Statistical Analysis Plan for Study M16-824

Safety and Efficacy of Elagolix for the Management of Heavy Menstrual Bleeding Associated with Uterine Fibroids in Premenopausal Women

Date: 06 October 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 elagolix Study M16-824 titled "Safety and Efficacy of Elagolix for the Management of Heavy Menstrual Bleeding Associated with Uterine Fibroids in Premenopausal Women."

Study M16-824 examines the efficacy and safety of elagolix for the management of heavy menstrual bleeding associated with uterine fibroids in premenopausal women.

The analyses of pharmacokinetic endpoints, pharmacodynamic endpoints, and biomarker 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 9.3 (SAS Institute Inc., Cary, NC 27513) or later under the UNIX operating system.

This SAP includes changes to analyses based on the changes in study design and procedures in the protocol amendment 4. Details are outlined in Section 13.0.

2.0 Study Design and Objectives

2.1 Objectives and Hypotheses

The objective of this study is to assess the safety and efficacy of elagolix 150 mg QD compared to placebo in reducing HMB associated with uterine fibroids in premenopausal women.

The primary hypothesis is that elagolix 150 mg QD, compared to placebo, reduces HMB associated with uterine fibroids in premenopausal women.

2.2 Study Design Overview

This is a Phase 4, randomized, double-blind, 6-month placebo-controlled, parallel-group, multicenter study. Premenopausal women 18 to 51 years of age with uterine fibroids and HMB, as evidenced by MBL > 80 mL during 1 menses in Screening as measured by the alkaline hematin method, will be selected for this study.

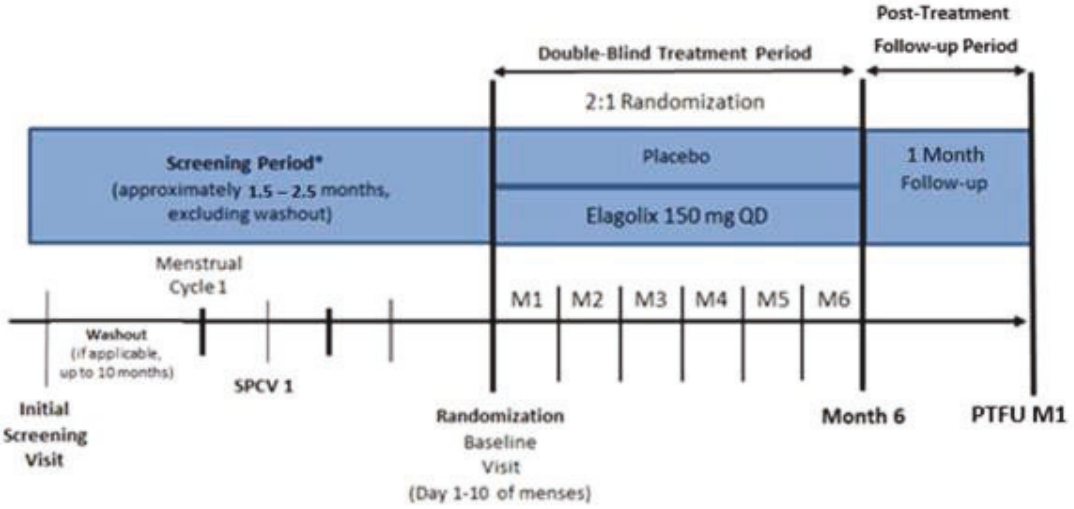
Once eligibility is established during the Screening Period, eligible subjects will be randomly assigned in the double-blind period at the baseline visit to receive either elagolix 150 mg QD or placebo in a 2:1 ratio for 6 months. Following the double-blind Treatment Period, there will be a 1-month Post-Treatment Follow-Up (PTFU) Period. The study will be conducted in approximately 45 sites in the US (including Puerto Rico).

Total study duration, excluding the Washout Period, can be up to 9.5 months (including a Screening Period of approximately 1.5 to 2.5 months prior to Study Day 1 [Baseline]). The study consists 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).
- Screening Period: approximately 1.5 to 2.5 months prior to Study Day 1 (Baseline).
- Treatment Period: duration of 6 months (double-blind).
- PTFU Period: One month following the last dose of study drug. All subjects are expected to enter the 1-month PTFU after completing Treatment Month 6 or once a subject prematurely discontinues from the treatment period.

The schematic of the study is shown in Figure 1.

Figure 1. Study Schematic



M = month; QD = once daily; PTFU = Post-Treatment Follow-up; SPCV = Screening Product Collection Visit

* The Screening Period may be extended in certain circumstances (e.g., heavy menstrual bleeding < 80 mL) by obtaining sponsor approval.

2.3 Treatment Assignment and Blinding

Subjects will be randomly assigned in the double-blind period at the baseline visit to receive either elagolix 150 mg QD or placebo in a 2:1 ratio for 6 months.

2.4 Sample Size Determination

Approximately 48 subjects will be randomly assigned (in a 2:1 ratio) to receive either elagolix 150 mg QD (N = 32) or placebo (N = 16). The sample size will provide at least 80% power to detect a difference between the elagolix 150 mg QD group and the placebo group in the percentage of subjects meeting the primary endpoint on MBL at the Final Month of the placebo-controlled Treatment Period, with a final MBL volume < 80 mL and a reduction in MBL volume of ≥ 50% from Baseline, assuming response rates of 10% for the placebo group and 55% for the elagolix 150 mg QD group with a 2-sided $\alpha = 0.05$. The above sample size was calculated using nQuery Advanced 8.1.

3.0 Endpoints

3.1 Primary Endpoint(s)

The primary endpoint will be the proportion of subjects meeting both of the following conditions:

- Menstrual blood loss (MBL) volume < 80 mL at the Final Month (the last 28 days of treatment), and
- ≥ 50% reduction in MBL volume from Baseline to the Final Month (the last 28 days of treatment).

3.2 Other Efficacy Endpoint(s)

The additional efficacy endpoints are:

- Proportion of subjects with suppression of bleeding (as defined in Section 8.4) at the Final Month and each month during the Treatment Period
- Proportion of subjects with amenorrhea (as defined in Section 8.4) at the Final Month and each month during the Treatment Period
- Change and percent change from Baseline in MBL volume at the Final Month and each month during the Treatment Period
- Change and percent change from Baseline in hemoglobin concentration at each month during the Treatment Period
- Patient Global Impression of Change – Menstrual Bleeding (PGIC-MB) at each month during the Treatment Period
- Change from Baseline in Uterine Fibroid Symptoms Quality of Life (UFS-QoL) at each scheduled assessment.

3.3 Safety Endpoint(s)

The safety endpoints will be based on the following evaluations:

- Adverse events monitoring

- Clinical laboratory testing
- Vital signs measurements
- Electrocardiogram (ECG) variables
- Columbia Suicide Severity Rating Scale (C-SSRS)
- Pregnancy

3.4 Additional Endpoint(s)

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

4.0 Analysis Populations

The following population sets will be used for the analyses.

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 baseline and efficacy 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 presented by the treatment actually received regardless of which treatment group was assigned at the time of randomization. If a subject takes more than 1 treatment, the subject will be analyzed in the safety analysis set as taking the treatment to which she is randomized. All safety analyses will be performed based on Safety Analysis Set.

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 each treatment group:

- Subjects randomized in the study;
- Subjects who took at least one dose of study drug;
- Subjects who completed protocol-specified treatment;
- Subjects who prematurely discontinued study drug (all reasons and primary reason);
- Subjects in each analysis population, as applicable.

For end of study participation, the number and percentage of subjects who completed the protocol defined follow-up period (or did not with associated reasons) will be summarized overall and by treatment group.

6.0 Study Drug Duration

For the safety analysis set population, duration of treatment will be summarized for each treatment group and for both groups combined. Duration of treatment is defined for each subject as last dose date minus first dose date + 1. Duration of treatment will be summarized using the number of subjects treated, mean, standard deviation, median, minimum and maximum. In addition, the number and percentage of subjects treated with study drug for a duration 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.

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

Demographics, baseline or disease characteristics, medical history, and prior and concomitant medications will be summarized for the FAS overall and by treatment group. 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 (≤ 35, > 35 years), BMI (< 25, 25 – < 30, ≥ 30 kg/m^2), tobacco user (current, former, never, unknown), and alcohol user (current, former, never, unknown).

Continuous baseline characteristics include MBL volume, hemoglobin concentration, International Federation of Gynecology and Obstetrics (FIGO) classification,[1] fibroid and uterine volume. Categorical baseline characteristics include Physician Surgery Questionnaire (PSQ), Reason for Study Participation Questionnaire, and hemoglobin concentration (≤ 10.5, > 10.5 – ≤ 12, and > 12 g/dL).

Baseline Fibroid and Uterine Volume

Fibroid and uterine volume data will be summarized at Baseline with data as observed; missing data will not be imputed.

For each subject, the fibroid with the largest volume at Baseline is considered as the primary fibroid for the subject. Baseline average fibroid volume is calculated using the average volume of fibroids with measurements available at Baseline.

FIGO Classification

The number and percentage of subjects with fibroids at each location (e.g., intramural, subserosal, submucosal non-pedunculated, and subserosal pedunculated) according to the FIGO classification[1] will be summarized for each treatment group at Baseline. For each subject, the lowest FIGO classification among all measurements of all fibroids before or on Day 1 (or before Day 8 if there is no measurement before or on Day 1) will be considered as the Baseline FIGO for the subject.

7.2 Medical History

Medical 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 overall and by treatment group. 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). Non-gynecological medical history, gynecological medical history, menstrual and obstetrical history will be summarized and presented separately.

7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name. A prior medication is defined as any medication taken prior to the date of the first dose of study drug. A concomitant medication is defined as any medication that started prior to the date of the first dose of study drug and continued to be taken after the first dose of study drug or any medication that started on or after the date of the first dose of study drug, but not after the date of the last dose of study drug. A medication will be considered a concomitant medication where one of the following three cases occurs (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 prior to or on the last dose of study drug and the end date is missing; (3) both the start date and the end date are missing.

The number and percentage of subjects taking medications will be summarized by generic drug name based on the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications.

8.0 Efficacy Analyses

8.1 General Considerations

All efficacy analyses will be conducted in the FAS Population. All tests will be 2-sided at an alpha level of 0.05.

Unless otherwise specified, categorical data will be summarized by frequency and percentage; descriptive summaries of continuous data will display the mean, SD, median, minimum, and maximum.

For continuous variables, when the analyses of change and/or percent change from Baseline to post-baseline visit(s) are performed, the within-group change from Baseline to each applicable visit will be summarized by treatment group with the mean, SD (or SE), and 95% Confidence Intervals (CIs). The between-group differences will be summarized with the mean, SE, 95% CIs, and P value when applicable. At each post-baseline visit, the Baseline mean and post-baseline visit mean will be calculated for all subjects with baseline and post-baseline value at that visit by treatment group.

Final Month is defined as the last 28 days prior to and including the last dose day.

Unless otherwise specified, there is no hypotheses testing for efficacy endpoints in the Post-Treatment Follow-up Period.

Information regarding corresponding statistical methods for analyses and any additional statistical measures required for a specific variable/endpoint will be provided in relevant sections.

8.2 Handling of Missing Data

Multiple Imputation (MI) will be used in the primary analysis for the primary endpoint.

Missing Final Month MBL volume will be imputed using multiple imputation. First, *M* "semi complete" datasets of monthly MBL volume (Month 1 [Day 2 – Day 28], Month 2

[Day 29 – Day 56], Month 3 [Day 57 – Day 84], Month 4 [Day 85 – Day 112], Month 5 [Day 113 – Day 140], and Month 6 [Day 141 – Day 168]) with monotone missing data pattern will be generated via MCMC option using SAS PROC MI.[2] Then, for each of the *M* "semi complete" datasets with monotone missing data, a "complete" dataset of monthly MBL volume from Month 1 to Month 6 will be generated via MONOTONE REG option using SAS PROC MI. *M* represents the number of imputations and we set $M = 30$. The random seed 12345 will be used.

Mixed-Effect Model Repeat Measurement (MMRM) will be used for the analyses of change and percent change from Baseline in MBL volume to each month. The repeated measures analysis will be conducted using a mixed model including observed measurements at all visits for each month being analyzed. The mixed model includes the categorical fixed effects of treatment, month and treatment-by-month interaction, and the continuous fixed covariate of baseline measurement. The REPEATED statement will be used for month in PROC MIXED with blocks in the covariance matrix identified by subject nested within treatment group. 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).

Missing data will be imputed using the following methods for the sensitivity analyses of the primary efficacy endpoint:

- Non-Responder Imputation (NRI): All subjects who have missing Final Month MBL volume will be considered as non-responders.
- Observed Cases (OC): The OC analysis will not impute values for missing evaluations, and thus subjects who have missing Final Month MBL volume will be excluded from this analysis.

8.3 Primary Efficacy Endpoint(s) and Analyses

8.3.1 Primary Efficacy Endpoint(s)

The primary endpoint is the percentage of responders, defined as subjects meeting the following two conditions:

- MBL volume < 80 mL at the Final Month, and
- 50% or greater reduction in MBL volume from Baseline to the Final Month

Subjects whose primary reason for prematurely discontinuing study drug is "lack of efficacy," or "requires surgery or invasive intervention for treatment of uterine fibroids" will be considered as non-responders.

8.3.2 Handling of Missing Data for the Primary Efficacy Endpoint(s)

See Section 8.2 for the handling of missing data for the primary analysis and sensitivity analyses for the primary efficacy endpoint.

8.3.3 Primary Efficacy Analysis

The primary analysis of the primary endpoint will be performed using the full analysis set population.

If a subject has missing observed Final Month MBL volume, the Final Month MBL volume will be imputed using multiple imputation as described in Section 8.2.

The primary analysis of the primary efficacy endpoint with multiple imputation will be carried out in the following steps:

1. **Multiple Imputation:** 30 "complete" datasets of monthly MBL volume from Month 1 to Month 6 will be generated using SAS PROC MI as described above. The following covariates will be used in the imputation model:
    a. Baseline MBL volume

b. Randomized treatment group
c. Baseline hemoglobin
d. MBL volume in prior months
e. Age of the subject at Baseline

2. **Impute Final Month MBL Volume:** In each of the 30 generated datasets, subject's missing Final Month MBL volume will be imputed using the MBL volume from the "complete" dataset with Month 1 – 6 MBL volume by looking at the corresponding month of the last dose day using analysis time window. Subjects whose last dose day are the same as Study Day 1 will have their Final Month MBL volume imputed using their Month 1 MBL volume.

3. **Impute Responder Status:** The responder status (yes/no) will be derived from "complete" (observed or imputed) Final Month MBL volume, using the criteria as described in Section 8.3.1. Subjects whose primary reason for prematurely discontinuing study drug is "lack of efficacy," or "requires surgery or invasive intervention for treatment of uterine fibroids" will be considered as non-responders.

4. **Analysis:** Each of the 30 imputed datasets will be analyzed separately using a logistic regression model including treatment as the main effect and Baseline MBL volume as a covariate to compare the elagolix treatment group to placebo.

5. **Pooling:** Estimates of the proportions of responders in each treatment group and the difference between the proportions from the 30 imputed datasets obtained from step 3 will be combined into one overall result using PROC MIANALYZE in SAS.[3,4]

8.3.4 Additional Analyses of the Primary Efficacy Endpoint(s)

Sensitivity Analyses

The sensitivity analyses for the primary endpoint use different approaches to handle prematurely discontinued subjects, missing Final Month MBL volume, and AH data from validated versus unvalidated products.

The analysis set used for sensitivity analyses is the full analysis set. Each of the following sensitivity analyses for the primary endpoint will be performed using a logistic regression model including treatment as the main effect and Baseline MBL volume as a covariate to compare the elagolix treatment group to placebo.

Unless otherwise specified, subjects whose primary reason for prematurely discontinuing study drug is "lack of efficacy," or "requires surgery or invasive intervention for treatment of uterine fibroids" will be considered as non-responders.

1. The primary analysis will be repeated with all subjects categorized as responders/non-responders based on observed or imputed MBL volume data only (without taking into account their reasons for premature discontinuation of study drug). Multiple imputation will be performed the same way as in the primary analysis.
2. Non-responder imputation (NRI): All subjects who have missing Final Month MBL volume will be considered as non-responders. No multiple imputation will be performed.
3. Observed cases: The primary analysis will be repeated with the observed Final Month MBL volume. Subjects who have missing Final Month MBL volume will be excluded from this analysis.
4. The primary analysis will be repeated using the total MBL volume collected from validated products only. All subjects will be categorized as responders/non-responders in the same manner as done in the primary analysis (i.e., using the

multiple imputation described in Section 8.3.3) with the exception that all AH data (including that for Baseline MBL volume) will be based on the total MBL volume collected from validated products only.

Additional Analyses

Each of the two response criteria that is a component of the primary endpoint, i.e., (1) MBL volume of < 80 mL at the Final Month; and (2) 50% or greater reduction in MBL volume from Baseline to the Final Month will be analyzed separately in the same way as for the primary analysis. Baseline and Final Month MBL volumes obtained for the primary analysis via MI will be used. The number and percentage of subjects meeting each criterion at the Final Month will be summarized by treatment group. The comparison between the elagolix treatment group and placebo will be made using a logistic regression model including treatment as the main effect and Baseline MBL volume as a covariate.

Besides the method in the primary analysis, the number and percentage of subjects meeting each criterion at the Final Month will also be analyzed in the same way as in each sensitivity analysis for the primary endpoint.

The number and percentage of subjects satisfying either (1) MBL volume < 80 mL, or (2) 50% or greater reduction in MBL volume from Baseline, or both will be summarized for every 28-day interval by treatment group based on observed MBL volume. This summary will also be provided at Final Month of the Treatment Period for subjects who prematurely discontinue study drug due to "lack of efficacy," or "requires surgery or invasive intervention for treatment of uterine fibroids" as the primary reason. Cross-tabulation of the two bleeding criteria above for each 28-day interval in the Treatment Period will be provided by treatment group based on observed MBL volume.

Plots will be provided by treatment group using the MBL volume data for the primary analysis (after multiple imputation):

- Proportion of responders at Final Month.

Additional supportive analysis of the primary efficacy endpoint using Bayesian historical borrowing may be explored. Historical data of placebo subjects from AbbVie M12-815 and M12-817 studies may be incorporated in the analysis based on the consistency and similarity between historical placebo data and current placebo data. Details of the historical borrowing method are provided in Appendix D.

8.4 Additional Efficacy Analyses

MBL Volume

The change and percent change from Baseline to the Final Month in MBL volume obtained from the primary analysis after multiple imputation will be summarized by treatment group and compared between the elagolix treatment group and placebo, using one-way ANCOVA with treatment as the main effect and Baseline MBL volume as a covariate. Baseline and Final Month MBL volumes obtained for the primary analysis will be used.

The comparison of change or percent change from Baseline in MBL volume to each month between the elagolix treatment group and placebo will be performed using a Mixed Model Repeated Measures (MMRM) model with observed MBL volume, as described in Section 8.2.

For the analysis above using observed MBL volume in each 28-day interval, only subjects who did not prematurely discontinue in or before this 28-day interval will be included. For example, for Study Days 29 – 56, it includes subjects who were on treatment for at least 56 days.

Plots will be provided by treatment group using the MBL volume data for the primary analysis (multiple imputation):

- Cumulative distribution function for MBL volume at Final Month;

- Cumulative distribution function for percent reduction from Baseline to Final Month in MBL volume;
- Cumulative distribution function for reduction from baseline to Final Month in MBL volume.

Plots will be provided by treatment group using observed MBL volume data:

- Mean change from Baseline in MBL volume over time.

Amenorrhea

The number and percentage of subjects who achieve amenorrhea will be calculated for each treatment group. The amenorrhea analysis includes subjects on study drug for at least 38 days. For each subject, amenorrhea at the Final Month is defined as having 0 days of bleeding or spotting during the last 28 days prior to and including the last dose day, with the interval starting at or beyond Study Day 11. If a subject does not have evaluable AH data during the last 28 days and no bleeding or spotting or "There was no visible blood on the sanitary products" is indicated on UBQ, then she is considered amenorrheic. The denominator will be the number of subjects with at least 38 days on study drug and having an amenorrhea status at the Final Month. The number and percentage of subjects with amenorrhea at Final Month will be summarized by treatment group and compared between the elagolix treatment group and placebo using the Miettinen-Nurminen (M-N) method.

Time to amenorrhea is defined as the number of days from the first study drug dose date to the day a subject achieves amenorrhea (i.e., the day a subject achieves amenorrhea – the first study drug dose date + 1). For a subject who achieves amenorrhea, the day she achieves amenorrhea is defined as the next day of the last bleeding/spotting day during the Treatment Period. The median, Q1, and Q3 of time to amenorrhea will be calculated for subjects who achieve amenorrhea using Kaplan-Meier method.

In the monthly analysis of bleeding for amenorrhea, the numerator will be the number of subjects on treatment who do not bleed or have spotting during the specified time window specified in Table 1 (but may have begun bleeding or spotting thereafter) and the denominator will be the number of subjects on study drug for the full window and having an amenorrhea status for that month.

Table 1. Monthly Analysis of Bleeding for Amenorrhea

Time Interval	Amenorrheic Status
Month 1 (Days 2 – 28)	Subject do not bleed or have spotting during Study Days 2 – 28 (Truncated, 27-day window)
Month 2 (Days 29 – 56)	Subject do not bleed or have spotting during Study Days 29 – 56
Month 3 (Days 57 – 84)	Subject do not bleed or have spotting during Study Days 57 – 84
Month 4 (Days 85 – 112)	Subject do not bleed or have spotting during Study Days 85 – 112
Month 5 (Days 113 – 140)	Subject do not bleed or have spotting during Study Days 113 – 140
Month 6 (Days 141 – 168)	Subject do not bleed or have spotting during Study Days 141 – 168
Final Month (last 28 days prior to and including the last dose day)	Subject do not bleed or have spotting during the last 28 days prior to and including the last dose day

Incidence and cumulative incidence of amenorrhea by last bleeding/spotting day at each month will be reported using the categories specified in Table 2.

The summary will be provided including all subjects who have at least 38 days on study drug. If a subject meets the criteria for amenorrhea but discontinues, this subject's amenorrheic status will be carried forward to the time points after the subject has discontinued study drug. The denominator will be the number of subjects who have at least 38 days on study drug and have an amenorrhea status at Final Month.

Plots will be provided by treatment group for percentage of subjects with incidence and cumulative incidence of amenorrhea during the Treatment Period, respectively.

In all the analysis for amenorrhea above, if the observed MBL volume over a window (but treat "Subject only had spotting that did not require the use of sanitary products" on UBQ the same as bleeding) is 0 mL for a subject, then the subject is considered as no bleeding/spotting during that interval. Last bleeding/spotting day refers to the last day with observed AH data > 0 mL or the last day with UBQ indicating bleeding or spotting.

Table 2. Categorical Summary of Incidence and Cumulative Incidence of Amenorrhea

Analysis	Time Interval	Numerator Calculation
Incidence	Month 2 (≤ Study Day 56)	Last bleeding/spotting day occurs during Study Days 1 – 28
	Month 3 (Study Day 57 – 84)	Last bleeding/spotting day occurs during Study Days 29 – 56
	Month 4 (Study Day 85 – 112)	Last bleeding/spotting day occurs during Study Days 57 – 84
	Month 5 (Study Day 113 – 140)	Last bleeding/spotting day occurs during Study Days 85 – 112
	Month 6 and after (Study Day ≥ 141)	Last bleeding/spotting day occurs after Study Day 113
Cumulative	Month 2 (≤ Study Day 56)	Last bleeding/spotting day occurs before Study Day 29
	Month 3 (Study Day 57 – 84)	Last bleeding/spotting day occurs before Study Day 57
	Month 4 (Study Day 85 – 112)	Last bleeding/spotting day occurs before Study Day 85
	Month 5 (Study Day 113 – 140)	Last bleeding/spotting day occurs before Study Day 113
	Month 6 and after (Study Day ≥ 141)	Last bleeding/spotting day occurs after Study Day 113

Suppression of bleeding

Suppression of bleeding is defined similarly to amenorrhea except that spotting is allowed. For each subject, achieving suppression of bleeding at the Final Month is defined as having 0 days of bleeding (spotting allowed) during the last 28 days prior to and including the last dose day with the interval starting at or beyond Study Day 11. The

suppression of bleeding analysis includes subjects with at least 38 days on study drug. If a subject does not have evaluable AH data during the last 28 days and no bleeding is indicated on UBQ ("Subject only had spotting that did not require the use of sanitary products" or "There was no visible blood on the sanitary products" is allowed), then the subject will be considered having achieved suppression of bleeding. The denominator will be the number of subjects with at least 38 days on study drug and having a suppression of bleeding status at the Final Month.

The number and percentage of subjects achieving suppression of bleeding at Final Month will be summarized by treatment group and compared between the elagolix treatment group and placebo using the Miettinen-Nurminen (M-N) method.

Time to suppression of bleeding is defined as the number of days from the first study drug dose date to the day a subject achieves suppression of bleeding (i.e., the day a subject achieves suppression of bleeding – the first study drug dose date + 1). For a subject who achieves suppression of bleeding, the day she achieves suppression of bleeding is defined as the next day of the last bleeding day during the Treatment Period. The median, Q1, and Q3 of time to suppression of bleeding will be calculated for subjects who achieve suppression of bleeding using Kaplan-Meier method.

Monthly analysis of bleeding for suppression of bleeding, incidence and cumulative incidence of suppression of bleeding will be provided in the same way as will be done for amenorrhea in Table 1 and Table 2. Plots will be provided by treatment group for percentage of subjects with incidence and cumulative incidence of suppression of bleeding during the Treatment Period, respectively.

In all the analysis for suppression of bleeding above, if the observed MBL volume over a window is 0 mL for a subject, then the subject is considered as no bleeding during that interval; if the observed MBL volume over a window is > 0 mL but there is no single daily observed AH data > 2 mL, the subject is also considered as no bleeding during that interval. Last bleeding day refers to the last day with observed AH data > 2 mL or the last day with UBQ indicating bleeding.

Hemoglobin Concentration

Hemoglobin (Hgb) concentration data will be summarized as observed; missing data will not be imputed.

Baseline Hgb concentration is defined as the last measurement prior to or on the first dose date of study drug. The change and percent change from Baseline in Hgb concentration to each of Months 1 – 6 during the Treatment Period will be analyzed with one-way ANCOVA with treatment group as the main effect and Baseline as a covariate.

PGIC on Menstrual Bleeding Questionnaire

For PGIC on menstrual bleeding (PGIC – MB), the number and percentage of subjects in each response category will be summarized at each applicable visit by treatment group. Missing data will not be imputed. No statistical tests will be performed.

For PGIC – MB, the response categories of "Very Much Improved" and "Much Improved" will be combined and the remaining five categories of the PGIC scale will be combined and labeled "Otherwise." The number and percentage of subjects with response of (a) "Very Much Improved" or "Much Improved," and (b) "Otherwise" will be summarized at each applicable visit by treatment group. Comparisons will be made using the Miettinen-Nurminen (M-N) method for each applicable visit in the Treatment Period.

Uterine Fibroid Symptoms Quality of Life (UFS-QoL)

Improvement in quality of life will be assessed on the UFS-QoL Questionnaire (4-week recall). The change from Baseline to each applicable visit during the Treatment Period will be calculated and summarized by treatment group for each of the UFS QoL subscales (symptom severity, concern, activities, energy/mood, control, self-conscious, and sexual function) and the HRQL total. The change from Baseline to each applicable visit in the Treatment Period will be analyzed using ANCOVA with treatment as the main effect and corresponding Baseline UFS-QoL as a covariate to compare the elagolix treatment group

to placebo. Missing items within a subscale will be handled using the methods specified in the UFS QoL Scoring Manual. Missing data across visits will not be imputed.

8.5 Efficacy Subgroup Analyses

For the primary endpoint, the following subgroups will be explored to assess potential differences in treatment effect across subgroup levels.

- Baseline weight (< median, ≥ median)
- Baseline BMI (< median, ≥ median)
- Baseline BMI (< 25 kg/ m^2, 25 – < 30 kg/m^2, ≥ 30 kg/m^2)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Race (Black or African American, Not Black or African American)
- Baseline MBL volume (< median, ≥ median)
- Baseline FIGO classification as determined by TAU/TVU (0 – 3, 4, 5 – 8)
- Baseline uterine volume as determined by TAU/TVU (< median, ≥ median)
- Baseline primary fibroid volume as determined by TAU/TVU (< median, ≥ median)

Within each level of a subgroup variable, the proportions of responders, difference between proportions, and 95% CIs will be summarized within each subgroup level. No p-value will be provided for subgroup analysis. The data set after multiple imputation for the primary analysis as in Section 8.3.3 will be used for the subgroup analyses. For Baseline BMI, in case of small sample size at a subgroup level, the level of "< 25 kg/ m^2" and "25 – < 30 kg/m^2" may be collapsed.

9.0 Safety Analyses

9.1 General Considerations

Safety data will be summarized for the Safety Analysis Set. Safety summaries will be presented by treatment group. For the safety analysis, subjects are assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized. If

a subject received more than one type of study drug, safety data will be analyzed in the treatment group to which she was randomized to. All safety analyses will be based on observed data. For analyses of safety endpoints, subjects who are missing an evaluation will not be included in the analysis of that particular parameter/visit, unless otherwise specified.

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 AEs are defined as any AE with the onset that is after the first dose of study drug. AEs starting more than 30 days following discontinuation of study drug will not be included in summaries of treatment-emergent AEs. Events where the onset date is the same as the study drug start date are assumed to be treatment-emergent. All treatment-emergent AEs will be summarized overall, as well as by primary MedDRA SOC and PT. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC. The number and percentage of subjects experiencing treatment-emergent AEs will be summarized.

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
- All deaths

9.2.3 Treatment-Emergent Adverse Events by SOC and/or 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.

When summarizing AEs by relationship or severity, if a subject has an event with unknown severity or relationship then the subject will be counted in the severity/relationship category of "unknown" even if the subject has a second occurrence of the same event with a severity/relationship present. The only exception is if the subject has a second occurrence of the same event with the most extreme severity (i.e., "severe") or the most extreme relationship category (i.e., "reasonable possibility"). In this case, the subject will be counted under these most extreme severity/relationship categories.

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

9.2.4 SAEs (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation

Treatment-emergent SAEs (including deaths) and AEs leading to study drug discontinuation will be summarized by SOC and PT and in listing format.

A listing of all deaths will be generated.

9.2.5 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, CMQs, or PMQs). Adverse events of special interest are listed in Appendix B.

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 (e.g., 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.

Laboratory Data in the Treatment Period

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 treatment groups (elagolix vs. placebo). The selected laboratory variables include the following: lipid variables, liver variables (alkaline phosphatase,

ALT, AST, bilirubin), glucose, 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, and triglycerides (TG), 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 either by NCI CTC criteria or categorized as low, normal, or high based on the normal ranges of the laboratory used for each sample. For those analytes that are not reflected in the NCI CTC criteria, exceptions to standard CTC lab grading criteria for elagolix studies will be used as specified in Appendix C. 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 and 4 will be summarized.

The number and percentage of subjects meeting the following criteria will be summarized at Baseline and each applicable visit by treatment group in the Treatment Period:

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

Laboratory abnormalities will be evaluated based on Potentially Clinically Important (PCI) criteria as below. For each laboratory PCI criterion, the number and percentage of subjects who have a laboratory value meeting the criteria will be summarized by treatment

group. Listings will be provided to summarize subject-level laboratory data for subjects meeting PCI criteria.

The PCI criteria for lipid parameters during the Treatment Period are:

- 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.

Plots will be provided by treatment group for HDL-C, LDL-C, triglycerides, and hemoglobin during the 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

Plots will be provided by treatment group for HDL-C, LDL-C, and triglycerides during the Treatment Period:

- Mean percent change from Baseline in lipid parameters over time.

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

Listings of subjects with at least one ALT/AST assessment meeting CTC grade 3 or 4 during the Treatment Period, subjects with at least one HDL C/LDL-C/TG/total cholesterol assessment meeting CTC grade 3 or 4 during the Treatment Period will be provided.

Assessment of Hepatotoxicity

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) will be summarized by treatment group during the Treatment Period to assess potential hepatotoxicity.

- ALT ≥ 3 × ULN and total bilirubin ≥ 2 × ULN
- AST ≥ 3 × ULN and total bilirubin ≥ 2 × ULN
- ALT and AST ≥ 3 × ULN and total bilirubin ≥ 2 × ULN
- ALT ≥ 3 × ULN and total bilirubin < 2 × ULN
- AST ≥ 3 × ULN and total bilirubin < 2 × ULN
- ALT and AST ≥ 3 × ULN and total bilirubin < 2 × ULN
- ALT ≥ 3 × ULN and total bilirubin ≥ 1.5 × ULN
- AST ≥ 3 × ULN and total bilirubin ≥ 1.5 × ULN
- ALT and AST ≥ 3 × ULN and total bilirubin ≥ 1.5 × ULN
- ALT ≥ 3 × ULN, ≥ 5 × ULN, ≥ 10 × ULN, ≥ 20 × ULN
- AST ≥ 3 × ULN, ≥ 5 × ULN, ≥ 10 × ULN, ≥ 20 × ULN
- Total bilirubin ≥ 1.5 × ULN, ≥ 2 × ULN.

The maximum ratio relative to the ULN will be used to determine if subjects meet 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 will be 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.

eDISH (evaluation of drug-induced serious hepatotoxicity) plots will be provided by treatment group for values during the Treatment Period:

- Peak AST vs. peak bilirubin;
- Peak ALT vs. peak bilirubin.

9.4 Analysis of Vital Signs

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

Each vital sign 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 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 treatment groups (elagolix vs. placebo).

The number and percentage of subjects who have PCS vital sign values meeting the following criteria will be summarized by treatment group and a listing of these subjects will be provided. All increase/decrease will be calculated from Baseline to a post-baseline visit in the Treatment Period.

The number and percentage of subjects who have a sustained PCS vital sign value 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 in the Treatment Period.

- Diastolic blood pressure
  - $\leq$ 50 mmHg and $\geq$ 15 mmHg decrease

- > 90 mmHg and ≥ 15 mmHg increase
- ≥ 100 mmHg

- Systolic blood pressure
  - ≤ 90 mmHg and ≥ 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

9.5 Endometrial Biopsy

The number and percentage of subjects in each category of endometrial biopsy results will be summarized at Baseline by treatment group. If multiple assessments exist for a subject, all assessments will be displayed.

9.6 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 by treatment group. Other analyses of C-SSRS only include subjects who have at least 1 post-baseline C-SSRS measurement, regardless of whether they have 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 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 Treatment Period will be summarized. No statistical test 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 provided.

9.7 Pregnancy Results

A listing of subjects with positive pregnancy test results during the Treatment and Post-Treatment Follow-up Periods will be provided. In addition, listings of pregnancy information, fetal/newborn, infant post-delivery assessments, and infant during 6 - 12 month follow-up assessment will be provided.

The annualized pregnancy rate during the Treatment Period will be reported by treatment group.

10.0 Interim Analyses

There is no interim analysis for this study.

11.0 Overall Type-I Error Control

No adjustment of the type I error rate (alpha) for primary analysis of the single primary endpoint is needed.

12.0 Version History

Table 3. SAP Version History Summary

Version	Date	Summary
1.0	02 Dec 2019	Original version
2.0	06 Oct 2020	Updates made based on protocol amendment 4 (dated 12 Jun 2020) • Description of study design and objectives updated to match protocol • Secondary efficacy endpoints removed to match protocol; Change and percent change from Baseline in fibroid and uterine volume at Month 6 removed from other efficacy endpoints since post-baseline TVU/TAU not performed • Safety endpoints updated to match protocol • End-of-Treatment Period analysis removed since not to be performed • Criteria for primary efficacy endpoint updated to match protocol • Efficacy subgroup analyses updated due to clinical relevance of subgroup variables and reduced sample size • Analyses for post-treatment AEs and post-treatment labs removed since Post-Treatment Follow-up changed to only 1 month • Analyses of bone mineral density removed since DXA not performed • Analyses of ovarian cysts and endometrial thickness removed since post-baseline TVU/TAU not performed. Supportive analysis of the primary efficacy endpoint using Bayesian historical borrowing method added to support analysis of the primary endpoint. List of AESIs is updated for consistency with other Elagolix studies.

13.0 References

1. Munro MG, Critchley HO, Broder MS, et al. FIGO Working Group on Menstrual Disorders. FIGO classification system (PALM-COEIN) for causes of abnormal uterine bleeding in nongravid women of reproductive age. Int J Gynaecol Obstet. 2011;113(1):3-13.
2. SAS Institute Inc. 2015. SAS/STAT® 14.1 User's Guide, The MI Procedure. Cary, NC: SAS Institute Inc.
3. SAS Institute Inc. 2015. SAS/STAT® 14.1 User's Guide, The MIANALYZE Procedure. Cary, NC: SAS Institute Inc.
4. Ratitch B, Lipkovich I, O'Kelly M. Combining Analysis Results from Multiply Imputed Categorical Data. PharmaSUG 2013-Paper SP03, 2013.
5. Schlaff WD, Ackerman RT, Al-Hendy A, et al. Elagolix for Heavy Menstrual Bleeding in Women with Uterine Fibroids. N Engl J Med. 2020;382:328-40.
6. Neuenschwander B, Capkun-Niggli G, Branson M, et al. Summarizing historical information on controls in clinical trials. Clin Trials. 2010;7(1):5-18.
7. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177-88.

Appendix A. Protocol Deviations

The number and percentage of subjects who reported at least one of the following protocol deviation categories will be provided.

- Subject entered into the study even though she did not satisfy entry criteria.
- Subject developed withdrawal criteria during the study and was not withdrawn.
- Subject received wrong treatment or incorrect dose of study.
- Subject took prohibited concomitant medication.

Appendix B. Definition of Adverse Events of Special Interest

Item of Safety Interest	Method of Surveillance
Hot flashes/night sweats	Non-bone related hypoestrogenic effects CMQ
Bone mineral density loss	Osteoporosis/Osteopenia SMQ DXA results from clinical trials
Anemia	Cases are identified through the Non-Hemolytic and Non-Aplastic Anemias CMQ Haematopoietic erythropenia SMQ
Bone Fractures	Osteoporosis/osteopenia SMQ Bone fracture CMQ
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 Endometrial biopsy results
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
Teratogenicity	Review of pregnancy outcomes All pregnancies will be followed up to 6 to 12 months post-delivery and reviewed at least quarterly and as they occur
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
Alopecia	Alopecia, diffuse alopecia, androgenic alopecia
Hepatic events	Drug related hepatic disorders comprehensive SMQ, Enhanced pharmacovigilance hepatic terms CMQ
Hormonally mediated malignancies	Breast neoplasms, malignant and unspecified SMQ Uterine and fallopian tube neoplasms, malignant and unspecified SMQ Ovarian neoplasms, malignant and unspecified SMQ

Appendix C. Exceptions to Standard CTC Lab Grading Criteria for Elagolix Studies

ANALYTE	GRADE 1 TOXICITY	GRADE 2 TOXICITY	GRADE 3 TOXICITY	GRADE 4 TOXICITY
HEMATOLOGY				
EOSINOPHIL COUNT INCREASED*	650 – 1500 cells/mm^3	1501 – 5000 cells/mm^3	> 5000 cells/mm^3	No Grade 4
HEMATOCRIT DECREASED**	Decrease in > 0% – 5% below LLN or below BL if BL below LLN	Decrease in > 5% – 10% below LLN or below BL if BL below LLN	Decrease in > 10% below LLN or below BL if BL below LLN	No Grade 4
WHITE BLOOD CELL COUNT INCREASED*	10,800 – 15,000 cells/mm^3	> 15,000 – 20,000 cells/mm^3	> 20,000 – 25,000 cells/mm^3	> 25,000 cells/mm^3
CHEMISTRIES				
BUN**	1.25 – 2.5 × ULN	> 2.5 – 5.0 × ULN	> 5 – 10.0 × ULN	> 10 × ULN
LDL CHOLESTEROL HIGH***	130 – 159 mg/dL	≥ 160 – 189 mg/dL	≥ 190 mg/dL	No Grade 4
HDL CHOLESTEROL LOW***	Low abnormal: < 40 mg/dL***			
PROTEIN, SERUM, LOW*	5.5 – 6.0 g/dL	5.0 – < 5.5 g/dL	< 5.0 g/dL	No Grade 4

* US Department of Health and Human Services Food and Drug Administration Center for Biologics Evaluation and Research. September 27. Guidance for Industry Toxicity Grading Scale for Healthy Adult and Adolescent Volunteers Enrolled in Preventative Vaccine Clinical Trials.

** Elagolix program criteria based on patient population, the disease states under study, and previous clinical trial experience.

*** National Institutes of Health National Cholesterol Education Program; Detection Evaluation and Treatment of High Blood Cholesterol in Adults (Adult treatment Panel III) Final Report NIH Publication No. 02-5215 September 2002.

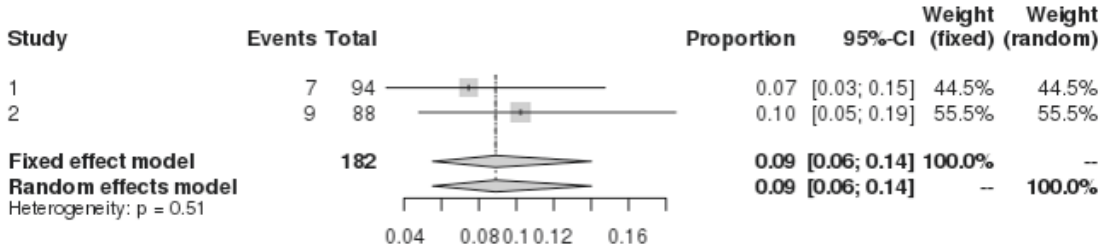
National Cancer Institute's Common Terminology Criteria for Adverse Events (NCI CTC) is applied to all hematology and chemistry analytes across the elagolix program where quantitative criteria are available. In the CTC where similar quantitative values are assigned to multiple grades and a qualitative criterion distinguishes between the grades, the more conservative quantitative grade is applied.

The table above includes those instances where CTC criteria are not provided for applicable lab parameters and alternative references are applied.

Appendix D. Details for Historical Data Borrowing

For the placebo control group, historical data of the observed proportion of responders for the primary endpoint from appropriate studies (M12-815 (n=94, proportion of responders=0.074),[5] M12-817 (n=88, proportion of responders=0.102)[5] were used for meta-analysis-predictive summary.[6] The meta-analytic predicted proportion of responders is 0.09 (95% CI [0.06, 0.14]) based on DerSimonian & Laird estimator.[7] Between-trial variability is small, and the historical effective sample size (the maximum number of subjects we could borrow) is 182.

Figure D-1. Meta-Analysis for Historical Placebo Data



For this supportive analysis of the primary endpoint, the principle for borrowing historical sample size is not to exceed the total sample size in the current study while keeping reasonable type I error rate and power with borrowing. Thus, for placebo control borrowing, assuming approximately 90% of subjects with observed primary endpoint data, the historical placebo borrowing sample size will not exceed 42.

Successful hypothesis testing for the primary endpoint will be claimed from Bayesian approach if the posterior probability of rate difference > 0 is larger than 97.5%, i.e.,:

$$P\left(\theta_{elgolix} - \theta_{Placebo} > 0 \middle| data\right) > 0.975.$$

Table D-1. Bayesian Type I Error (One-Sided) and Power for Elagolix vs. Placebo (sample size = 28:14, placebo arm rate = 10%, elagolix arm rate = 55%)

Control Rate	Bias	Error/Power Borrowing 7 subjects	Error/Power Borrowing 14 subjects	Error/Power Borrowing 21 subjects	Error/Power Borrowing 28 subjects	Error/Power Borrowing 42 subjects
0.055	-0.034	0.008/0.985	0.002/0.992	0.002/0.995	0.002/0.998	0.002/0.998
0.064	-0.025	0.012/0.983	0.003/0.992	0.004/0.995	0.004/0.998	0.004/0.998
0.072	-0.017	0.017/0.981	0.005/0.991	0.006/0.995	0.006/0.998	0.006/0.998
0.081	-0.008	0.022/0.979	0.007/0.991	0.009/0.995	0.009/0.998	0.009/0.998
0.089	0.000	0.028/0.977	0.010/0.991	0.013/0.995	0.013/0.998	0.013/0.999
0.098	0.009	0.033/0.976	0.014/0.991	0.018/0.996	0.019/0.998	0.019/0.999
0.106	0.017	0.038/0.974	0.017/0.991	0.023/0.996	0.025/0.998	0.025/0.999
0.115	0.026	0.043/0.973	0.022/0.991	0.030/0.996	0.033/0.999	0.033/0.999
0.123	0.034	0.046/0.971	0.026/0.991	0.036/0.996	0.040/0.999	0.041/0.999
0.132	0.043	0.050/0.970	0.031/0.991	0.044/0.996	0.050/0.999	0.050/0.999
0.140	0.051	0.053/0.969	0.036/0.991	0.051/0.997	0.059/0.999	0.060/0.999

Table D-1 shows the Bayesian type I error rate (one-sided) and power for different number of borrowing and different bias from -3.4% to 5.1% (corresponding to true unknown placebo group rate from 5.5% to 14.0%) given sample size = 28:14 (assuming 90% of subjects with observed primary endpoint data), placebo arm rate = 10%, and elagolix arm rate = 55%. With the sample size of 28:14 for elagolix and placebo respectively, 14 historical placebo subjects could be borrowed while controlling one-sided type I error rate < 2.5% (when bias is not exceeding 2.6%) and not exceeding the study total sample size. For example, for true placebo arm rate =10.6%, borrowing 14 will control one-sided type I error rate at 1.7%.

Therefore, for this supportive analysis with historical data borrowing, we would borrow 14 historical subjects (smaller than effective sample size 182 and the current study placebo assumed observed sample size 42). *Beta* (1.24, 12.76), with mean 0.09, will be assumed as informative prior for placebo group. For the elagolix group, a Jeffrey's prior for binomial distribution *Beta* (0.5, 0.5) will be assumed.