care Practitioner who administered care to the subject. In addition, number of Non-Study Health Care Practitioner visits and endometriosis-related Non-Study Health Care Practitioner visits by practitioner will be summarized.

The number and percentage of subjects will be presented by the type of diagnostic or therapeutic procedures and endometriosis-related diagnostic or therapeutic procedures by visit during the Treatment Period. No statistical tests will be performed.

8.5.9 Hospitalization

Hospitalization related data will be summarized using observed cases. Missing data will not be imputed. Hospitalization data will be summarized with descriptive statistics during the Treatment Period based on the Adverse Events eCRF. No statistical tests will be performed.

The number and percentage of subjects who were hospitalized or had prolonged hospitalization will be summarized by treatment group. If a subject was hospitalized or

had prolonged hospitalization multiple times during the Treatment Period, she will be counted only once.

The number of hospitalizations will be summarized by treatment group. In addition, the number of days in hospital (i.e., discharge date – admission date + 1) will be summarized by treatment group. Only hospitalizations with admission date on or after the first dose date and within 30 days following the last dose date (i.e., first dose date ≤ admission date ≤ last dose date + 30) will be included in the analysis of the number of days in hospital. If discharge date or admission date is missing, then the hospitalization will be excluded from the analysis of the number of days in hospital.

8.6 Efficacy Subgroup Analyses

For each co-primary endpoint, the following subgroups will be explored to assess potential differences in treatment effect across subgroup levels. The subgroup analyses will be performed at Month 6, which represents the study visit for statistical evaluation of the co-primary efficacy endpoints.

- Baseline BMI: < median, ≥ median
- Baseline BMI category: Normal (< 25 kg/m2), Overweight (25 – < 30 kg/m2), Obese (≥ 30 kg/m2)
- Race: Black or African American, Not Black or African American
- Baseline Age: < 25 years old, 25 – ≤ 35 years old, > 35 years old.

The subgroup analysis will consist of a logistic regression with overall responder as the response variable, baseline value as a covariate, subgroup and treatment as main effects, and a treatment by subgroup interaction term.

9.0 Safety Analyses

9.1 General Considerations

Safety summaries will be provided separately using the Safety Analysis Set during the placebo-controlled Treatment Period, Safety Analysis Set during the open-label Treatment

Period, and All Elagolix plus E2/NETA Analysis Set as defined in Section 4.0. Unless otherwise specified, no statistical comparisons or tests will be performed on Safety Analysis Set during the open-label Treatment Period or the All Elagolix plus E2/NETA Analysis Set.

Unless otherwise specified, the baseline for the 12-month placebo-controlled Treatment Period will be the last non-missing value obtained prior to or on Study Day 1. The baseline during the open-label Treatment Period will be the last non-missing value obtained prior to or on the initiation day of elagolix (either elagolix alone or elagolix plus E2/NETA). The baseline for All Elagolix plus E2/NETA Analysis Set will be the last non-missing value obtained prior to or on the initiation day of elagolix plus E2/NETA.

All safety analyses will be based on observed data. Unless otherwise specified, missing data will not be imputed.

9.2 Adverse Events

Adverse events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) according to the version of the MedDRA coding dictionary used for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

9.2.1 Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) for each safety population are defined as follows:

TEAEs during the 12-month placebo-controlled Treatment Period on Safety Analysis Set during the placebo-controlled Treatment Period

TEAEs during the 12-month placebo-controlled Treatment Period are defined as any AEs with the onset that is on or after the first dose of study drug during the 12-month placebo-controlled Treatment Period and no more than 30 days after the last dose of study drug for subjects who premature discontinue during the 12-month double-blind Treatment Period or until the first dose of study drug in the open-label Treatment Period for subjects who enter the open-label Treatment Period. AEs starting more than 30 days following discontinuation of study drug will not be included in summaries of treatment-emergent AEs but will be summarized separately.

TEAEs during the open-label Treatment Period on the Safety Analysis Set during the open-label Treatment Period

TEAEs during the open-label Treatment Period are defined as any AEs with the onset that is on or after the first dose of study drug during the open-label Treatment Period and no more than 30 days after the last dose of study drug in the study. AEs starting more than 30 days following discontinuation of study drug will not be included in summaries of treatment-emergent AEs but will be summarized separately.

TEAEs for All Elagolix plus E2/NETA Analysis Set

TEAEs for All Elagolix plus E2/NETA Analysis Set are defined as any AEs with the onset that is on or after the first dose of elagolix plus E2/NETA (either in the placebo-controlled Treatment Period or open-label Treatment Period) and no more than 30 days after last dose of elagolix plus E2/NETA. AEs starting more than 30 days following discontinuation of study drug will not be included in summaries of treatment-emergent AEs but will be summarized separately.

The number and percentage of subjects experiencing treatment-emergent AEs will be summarized overall, as well as by primary MedDRA SOC and PT by treatment group.

The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

9.2.2 Adverse Event Overview

An overview of treatment-emergent AEs will be presented consisting of the number and percentage of subjects experiencing at least one event for each of the following AE categories:

- Any treatment-emergent AE
- Any treatment-emergent AE related to study drug according to the investigator
- Any severe treatment-emergent AE
- Any serious treatment-emergent AE
- Any treatment-emergent AE leading to discontinuation of study drug
- Any treatment-emergent AE leading to death
- Any treatment-emergent AE of special interest as specified in Appendix B
- All deaths

In addition, for subjects who received elagolix 200 mg BID alone at Study Day 1, an overview of AEs during the placebo-controlled Treatment Period will be summarized separately for the first 6-month and second 6-month.

9.2.3 Treatment-Emergent Adverse Events by SOC and PT

Treatment-emergent adverse events will be summarized by SOC and PT; by maximum relationship to study drug as assessed by the investigator (e.g., reasonable possibility or no reasonable possibility) and SOC and PT; by maximum severity and SOC and PT; and by subject number and SOC and PT. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

In addition, treatment-emergent adverse events will be summarized by PT and sorted by decreasing frequency for the elagolix plus E2/NETA group.

In addition, for subjects who received elagolix 200 mg BID alone at Study Day 1, TEAEs during the placebo-controlled Treatment Period will be summarized separately for the first 6-month and second 6-month by SOC and PT.

9.2.4 Treatment-Emergent Adverse Events per Patient-Years of Exposure

An overview of AEs for All Elagolix plus E2/NETA Analysis Set will be presented per 100 patient-years. In addition, TEAEs for All Elagolix plus E2/NETA Analysis Set will be presented per 100 patient-years by SOC and PT. AEs per 100 patient-years of exposure are defined as the number of AEs divided by the total exposure in 100 patient-years. Note that one event per preferred term per day per subject will be counted in the calculation of the number of AEs (i.e., a preferred term will not be counted twice on the same day for the same subject).

$$\text{TEAEs per 100 patient-years of exposure} = 100 \times \frac{\text{Number of TEAEs}}{\text{Total Patient Years}}$$

where total patient years are defined as the sum of the study drug exposure (defined as date of last dose of elagolix plus E2/NETA – date of first dose of elagolix plus E2/NETA + 30 days) of all subjects normalized by 365.25, and rounded to one decimal place.

9.2.5 Serious Adverse Events (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation

The number and percentage of subjects experiencing the following categories will be tabulated according to the primary MedDRA SOC and PT. In addition, listing of subjects experiencing the following categories will be generated.

- Treatment-emergent SAEs

- Treatment-emergent AEs leading to discontinuation of study drug
- Treatment-emergent AEs leading to death

9.2.6 Adverse Events of Special Interest

Adverse events of special interest will be summarized by SOC and PT and will be based on standardized or company MedDRA queries (SMQs or CMQs), or based on adjudication results.

Detailed information about the search criteria are provided in Appendix B.

9.2.7 Hot Flushes and Night Sweats

Summary of total number of hot flushes and night sweats in the last 24 hours by subject-reported severity categories and summary of total number of hot flushes, night sweats, and hot flushes or night sweats in the last 24 hours by maximum subject-reported severity will be provided by treatment group at each time point during the Treatment Period and Follow-up Period. Listings of subject IDs associated with hot flush and night sweats in the last 24 hours by subject-reported severity categories and by maximum subject-reported severity categories will be provided separately.

In addition, summary of time to first onset of treatment-emergent AE of hot flush, night sweats, hot flush and/or night sweats will be provided separately.

9.2.8 Post-Treatment Adverse Events

Post-Treatment Adverse Events are defined as AEs starting more than 30 days following discontinuation of study drug in the Treatment Period. The Post-Treatment AEs will be summarized as follows:

- Overview of Post-Treatment AEs
- Post-Treatment AEs by SOC and PT
- Post-Treatment SAEs by SOC and PT

In addition, listing of subjects with Post-Treatment AEs by SOC and PT and Post-Treatment SAEs by SOC and PT will be provided.

9.3 Analysis of Laboratory Data

Data collected from central and local laboratories, including additional laboratory testing due to an SAE, will be used in all analyses, except for Baseline where SAE-related laboratory assessments on or before the first dose of study drug will be excluded. The clinical laboratory tests defined in the protocol operations manual (i.e., hematology and clinical chemistry) will be summarized.

Each laboratory variable will be summarized for all time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from baseline to each applicable post-baseline visit will be summarized for selected laboratory variables, with the number of observations, baseline mean, and visit mean. The change from baseline mean, standard error, and 95% confidence interval will be presented for the mean change from baseline within each treatment group and difference between each elagolix group and placebo. Selected laboratory variables are lipid variables, liver variables (alkaline phosphatase, ALT, AST, bilirubin), hemoglobin, hematocrit, platelet count, and red blood cell (RBC) count. For lipid variables, in addition to low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), total cholesterol, triglycerides (TG)m and apolipoprotein A and B, the following ratios will be included: the ratio of total cholesterol to HDL-C, the ratio of LDL-C to HDL-C, the ratio of TG to HDL-C, and the ratio of non-HDL-C (calculated as total cholesterol minus HDL-C) to HDL-C.

Changes in selected laboratory parameters will be tabulated using shift tables by NCI CTC criteria and categorized as low, normal, or high based on the normal ranges of the laboratory used for each sample. A shift table from baseline either to the worse value (based on NCI CTC criteria) during treatment or to minimum and maximum value (based on normal range), will be created. A similar shift table will be provided to summarize shifts from baseline to the final post-baseline value.

Laboratory abnormalities meeting CTC criteria grade 3 or 4 will be summarized.

The number and percentage of subjects meeting the following sponsor-defined potentially clinically significant lipid values will be summarized at baseline and each relevant visit by treatment group in the appropriate Treatment Period:

- Total cholesterol: ≤ 300, > 300 – ≤ 400, > 400 – ≤ 500, and > 500 mg/dL
- HDL-C: < 40 and ≥ 40 mg/dL
- LDL-C: < 130, ≥ 130 – < 160, ≥ 160 – < 190, and ≥ 190 mg/dL
- TG: ≤ 150, > 150 – ≤ 300, > 300 – ≤ 500, > 500 – ≤ 1000, and > 1000 mg/dL
- LDL-C/HDL-C ratio: ≤ 3, and > 3
- Total cholesterol/HDL-C ratio: ≤ 4.5, and > 4.5

In addition, the number and percentage of subjects who have potentially clinically significant (PCS) lipid values meeting the following criteria any time during the appropriateTreatment Period will be summarized by treatment group. Listings will be provided to summarize subject-level laboratory data for subjects meeting PCS criteria.

- Total cholesterol > 200 mg/dL
- LDL-C > 130 mg/dL
- LDL-C > 160 mg/dL
- HDL-C < 40 mg/dL
- TG > 150 mg/dL
- TG/HDL-C ratio > 3.5
- LDL-C/HDL-C ratio > 4

The following plots will be provided by treatment group for HDL-C, LDL-C, and triglycerides during the appropriate Treatment Period:

- Final post-baseline lab values vs. Baseline lab values
- Final post-baseline lab values vs. Baseline lab values for subjects with NCI CTC Grade 3 or 4

- Worst post-baseline lab values vs. Baseline lab values
- Worst post-baseline lab values vs. Baseline lab values for subjects with NCI CTC Grade 3 or 4
- Mean percent change from baseline in lipid parameters over time

In addition, plots of mean lab values over time during the Treatment Period will be provided for hemoglobin, total cholesterol, HDL-C, LDL-C, triglycerides, LDL-C/HDL-C ratio, ALT, AST, bilirubin, apolipoprotein A and apolipoprotein B.

The number and percentage of subjects in each treatment group with maximum on-treatment laboratory values meeting the following criteria compared to the upper limit of normal (ULN) during the appropriate Treatment Period will be summarized to assess potential hepatotoxicity.

- ALT $\geq 3 \times$ ULN and total bilirubin $\geq 2 \times$ ULN
- AST $\geq 3 \times$ ULN and total bilirubin $\geq 2 \times$ ULN
- ALT and AST $\geq 3 \times$ ULN and total bilirubin $\geq 2 \times$ ULN
- ALT $\geq 3 \times$ ULN and total bilirubin $< 2 \times$ ULN
- AST $\geq 3 \times$ ULN and total bilirubin $< 2 \times$ ULN
- ALT and AST $\geq 3 \times$ ULN and total bilirubin $< 2 \times$ ULN
- ALT $\geq 3 \times$ ULN and total bilirubin $\geq 1.5 \times$ ULN
- AST $\geq 3 \times$ ULN and total bilirubin $\geq 1.5 \times$ ULN
- ALT and AST $\geq 3 \times$ ULN and total bilirubin $\geq 1.5 \times$ ULN
- ALT $\geq 3 \times$ ULN, $\geq 5 \times$ ULN, $\geq 10 \times$ ULN, $\geq 20 \times$ ULN
- AST $\geq 3 \times$ ULN, $\geq 5 \times$ ULN, $\geq 10 \times$ ULN, $\geq 20 \times$ ULN
- Total bilirubin $\geq 1.5 \times$ ULN, $\geq 2.0 \times$ ULN

The maximum ratio relative to the ULN is used to determine if subjects met the criteria listed above. The ALT, AST, and total bilirubin values do not need to be concurrent in order to meet the defined criteria. For ALT, AST, and total bilirubin, a subject is counted if the post-baseline laboratory value during the Treatment Period meets the above criteria

regardless of Baseline laboratory value (i.e., the post-baseline laboratory value does not need to be worse than Baseline laboratory value).

A listing of all ALT, AST, total bilirubin, and alkaline phosphatase values will be provided for each subject who meets any of the criteria defined above.

The following plots will be generated: peak AST vs. peak bilirubin; peak ALT vs. peak bilirubin.

9.4 Analysis of Vital Signs and Weights

Vital sign measurements of pulse rate, sitting systolic blood pressure, and sitting diastolic blood pressure will be summarized.

Each vital sign variable and weight will be summarized for all time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from baseline to each applicable post-baseline visit will be summarized for each vital sign variable, with the number of observations, baseline mean, and visit mean. The change from baseline mean, standard error, and 95% confidence interval will be presented for the mean change from baseline within each treatment group and difference between each elagolix group and placebo.

The number and percentage of subjects who have potentially clinically significant (PCS) vital sign and weight values meeting the following criteria will be summarized by treatment group. All increase/decrease is calculated from Baseline to a post-baseline visit in the Treatment Period.

- Diastolic blood pressure
  - $\leq$ 50 mmHg and $\geq$ 15 mmHg decrease
  - $>$ 90 mmHg and $\geq$ 15 mmHg increase
  - $\geq$ 100 mmHg
- Systolic blood pressure
  - $\leq$ 90 mmHg and $\geq$ 20 mmHg decrease

- ≥ 140 mmHg and ≥ 20 mmHg increase
- ≥ 160 mmHg

- Pulse rate
  - ≤ 45 bpm and ≥ 15 bpm decrease
  - > 100 bpm and ≥ 15 bpm increase
  - ≥ 120 bpm
- Weight
  - ≥ 5% decrease
  - ≥ 7% increase.

Listings will be provided to summarize subject-level vital sign and weight data for subjects meeting PCS criteria.

The number and percentage of subjects who had a sustained PCS vital sign and weight values will be summarized by treatment group and a listing of these subjects will be provided. A sustained PCS value is defined as 3 consecutive PCS values.

9.5 Other Safety Analyses

9.5.1 Columbia Suicide Severity Rating Scale (C-SSRS)

The C-SSRS data will be summarized as observed by treatment group.

The C-SSRS – Baseline/Screening measured at Day 1 will be considered as Baseline C-SSRS. Baseline C-SSRS will be summarized. Other analysis of C-SSRS will only include subjects who have at least 1 post-baseline C-SSRS measurement, regardless of whether they had a baseline C-SSRS measurement. The number of subjects with suicidal ideation, suicidal behavior, and self-injurious behavior without suicidal intent at each scheduled assessment during the appropriate Treatment Period will be summarized. In addition, this table will be repeated for providing a summary of lifetime outcomes and past year outcomes at screening and Day 1. No statistical test will be performed.

The number of subjects with suicide-related treatment-emergent events based on the C-SSRS during the appropriate Treatment Period will be summarized. No statistical tests will be performed.

A listing of subjects with suicidal ideation, suicidal behavior, or self-injurious behavior without suicidal intent based on the C-SSRS during treatment period will be provided.

9.5.2 Bone Mineral Density

All analyses and summaries of bone mineral density (BMD) will be performed for each anatomic location, i.e., femoral neck, lumbar spine, and total hip. For subjects who had a right-side scan performed (rare instances), their data for right femoral neck and right femoral total hip will be included in the analysis with the data for the left femoral neck and left total hip, respectively (available for the majority of subjects). If more than one scan is reported for an anatomic location within an analysis window, the worse (the lower value) of the multiple measurements will be used for analysis for each anatomic location. Unless otherwise specified, the analysis of BMD will exclude subjects who switch machine manufacturer type (Lunar or Hologic).

9.5.2.1 Bone Mineral Density during the Treatment Period

BMD

A summary of continuous BMD at each scheduled assessment will be provided during the appropriate Treatment Period. This summary will include the sample size, mean, SD, median, minimum, and maximum. This analysis will be repeated including subjects who switch machine manufacturer type (Lunar or Hologic), for subjects with scans only on Lunar machines, and for subjects with scans only on Hologic machines.

The analysis of percent change in BMD from baseline to each relevant visit during the Treatment Period will be based on an ANCOVA model with treatment as the main effect and baseline value of corresponding parameter as a covariate. Elagolix plus E2/NETA group will be compared to placebo at Months 6 and 12 and will be compared to the

elagolix alone group at Month 6 during the placebo-controlled Treatment Period. In addition, elagolix alone group will be compared to placebo at Months 6 and 12 during the placebo-controlled Treatment Period. This analysis will be repeated including subjects who switch machine manufacturer type (Lunar or Hologic), for subjects with scans only on Lunar machines, and for subjects with scans only on Hologic machines.

The number and percentage of subjects with percent change from Baseline to each scheduled assessment during the Treatment Period in mutually exclusive categories of BMD decrease: $\leq 0\%$, $> 0\% - \leq 1.5\%$, $> 1.5\% - \leq 3\%$, $> 3\% - \leq 4\%$, $> 4\% - \leq 5\%$, $> 5\% - < 8\%$, and $\geq 8\%$ will be summarized by treatment group. Elagolix plus E2/NETA group will be compared to placebo at Months 6 and 12 and will be compared to elagolix alone group at Month 6 during the placebo-controlled Treatment Period using chi-square tests. In addition, elagolix alone group will be compared to placebo at Months 6 and 12 during the placebo-controlled Treatment Period.

The analysis of percent change in BMD from Month 6 to Month 12 during the Treatment Period will be based on an ANCOVA model with treatment as the main effect and Month 6 BMD values of corresponding parameter as a covariate for subjects who have both Month 6 and Month 12 BMD values. Same analyses will be provided for percent change in BMD from Month 6 to Month 18 and Month 24. Each elagolix group will be compared to placebo.

Cross-tabulation of the DXA scans performed versus Treatment Period completion will be provided after all subjects complete the Treatment Period or premature discontinue from the Treatment Period.

Listings of subjects meeting the following criteria during the Treatment Period will be provided:

- Listing of subjects with BMD decrease $\geq 8\%$

In addition, the following plot will be provided as needed:

- Percent change from Baseline to each scheduled assessment (Months 6, 12, 18, 24, 30, 36, 42, 48) during the Treatment Period in BMD vs. Baseline BMD values for subjects with greater than 8% BMD decrease at any location during the Treatment Period.
- Cumulative distribution function of percent change from baseline in BMD at each scheduled assessment (Months 6, 12, 18, 24, 30, 36, 42, 48) during the Treatment Period.
- Percent change from Baseline to each scheduled assessment (Months 6, 12, 18, 24, 30, 36, 42, 48) in BMD during the Treatment period.

Z-Score and T-Score

A continuous summary of the Z-score and T-score at each scheduled assessment will be provided by treatment group during the Treatment Period. This summary will include the sample size, mean, SD, median, minimum, and maximum.

A categorical summary of Z-score will be provided at each scheduled assessment during the Treatment Period for the following categories: ≤ -2.0, > -2.0 to ≤ -1.5, > -1.5 to ≤ -1.0, and > -1.0. A categorical summary of T-score will be provided at each scheduled assessment during the Treatment Period for the following categories: ≤ -2.5, > -2.5 to < -1.0, and ≥ -1.0. In addition, a categorical summary of the worst Z-score and T-score during the Treatment Period for the above categories will be provided.

9.5.2.2 Bone Mineral Density during the Follow-up Period

BMD

A summary of continuous BMD at each scheduled assessment will be provided by treatment group during the Follow-up Period. This summary will include the sample size, mean, SD, median, minimum, and maximum. This analysis will be repeated including subjects who switch machine manufacturer type (Lunar or Hologic).

The analysis of percent change in BMD from baseline to each relevant visit during the Follow-up Period will be based on an ANCOVA model with treatment as the main effect

and baseline value of corresponding parameter as a covariate. No comparison will be performed during the Follow-up Period. This analysis will be repeated including subjects who switch machine manufacturer type (Lunar or Hologic).

The analysis of percent change in BMD from baseline will be repeated for the following subset of subjects:

- subjects who have a Treatment Month X (e.g., Month 12) scan and a Follow-up Month Y (e.g., Follow-up Month 6) scan and premature discontinue prior to next scheduled assessments.

The number and percentage of subjects with percent change from Baseline to each scheduled assessment during the Follow-up Period in mutually exclusive categories of BMD decrease: $\leq 0\%$, $> 0\% - \leq 1.5\%$, $> 1.5\% - \leq 3\%$, $> 3\% - \leq 4\%$, $> 4\% - \leq 5\%$, $> 5\% - < 8\%$, and $\geq 8\%$ will be summarized by treatment group.

In addition, the number and percentage of subjects with percent change from final on-treatment to each scheduled assessment during the Follow-up Period in mutually exclusive categories of BMD decrease: $\leq 0\%$, $> 0\% - \leq 1.5\%$, $> 1.5\% - \leq 3\%$, $> 3\% - \leq 4\%$, $> 4\% - \leq 5\%$, $> 5\% - < 8\%$, and $\geq 8\%$ will be summarized by treatment group.

For the purpose of assessing post-treatment BMD recovery, the following summaries will be provided. The "recovery" statistic was defined as

Recovery at Follow-up Month X = 100 × ((% change from baseline to final) – (% change from baseline to Follow-up Month X))/(% change from baseline to final)

The "recovery" statistic is the proportion of BMD decrease at the final treatment scan recovered at Follow-up Month X. For example, a subject who has a change of –2% at the end of treatment and has a change of –1% at Follow-up Month 6 has a recovery value of 50% at Follow-up Month 6. This statistic will only be defined for subjects who experienced a decrease from baseline at final treatment.

A continuous summary of recovery at each relevant visit during the Follow-up Period will be provided including mean, standard deviation, median, and a within-group 95% confidence interval for the following subset of subjects:

- All subjects who experienced a decrease from baseline at final treatment;
- Subjects who premature discontinue after Month 6 and on or prior to Month 12;
- Subjects who premature discontinue after Month 12 and on or prior to Month 18;
- Subjects who premature discontinue after Month 18 and on or prior to Month 24;
- Subjects who premature discontinue after Month 24 and on or prior to Month 30;
- Subjects who premature discontinue after Month 30 and on or prior to Month 36;
- Subjects who premature discontinue after Month 36 and on or prior to Month 42;
- Subjects who premature discontinue after Month 42 or completed the Treatment Period.

Additionally, the number and percentage of subjects in each of the following categories will be provided for the subset of subjects listed above: < 0%, 0 – 25%, > 25 – 50%, > 50 – 75%, > 75 – 100%, > 100%.

Bone loss will be defined as > 3% at the spine, > 4% at the total hip, or > 5% at the femoral neck. The following listings will be provided:

- Listing of subjects with bone loss from baseline to final on-treatment visit and have continued bone loss from final on-treatment to Follow-up visits will be provided. Subjects subsequently with an increase from Follow-up Month 6 to a subsequent Follow-up assessment will be excluded from this listing.

In addition, the following plots will be provided:

- Cumulative distribution function of recovery at Follow-up Month 6 and Follow-up Month 12.
- Percent change from baseline to treatment and follow-up months of interest.

9.5.3 Uterine Bleeding

The following analyses will be conducted based on 28-day windows during the placebo-controlled Treatment Period: –28 – –1, 1 – 28, 29 – 56, 57 – 84, 85 – 112, 113 – 140, 141 – 168, 169 – 196, 197 – 224, 225 – 252, 253 – 280, 281 – 308, 309 – 336. For each window, only subjects who were on study drug for the entire window will be included in the analysis.

- Continuous summary (mean, standard deviation, median, minimum, maximum) of the bleeding intensity. This is calculated for any bleeding/spotting days only for subjects who experienced bleeding/spotting. Bleeding categories are represented numerically as follows: Spotting = 1, Light = 2, Medium = 3, Heavy = 4.
- Continuous summary (mean, standard deviation, median, minimum, maximum) of the number of bleeding or spotting days. This is calculated for any subjects who experienced bleeding and/or spotting.
- Continuous summary (mean, standard deviation, median, minimum, maximum) of the number of bleeding only (no spotting) days. This is calculated for any subject who experienced bleeding, excluding those who experienced spotting only with no bleeding days.
- Continuous summary (mean, standard deviation, median, minimum, maximum) of the number of spotting days. This is calculated for any subject who experienced only spotting or no bleeding (i.e., subjects with at least 1 day of spotting and no days of light, medium, or heavy bleeding) in the given 28-day interval.
- Average number of uterine bleeding days by intensity. For each time window, the number of days when subjects report uterine bleeding will be summarized for each bleeding category, i.e., missing, no bleeding, spotting, light, medium

or heavy bleeding, spotting to heavy bleeding combined, and light to heavy bleeding combined.

9.5.4 Amenorrhea

Determination of amenorrhea depends on a subject indicating both no menstrual period and no uterine bleeding as collected in the daily e-Diary. Missing responses for uterine bleeding question will be treated as "No Bleeding." Missing responses for the menstrual period question will be treatment as a response of "No" to the menstrual period question. The determination of amenorrhea also requires data indicating "No Bleeding" and the absence of a menstrual period for a minimum number of 56 days.

The number and percentage of subjects with amenorrhea by month based on a 56-day window (Table 2) will be summarized by treatment group. the numerator is the number of subjects on treatment who did not bleed during the specified time window and the denominator is the number of subjects on drug for the full window. To be considered amenorrheic, a subject must have answered "No Bleeding" at least once during each 28-day period of interest.

Table 2. Monthly Summary of Amenorrhea – 56 days

Basis	Time Interval	Amenorrheic Status
Amenorrheic at least 56 days prior to last dose date	Month 2 (Days 2 – 56)	Subjects did not bleed during Study Days 2 – 56 and subjects answered the e-Diary at least once during Days 2 – 28 and Days 29 – 56 (Truncated, 55 day window)
	Month 3 (Days 29 – 84)	Subjects did not bleed during Study Days 29 – 84 and subjects answered the e-Diary at least once during Days 29 – 56 and Days 57 – 84
	Month 4 (Days 57 – 112)	Subjects did not bleed during Study Days 57 – 112 and subjects answered the e-Diary at least once during Days 57 – 84 and Days 85 – 112
	Month 5 (Days 85 – 140)	Subjects did not bleed during Study Days 85 – 140 and subjects answered the e-Diary at least once during Days 85 – 112 and Days 113 – 140
	Month 6 (Days 113 – 168)	Subjects did not bleed during Study Days 113 – 168 and subjects answered the e-Diary at least once during Study Days 113 – 140 and Days 141 – 168
	Month 7 (Days 141 – 196)	Subjects did not bleed during Study Days 141 – 196 and subjects answered the e-Diary at least once during Days 141 – 168 and Days 169 – 196
	Month 8 (Days 169 – 224)	Subjects did not bleed during Study Days 169 – 224 and subjects answered the e-Diary at least once during Days 169 – 196 and Days 197 – 224
	Month 9 (Days 197 – 252)	Subjects did not bleed during Study Days 197 – 252 and subjects answered the e-Diary at least once during Days 197 – 224 and Days 225 – 252
	Month 10 (Days 225 – 280)	Subjects did not bleed during Study Days 225 – 280 and subjects answered the e-Diary at least once during Study Days 225 – 252 and Days 253 – 280
	Month 11 (Days 253 – 308)	Subjects did not bleed during Study Days 253 – 308 and subjects answered the e-Diary at least once during Study Days 253 – 280 and Days 281 – 308
	Month 12 (Days 281 – 336)	Subjects did not bleed during Study Days 281 – 336 and subjects answered the e-Diary at least once during Days 281 – 308 and Days 309 – 336
	Final (last 56 days including last dose day)	Subject did not bleed during the last 56 days on treatment and answered the e-Diary at least once during the periods [last dose date – 55, last dose date – 28] and [last dose date – 27, last dose date]

9.5.5 Post-treatment Assessment of Menstruation

Time to first post-treatment menses will be summarized based on 28-day windows during the Follow-up Period: Study End Days 1 – 28, 29 – 56, 57 – 84, 85 – 112, 113 – 140, 141 – 168, and ≥ 169. The number and percentage of subjects by who answered "Yes" to the question "Has the subject had a menstrual period since stopping study drug" with an onset date fall into the above time interval will be summarized by treatment group. The denominator is the number of subjects who answered the follow-up assessment of menstruation eCRF at each time interval.

9.5.6 Transvaginal Ultrasound

9.5.6.1 Ovarian Cysts

The presence of ovarian cysts will be noted at Baseline and each relevant post-baseline visit during the placebo-controlled and open-label Treatment Periods. Significant ovarian findings included complex ovarian cyst > 3.5 cm, simple ovarian cyst > 5 cm or endometriomas > 3 cm. The number and percentage of subjects with the pre-defined complex ovarian cysts and simple ovarian cysts will be summarized by treatment group at each relevant visit during the Treatment Period.

For each TAU/TVU assessment, information for significant ovarian findings from the cysts assessment may be available for more than one cyst at more than one ovary location (left and/or right). The subject with multiple cyst findings was counted once in the numerator and denominator when reporting the number and percentage of subjects in the relevant category of significant ovarian findings. A listing included all results across multiple cysts findings from multiple assessments (where available) from subjects who had significant ovarian findings.

9.5.6.2 Endometrial Thickness

The endometrial thickness at each relevant visit during the Treatment Period will be summarized with descriptive statistics. Additionally, analysis of change in endometrial thickness from Baseline to each relevant visit during the Treatment Period will be

performed. The mean change from Baseline each relevant visit during the 12-month placebo-controlled Treatment Period will be compared between each elagolix treatment group and placebo using one-way ANOVA with treatment as the main effect; no comparison will be conducted during the open-label Treatment period.

In addition, the number and percentage of subjects with endometrial thickness of < 8 mm, ≥ 8 mm and ≤ 12 mm, > 12 mm and ≤ 18 mm, and > 18 mm will be summarized by treatment group at each relevant visit during the Treatment Period. Each elagolix group will be compared to placebo using Fisher's exact test during the 12-month placebo-controlled Treatment Period; no comparison will be conducted during the open-label Treatment Period.

9.5.6.3 Uterine Fibroids

The number and percentage of subjects with uterine fibroids present indicated by TVU will be summarized by treatment group at each relevant visit during the placebo-controlled and open-label Treatment Periods. The categories into which the number of fibroids are as follows: 0, 1, 2, 3, 4, 5, > 5, diffuse, and not evaluable. Subjects with multiple uterine fibroids present will be counted once in the numerator and denominator when reporting the number and percentage of subjects with uterine fibroids.

Listing of subjects with uterine fibroids will be presented.

9.5.7 Endometrial Biopsy

The number and percentage of subjects in each category of endometrial biopsy results will be summarized by treatment group at each relevant visit during the placebo-controlled and open-label Treatment Periods. If multiple assessments existed in a specific time window, all assessments will be displayed.

9.5.8 Pregnancy Results

Pregnancies and outcomes will be summarized by treatment group for the Treatment Period (placebo-controlled and open-label) and Follow-up Period. Subjects with

conception date by sponsor more than 30 days after the last dose of study drug will be included in the summaries for the Follow-up Period.

Listings will be prepared of all pregnancy test results for any subject who ever had a positive pregnancy test at any time point during the study.

Positive pregnancy test results (serum, urine, and both) and a listing of all positive pregnancy test results will be provided.

Pregnancy outcome will be summarized by treatment group. Additionally, a listing of pregnancy information will be provided. Pregnancy outcome includes:

- Number of pregnancies
- Live births
  - Pre-term (extremely pre-term: < 28 weeks; very pre-term: 28 - < 32 weeks; moderate to late preterm: 32 - < 37 weeks)
  - Term (37 - < 42 weeks)
  - Post-term (≥ 42 weeks)
- Spontaneous abortion (< 6, 6 - 13, > 13 weeks gestation)
- Termination of pregnancy
- Ectopic pregnancy
- Still birth
- Congenital anomaly
- Lost to follow-up
- Subjects refused to provide information
- Other
- Ongoing (pregnancy ongoing at end of study)
- Maternal exposure to Elagolix/Study drug (0, > 0 - 4, > 4 - 6, > 6 - 12, > 12 - 23, > 23 weeks)
- Embryo/Fetal exposure to Elagolix/Study drug assessed by the sponsor (0, > 0 - 4, > 4 - 6, > 6 - 12, > 12 weeks)

Annualized pregnancy rate during treatment will be reported by treatment group. The annualized pregnancy rate will be calculated as

Annualized pregnancy rate = (number of pregnancies/days of exposure to treatment) × 365

In addition, listings of fetal/newborn, infant post-delivery assessments, and infant during 6 – 12 month follow-up assessment will be provided.

The above information will be summarized primarily based on data collected on the pregnancy case report form. Information not directly obtained from the case report form will use the following definitions.

Gestation day

Gestation Day 0 was calculated by subtracting the gestational age provided according to the first trimester ultrasound date from the first trimester ultrasound date. If no gestational age was provided, it will be imputed by subtracting 14 days from the midpoint between the subject's last negative pregnancy test date and the date of the positive pregnancy test.

Gestational age

Gestational age at delivery was calculated from gestation Day 0 to the date of delivery. Gestational age at abortion was calculated from gestation Day 0 to the date of abortion.

Conception date

Conception date by sponsor was calculated from the gestational age provided from the first trimester ultrasound, by adding 14 days to gestation Day 0. If no gestational age was provided, the conception date by sponsor was imputed as the midpoint between the subject's last negative pregnancy test date and the date of the positive pregnancy test.

Maternal exposure

Maternal exposure to study drug was calculated as the last dose date of study drug – the first dose date of study drug + 1.

Embryo/fetal exposure

Embryo/fetal exposure to study drug by sponsor was calculated from the conception date by site/sponsor to the date of maternal last dose of study drug.

10.0 Overall Type-I Error Control

The primary comparison for the co-primary will be made between elagolix plus E2/NETA and placebo. The elagolix alone arm will be served as a reference arm. Therefore, no adjustment of the type I error rate (alpha) for the primary endpoint is needed. Ranked secondary endpoints will be tested following a fixed-sequence testing procedure. Specifically, testing will begin with testing each of the co-primary endpoints using alpha of 0.05 (2-sided) for elagolix plus E2/NETA compared to placebo. If both co-primary endpoints achieve statistical significance with elagolix W2/NETA as compared to placebo, continued testing will be performed for the ranked secondary endpoints following a fixed-sequence testing procedure.

11.0 Version History

Table 3. SAP Version History Summary

Version	Date	Summary
1.0	29 April 2020	Original version

12.0 References

Appendix A. Receiver Operating Characteristics (ROC) Analysis for Determination of Thresholds for Reduction in Pain

The ROC analysis will be conducted prior to blind break and will contain all randomized subjects who received at least one dose of study drug. Subjects who prematurely discontinue at or before Month 6 will have their pain scores and PGIC response carried forward (LOCF) – from the previous time point when both PGIC and pain score data are observed. Subjects who a) prematurely discontinue at or before Month 6 and have no month with both a PGIC and e-Diary data (DYS, NMPP, or dyspareunia) to carry forward, or b) either prematurely discontinue after Month 6 or complete the Treatment Period but are missing Month 6 PGIC will be assigned a PGIC response of "No Change." Subjects who a) prematurely discontinue at or before Month 6 and have no month with both a PGIC and e-Diary data (DYS, NMPP, or dyspareunia) to carry forward, or b) either prematurely discontinue after Month 6 or complete the Treatment Period but are missing Month 6 pain scores will be assigned a change from baseline in pain score equal to 0.

The threshold for reduction in pain will be determined based on a ROC analysis using the PGIC assessment at Month 6 as an anchor and change from baseline in pain score (DYS, NMPP, or dyspareunia) at Month 6 as the independent variable. The PGIC is a 7-point response scale "Since I started taking the study medication, my endometriosis related pain has: very much improved (1), much improved (2), minimally improved (3), no change (4), minimally worse (5), much worse (6), very much worse (7)." The responses of "much improved" and "very much improved" on PGIC will used to define responders, and the threshold for response (the value of X) will be chosen to balance sensitivity and specificity. That is, it is the value that corresponds to the point on the ROC curve that is closest to the upper left corner, i.e., closest to 100% sensitivity and 100% specificity. No other covariates will be included in the ROC analysis.

The ROC analyses for DYS, NMPP, and dyspareunia will be conducted separately. The threshold obtained from the ROC analysis using PGIC at Month 6 and change from baseline to Month 6 in DYS will be fixed and used for all responder analyses at all time

points during the 12-month placebo-controlled Treatment Period. Same applies to NMPP and dyspareunia.

The results of the ROC analyses can be found in Appendix C.

Appendix B. Definition of Adverse Events of Special Interest (AESI)

Adverse Events of Special Interest (AESI) will be identified using the following search criteria:

Item of Safety Interest	Method of Surveillance
Hot flashes	Non-bone related hypoestrogenic effects CMQ
Bone mineral density loss	Osteoporosis/Osteopenia SMQ
Anemia	Cases are identified through the Non-Hemolytic and Non-Aplastic Anemias CMQ Haematopoietic erythropenia SMQ
Bone Fractures	Osteoporosis/Osteopenia SMQ
Rash and hypersensitivity reactions	Anaphylactic reaction SMQ Severe cutaneous adverse reactions SMQ Drug induced rash CMQ
Lipid abnormalities	Dyslipidemia SMQ
Uterine bleeding change	Female reproductive bleeds CMQ
Endometrial safety	Uterine and fallopian tube neoplasms, malignant and unspecified SMQ Reproductive Premalignant Disorders SMQ
Hypoestrogenic AEs (excluding hot flashes, BMD loss, and fractures)	Non-bone related hypoestrogenic effects CMQ
Spontaneous abortion	Termination of pregnancy and risk of abortion SMQ
Obstetrical complications (maternal and infant)	Pregnancy, labor, and delivery complications and risk factors (excluding abortions and stillbirths) SMQ
Psychiatric events	Depression and suicide/self-injury SMQ
Cardiovascular events	Cardiac arrhythmias SMQ, Cardiomyopathy SMQ, and Ischemic heart disease SMQ
Thromboembolic events	Embolic and thrombotic events SMQ
Elevated hepatic transaminases	Drug related hepatic disorders – comprehensive search SMQ Enhanced pharmacovigilance hepatic terms (elagolix product specific) PMQ

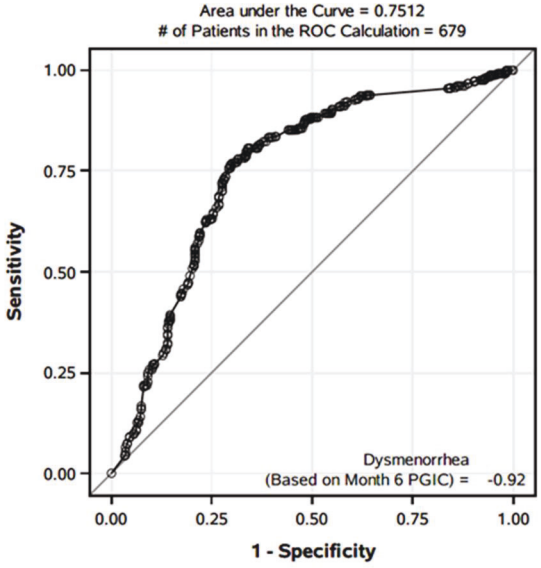
Appendix C. Receiver Operating Characteristics (ROC) Thresholds for Reduction in Pain

Receiver Operating Characteristics (ROC) Thresholds

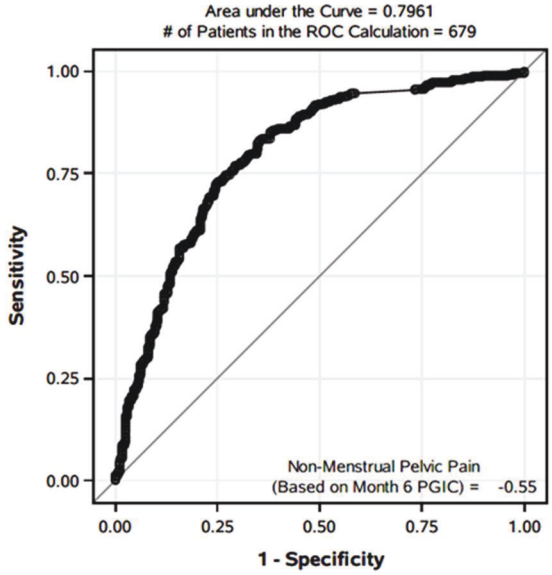
Pain Variable	Measure	ROC Threshold
Dysmenorrhea	Change from Baseline	-0.92
Non-Menstrual Pelvic Pain	Change from Baseline	-0.55
Dyspareunia	Change from Baseline	-0.30

Receiver Operating Characteristic Curves (ROC Curves)

Dysmenorrhea – Change from Baseline to Month 6



Non-Menstrual Pelvic Pain – Change from Baseline to Month 6



Dyspareunia – Change from Baseline to Month 6