1.0 Title Page

Statistical Analysis Plan

Study M12-816

Extension Study to Evaluate the Efficacy and Safety of Elagolix in Premenopausal Women with Heavy Menstrual Bleeding Associated with Uterine Fibroids

Date: 24 May 2018

Version 2.0

2.0 Table of Contents

List of Tables

List of Figures

List of Appendices

3.0 Introduction

This statistical analysis plan (SAP) describes the planned statistical analyses for elagolix (ABT-620) Study M12-816 titled "Extension Study to Evaluate the Efficacy and Safety of Elagolix in Premenopausal Women with Heavy Menstrual Bleeding Associated with Uterine Fibroids." The analysis plan was created based on the Study Protocol M12-816 Amendment 1 dated 18 December 2017.

This analysis plan describes both the efficacy and safety analyses. The pharmacokinetic data will be analyzed separately and is not addressed in this SAP.

The SAS System 9.2 or above will be used to perform the statistical analyses. No statistical tests will be performed unless otherwise specified.

4.0 Study Objectives, Design and Procedures

4.1 Objectives

The primary objective of this study is to evaluate the long-term efficacy and safety of elagolix administered alone and in combination with add-back therapy (estradiol/norethindrone acetate or E2/NETA) to reduce heavy menstrual bleeding (HMB) associated with uterine fibroids for up to 12 months (initial 6 months if on active treatment in the pivotal Study M12-815 or Study M12-817 and an additional 6 months in this extension study).

The study will also evaluate the effects of the treatment regimens on hypoestrogenic side effects, changes in bone mineral density (BMD) as assessed by Dual Energy X-Ray Absorptiometry (DXA), and vasomotor symptoms, such as hot flush.

4.2 Study Design

This is a Phase 3, double-blind, multicenter, extension study designed to:

- Obtain 6 month treatment data in subjects who were randomized to placebo in one of the two pivotal studies and subsequently randomized to receive 6 months of treatment with elagolix 300 mg BID or elagolix 300 mg BID plus E2/NETA (estradiol 1 mg/norethindrone acetate 0.5 mg) QD in this study.
- Obtain 12 months of continuous treatment data in subjects who received elagolix 300 mg BID or elagolix 300 mg BID plus E2/NETA (estradiol 1 mg/norethindrone acetate 0.5 mg) QD for 6 months in one of the two pivotal studies and will continue to receive an additional 6 months of the same treatment in this study.
- Obtain data on endometrial health (TAU, TVU and endometrial biopsy).
- Assess bone mineral density, bone health and general safety with long-term treatment and recovery during Treatment and Post-Treatment Follow-Up Period (up to 12 months of Treatment and 12 months Post-Treatment Follow-Up).

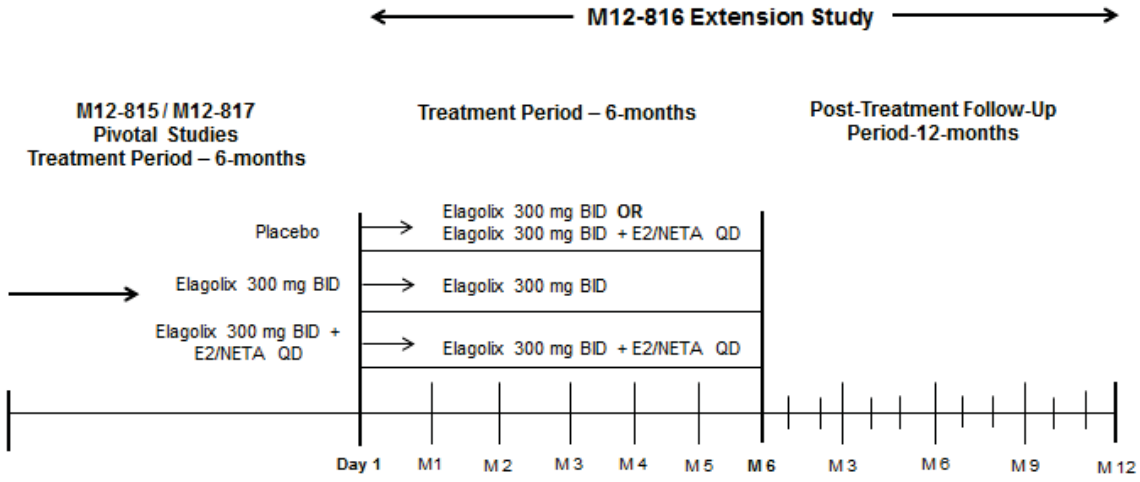
All subjects who completed the 6-Month Treatment Period in their respective pivotal study, who meet eligibility criteria and provide informed consent, will be eligible to enroll in this extension study. This study is designed to enroll approximately 400 subjects across approximately 250 clinical study sites to meet scientific and regulatory objectives without enrolling an undue number of subjects in alignment with ethical considerations.

The study consists of 2 periods as follows:

1. 6-Month Treatment Period
2. 12-Month Post-Treatment Follow-Up Period

Following the 6-Month Treatment Period, subjects will enter into a 12-Month Post-Treatment Follow-Up Period. The Study Periods are illustrated in Figure 1.

Figure 1. Study Schematic



4.3 Sample Size

The two pivotal Studies M12-815 and M12-817 have a planned enrollment of a total of 800 subjects. Based on assumptions related to discontinuation in the pivotal studies and the estimated roll over rate into this extension study, approximately 400 subjects are expected to be enrolled in this extension study.

4.4 End-of-Treatment Period Analysis

An end-of-treatment period analysis of efficacy, demographic and safety variables will be performed after all subjects enrolled in this study complete the 6-Month Treatment Period of this extension Study M12-816. This end-of-treatment period analysis will include all Treatment Period data from all subjects enrolled into the extension study. The database will be versioned and any discrepant data will be clarified before the lock. The analyses will be completed by the Clinical Statistics Department at AbbVie.

5.0 Analysis Populations

5.1 Definition for Analysis Populations

Full Analysis Set

The full analysis set is comprised of all randomized subjects who have received at least one dose of study drug in this study. The data from the full analysis set will be presented by the treatment group assigned at the time of randomization, even if the subject does not receive the correct treatment or does not follow the protocol until completion. The full analysis set will be used for all Baseline and efficacy analyses.

Safety Analysis Set

The Safety Analysis Set will include all randomized subjects who received at least one dose of study drug. The data from the safety analysis set will be presented by the treatment group as treated, independent of treatment group assignment at the time of randomization. If a subject receives more than one treatment, she will be analyzed in the treatment group to which she was randomized. The safety analysis set will be used for all safety analyses.

5.2 Variables for Stratification of Randomization

There is no variable used for stratification of randomization.

5.3 Analysis Groups

Data will primarily be summarized by group for each of the following four main groups of subjects listed here based on time of exposure to elagolix.

- Subjects randomized to elagolix 300 mg BID plus E2/NETA in the pivotal studies and continue on the same treatment in the extension study.
- Subjects randomized to elagolix 300 mg BID alone in the pivotal studies and continue on the same treatment in the extension study.

- Subjects randomized to placebo in the pivotal studies and randomized to elagolix 300 mg BID plus E2/NETA in the extension study.
- Subjects randomized to placebo in the pivotal studies and randomized to elagolix 300 mg BID alone in the extension study.

Where it is noted that the summaries for six analysis groups will be reported, this refers to the four groups above as well as the summaries combining the elagolix treatment groups based on exposure groups, seen below.

- Subjects randomized to placebo in the pivotal studies
- Subjects randomized to active treatment in the pivotal studies.

6.0 Protocol Deviation

Protocol deviations will be summarized and listed for the six analysis groups.

7.0 Analysis Conventions

7.1 Definition of Baseline

Subjects initially randomized to placebo in the pivotal Study M12-815 or Study M12-817 will have their baseline (except for the MBL volume, bleeding days, and some demographics) re-set to the last non-missing assessment collected prior to the first dose of elagolix 300 mg BID or elagolix 300 mg BID plus E2/NETA QD in the extension study, unless otherwise specified. If multiple measurements are recorded on the same day, the average of these measurements will be considered as the baseline value. This same baseline value will be used for the Treatment and Post-Treatment Follow Up Periods. Subjects initially randomized to elagolix 300 mg BID or elagolix 300 mg BID plus E2/NETA QD in the pivotal Study M12-815 or Study M12-817 will have their baseline referring to the baseline from the respective pivotal Study M12-815 or Study M12-817, unless otherwise specified.

MBL Volume and Bleeding Days Baseline

For the MBL volume and bleeding days, the baseline for all subjects will refer to the baseline from pivotal Study M12-815 or Study M12-817.

Fibroid and Uterine Volume Baseline

For subjects initially randomized to placebo in the pivotal Study M12-815 or Study M12-817, baseline for uterine volume, volume of the largest fibroid, total fibroid volume (3 largest fibroids), and endometrial thickness measured by ultrasound (transvaginal [TVU] or transabdominal [TAU]) or Magnetic Resonance Imaging (MRI) will be based on the last non-missing measurement collected on or prior to Study Day 1 in the extension study. If there is no measurement collected on or prior to Study Day 1 in the extension study, the first measurement collected prior to Study Day 8 (Study Day 8 is not included) in the extension study will be used as Baseline.

Adenomyosis Baseline

For subjects initially randomized to placebo in the pivotal Study M12-815 or Study M12-817, baseline for adenomyosis measured by TVU/TAU or MRI will be based on the last non-missing measurement collected on or prior to Study Day 1 in the extension study. If there is no measurement collected on or prior to Study Day 1 in the extension study, the first measurement collected prior to Study Day 15 (Study Day 15 is not included) in the extension study will be used as Baseline.

Quality of Life Questionnaires Baseline

For subjects initially randomized to placebo in the pivotal Study M12-815 or Study M12-817, baseline for quality of life questionnaires including Uterine Fibroid Symptom Questionnaire (UFS-QoL), EuroQoL-5D (EQ-5D-5L) questionnaire, Work Productivity and Activity Impairment (WPAI) questionnaire and Health Care Resource Utilization (HCRU) questionnaire will be based on the measurement collected on or before Study Day 1 in the extension study. For UFS-QoL and EQ-5D-5L, if there is no measurement collected on or before Study Day 1 in the extension study, the

first measurement collected prior to Study Day 8 (Study Day 8 is not included) in the extension study will be used as Baseline.

Demographics

All subjects in the extension study will have their baseline demographics including height, weight, BMI, race, sex, ethinicity, age, alcohol and tobacco use referring to the baseline from the pivotal studies.

When weight is analyzed as a vital sign parameter, subjects initially randomized to placebo in the pivotal studies will have their baseline weight re-set to the last non-missing assessment collected prior to the first dose of elagolix 300 mg BID or elagolix 300 mg BID plus E2/NETA QD in the extension study.

7.2 Definitions of Final Month and Final Visit

Final Month is defined as the last 28 days prior to and including the last dose date in the extension study.

Final Visit is defined as the last non-missing assessment during the Treatment Period in the extension study.

Post-Treatment Final Visit is defined as the last non-missing assessment during the Post-Treatment Follow-up Period in the extension study.

7.3 Definition of Study Days (Days Relative to the First Dose of Study Drug) and Study End Days (Days Relative to the Last Dose of Study Drug)

For randomized subjects who received at least one dose of study drug in the extension study, the study day is defined as the number of days since (positive values) or prior to (negative values) the first study drug dose in the extension study. The day of the first study drug dose in the extension study is defined as Study Day 1 while the last day prior to the first dose of study drug is defined as Study Day –1. There is no Study Day 0. Study end days are calculated based on the number of days relative to the last dose of

study drug in the extension study. The first day following the last dose of study drug in the extension study is defined as Study End Day 1.

7.4 Definition of Analysis Window

7.4.1 Treatment Period

Data obtained more than 3 days after the subject's last dose of study drug in the extension study will be excluded for assessments in the Treatment Period with the following exceptions:

1. For BMD, TVU/TAU, MRI, and endometrial biopsy assessments, data obtained more than 28 days after the last dose of study drug in the extension study will be excluded from summaries/analyses in the Treatment Period.
2. For Alkaline Hematin (AH) data, data obtained after the last dose of study drug in the extension study will be excluded from the summaries/analyses in the Treatment Period.

7.4.2 Post-Treatment Follow-Up Period

Summaries for the Post-Treatment Follow-up Period will include data obtained more than 3 days after the subject's last dose of study drug with the following exceptions:

- For BMD, TVU/TAU, and MRI assessments, data obtained more than 28 days after the last dose of study drug in the extension study will be included in the summaries during the Post-Treatment Follow-up Period.
- For AH data, data obtained after the last dose of study drug in the extension study will be included in the summaries during the Post-Treatment Follow-up Period.

7.4.3 Details Regarding Visit Windows

For analyses in the Treatment Period in the extension study, time points and corresponding time windows are defined based on the time of exposure to study drug starting from Study Day 1 in the extension study.

The time windows described in this section will not be applied to bleeding assessments including AH data and Uterine Bleeding Questionnaire (UBQ). Rules described in Section 7.4.1 and Section 7.4.2 will be applied prior to defining time windows. Data considered as in the Treatment Period are not considered in the Post-Treatment Follow-up Period as described in Section 7.4.1 and Section 7.4.2. Any data considered as Baseline (see Section 7.1) will not be included in any post-baseline windows.

Table 1 will be used for analyses of endpoints collected monthly in the Treatment Period of the extension study.

Table 1. Analysis Time Windows for Non-Bleeding Measurements Collected Monthly

Visit	Nominal Day (Study Day)	Time Window (Study Days)
Month 1	28	> 1 – ≤ 42
Month 2	56	> 42 – ≤ 70
Month 3	84	> 70 – ≤ 98
Month 4	112	> 98 – ≤ 126
Month 5	140	> 126 – ≤ 154
Month 6	168	> 154 – ≤ 196

Table 2 will be used for analyses of endpoints collected every 3 months in the Treatment Period of the extension study. If there is a scheduled Month 1 visit for the endpoint (e.g., PGIC-MB and PGIC-NBUFS), the first subtable will be used. If there is no scheduled Month 1 visit for the endpoint (e.g., TAU/TVU, BMD, urinalysis, apolipoprotein A and B, reflexive TSH, thyroxine-binding globulin (TBG), UFS-QoL), the second subtable will be used.

Table 2. Analysis Time Windows for Non-Bleeding Measurements Collected Every 3 Months in the Treatment Period

Visit	Nominal Day (Study Day)	Time Window (Study Days)
Month 1	28	> 1 – ≤ 56
Month 3	84	> 56 – ≤ 140
Month 6	168	> 140 – ≤ 196
Visit	**Nominal Day (Study Day)**	**Time Window (Study Days)**
Prior to Month 3	28	> 1 – ≤ 56
Month 3	84	> 56 – ≤ 140
Month 6	168	> 140 – ≤ 196

Table 3 will be used for analyses of endpoints collected at Month 6 (biopsy, MRI, creatinine phosphokinase, EQ-5D-5L, WPAI) in the Treatment Period of the extension study.

Table 3. Analysis Time Windows for Non-Bleeding Measurements Collected Every 6 Months in the Treatment Period

Visit	Nominal Day (Study Day)	Time Window (Study Days)
Prior to Month 6	70	> 1 – ≤ 140
Month 6	168	> 140 – ≤ 196

Table 4 will be used for analyses of endpoints collected at Post-Treatment Month 1, 3, 6, 9, and 12 of the extension study (e.g., vital signs, clinical safety labs including chemistry, hematology, lipid panel and urinalysis, apolipoprotein A and B). Table 5 will be used for analysis of endpoints collected at Post-Treatment Month 3 and 6 of the extension study (e.g., TAU/TVU). Table 6 will be used for analysis of endpoints collected at Post-Treatment Month 3 of the extension study (e.g., MRI). Table 7 will be used for analysis of endpoints collected at Post-Treatment Month 6 and 12 of the extension study (e.g., BMD).

Table 4. Analysis Time Windows for Non-Bleeding Measurements Collected at Post-Treatment Month 1, 3, 6, 9, 12

	Nominal Day (Study End Day)	Time Window (Study End Days)
Post-Treatment Month 1	28	> 3 – ≤ 42
Post-Treatment Month 3	84	> 42 – ≤ 126
Post-Treatment Month 6	168	> 126 – ≤ 210
Post-Treatment Month 9	252	> 210 – ≤ 294
Post-Treatment Month 12	336	> 294

Table 5. Analysis Time Windows for Non-Bleeding Measurements Collected at Post-Treatment Month 3 and 6

	Nominal Day (Study End Day)	Time Window (Study End Days)
Post-Treatment Month 3	84	> 28 – ≤ 126
Post-Treatment Month 6	168	> 126 – ≤ 210
Post-Treatment Months 7 – 12	266	> 210

Table 6. Analysis Time Windows for Non-Bleeding Measurements Collected at Post-Treatment Month 3

	Nominal Day (Study End Day)	Time Window (Study End Days)
Post-Treatment Month 3	84	> 28 – ≤ 126
Post-Treatment Months 4 – 12	224	> 126

Table 7. Analysis Time Windows for Non-Bleeding Measurements Collected at Post-Treatment Month 6 and 12

	Nominal Day (Study End Day)	Time Window (Study End Days)
Prior to Post-Treatment Month 6	84	> 28 – ≤ 140
Post-Treatment Month 6	168	> 140 – ≤ 196
Post-Treatment Months 7 – 11	252	> 196 – ≤ 308
Post-Treatment Month 12	336	> 308

7.4.4 Handling of Multiple Assessments in a Specific Time Window

For all parameters except MBL volume, multiple assessments in a specific time window will be handled as follows:

- For parameters other than BMD and WPAI data (or otherwise specified), if more than one assessment is included in a time window, then the assessment performed closest to the scheduled study day (i.e., the nominal day) will be used in the analyses. If more than 1 day is of equal proximity to the nominal day, then the data collected after the nominal day will be used in the analyses.
- For BMD data, if multiple assessments exist in a specific time window, analyses of data will be based on the most conservative (worst) assessment.
- For WPAI data, if multiple assessments exist in a specific time window in the treatment period, analyses of data will be based on the last assessment during the treatment period.

7.5 Dealing with Multiple Values on the Same Day

If multiple measurements are made on the same day for a laboratory parameter or a vital signs parameter, the average of the values will be used in analyses. For summaries of shifts from Baseline and potentially significant values, multiple values on the same day will not be averaged; all values will be considered for these analyses.

If multiple measurements for a particular parameter other than a laboratory parameter and a vital signs parameter are made on the same day for the same subject, the most conservative (worst) measurement will be used in analyses.

7.6 Definitions and Conventions of MBL Volume and Bleeding Days

7.6.1 MBL Volume

The MBL volume is based on validated and non-validated sanitary products unless otherwise specified (for some sensitivity analysis, the MBL volume will be based on validated sanitary products only).

7.6.1.1 Observed MBL Volume Over a Window

If there were observed AH data between two study visits (Product Collection Visit or site visit with UBQ completed), then the MBL volume on the days with observed AH data will be the observed AH data and the MBL volume for the rest of days between the two study visits will be imputed as 0 by AH data.

UBQ responses indicate if a subject has any bleeding or spotting since the last study visit (Product Collection Visit or site visit with UBQ completed). If observed evaluable AH data is available over a window, then the UBQ will not be considered.

If the AH data is reported as "NVB" (no visible blood, discarded without assay), "NO BLEED" (no collection due to no bleeding as indicated by subject or clinical site on assay requisition form), or "TL BQL" (total below low limit of quantitation), the numeric value for the AH data will be imputed as 0. If the AH data is reported as other character value ("NULL," "IN ERROR," "NO DATA," and etc.), the corresponding numeric value for the AH data will be set as missing.

The observed MBL volume over a window (such as the Final Month, Month 1 [Study Days 1 – 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 [Study Days 141 – 168]) is defined as follows:

- If a subject has any evaluable AH data reported in the window, then the subject's MBL volume in this window will be the total combined observed AH data of validated and non-validated sanitary products over this window.
- If a subject is missing AH data or all AH data are unevaluable in the window and the UBQ is completed and indicates no bleeding or spotting, "Subject only had spotting that did not require the use of sanitary products" or "There was no visible blood on the sanitary products" and covers this window, then:
  - A value of 0 will be assigned to the MBL volume over this window.
- If a subject is missing AH data or all AH data are unevaluable in the window and the UBQ is completed with no bleeding or spotting or "Subject only had spotting that did not require the use of sanitary products" or "There was no visible blood on the sanitary products" indicated but UBQ does not cover the full window, then:
  - A value of 0 will be assigned to the MBL volume over this window if the days not covered by UBQ response as described above have AH imputed MBL volume (see first paragraph in Section 7.6.1.1) being 0 for those days.

If there is still no value assigned after the imputation procedure described above, the observed MBL volume over the window is considered missing.

7.6.2 Bleeding Days

A bleeding/spotting day is defined as a day having daily MBL volume of > 0 mL.

A spotting day is defined as a day having daily MBL volume of > 0 – 2 mL.

A bleeding day is defined as a day having daily MBL volume of > 2 mL.

Bleeding intensity categories are

- Daily MBL volume of > 2 – 10 mL.
- Daily MBL volume of > 10 – 40 mL.
- Daily MBL volume of > 40 – 80 mL.

- Daily MBL volume of > 80 mL.

Please note that only observed AH data will be used for daily MBL volume categroies above.

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

8.1 Demographic and Baseline Characteristics

Demographics and Baseline characteristics will be summarized for all randomized subjects who received at least one dose of study drug by the six analysis groups listed in Section 5.3. The number of missing values will also be summarized.

No statistical tests will be performed.

8.1.1 Demographics

Continuous variables such as age, height, weight, and BMI will be summarized with the mean, standard deviation (SD), median, minimum, and maximum by analysis group. Number and percentage will be computed by analysis group for the following demographic parameters: age group (18 – < 20 years old, 20 – < 25 years old, 25 – < 30 years old, 30 – < 35 years old, 35 – < 40 years old, 40 – < 45 years old, 45 – < 50 years old, ≥ 50 years old), BMI (≤ 18.5 kg/m^2, > 18.5 kg/m^2 – < 25 kg/m^2, 25 kg/m^2 – < 30 kg/m^2, 30 kg/m^2 – < 35 kg/m^2, 35 kg/m^2 – < 40 kg/m^2, ≥ 40 kg/m^2), sex, race, ethnicity, tobacco use (current, former, never, and unknown), and alcohol use (current, former, never, and unknown).

The number and percentage of subjects with Baseline presence of adenomyosis (as defined in Section 11.3.5.3) will be summarized.

8.1.2 Baseline Characteristics

Menstrual Blood Loss (MBL)

Baseline MBL measured by AH method as defined in Section 7.1 will be summarized with mean, SD, median, minimum, and maximum values.

Bleeding Days

Baseline for bleeding days will be based on the observed daily AH data as defined in Section 7.1 and Section 7.6. The numbers of bleeding days by intensity categories at Baseline will be summarized with mean, SD, median, minimum, and maximum.

Hemoglobin Concentration

Baseline hemoglobin concentration will be summarized with mean, SD, median, minimum, and maximum values. The number and percentage of subjects with Baseline hemoglobin concentration in the following categories will be summarized: ≤ 10.5 g/dL, $> 10.5 - \leq 12$ g/dL, and > 12 g/dL.

Fibroid and Uterine Volume

Baseline uterine volume, Baseline volume of the largest fibroid, and Baseline total fibroid volume (3 largest fibroids) will be summarized with mean, SD, median, minimum, and maximum.

8.2 Medical History

Medical/surgical, gynecological, menstrual and obstetrical history will be summarized and presented using system organ classes (SOCs) and preferred terms (PTs) from the Medical Dictionary for Regulatory Activities (MedDRA) Version 19.0 or higher for each of the six analysis groups listed in Section 5.3. Subjects reporting more than one medical history within a SOC will be counted only once for that SOC. Subjects reporting more than one medical history for a PT will be counted only once for that PT.

8.3 Prior, Concomitant, and Post-Treatment Medications

Prior, concomitant, and post-treatment medications will be summarized for the full analysis set for subjects who enrolled in the extension study by the six analysis groups listed in Section 5.3.

Prior medications based on data collected prior to the pivotal studies were summarized.

For prior medications administered to treat uterine fibroid symptoms, the following data are collected: dates of administration (including start and stop dates), dose, route, and reason for discontinuation. Prior medications will be summarized for all randomized subjects who received at least one dose of study drug with number and percentage for each analysis group using ATC Classification and preferred terms from the World Health Organization (WHO) Drug Dictionary.

Concomitant medications are those medications, other than study drugs, taken during the treatment period of the extension study with an end date on or after the first dose of study drug or ongoing at the end of study, and a start date prior to or on 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 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.

Concomitant medications for all randomized subjects who received at least one dose of study drug will be summarized using ATC Classification and preferred terms from the World Health Organization (WHO) Drug Dictionary with number and percentage for each analysis group.

Other medications taken during the Post-Treatment Follow-up Period of this extenstion study, which includes all medications with an end date after the last dose of study drug or ongoing at the end of study, will be summarized by ATC Classification and WHO preferred term with number and percentage for each analysis group.

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. Prior medications, concomitant medications, and other medications during the Post-Treatment Follow-up Period will be summarized separately.

9.0 Subject Disposition

The number of subjects for each of the following categories will be summarized by the six analysis groups listed in Section 5.3:

- All randomized subjects
- Subjects who took at least one dose of study drug
- Subjects who completed the Treatment Period
- Subjects who discontinued from the Treatment Period
- Subjects who completed the Post-Treatment Follow-Up Period
- Subjects who prematurely discontinued from the Post-Treatment Follow-Up Period

Premature discontinuation of the study drug by primary reason and by any reason will be summarized for each of the six analysis groups, with number and percentage and by reason for discontinuation for all subjects who received at least one dose of study drug. Subjects may have multiple reasons for prematurely discontinuing study drug, but will be counted no more than once for the total of "Any Reason."

Premature discontinuation of the study by primary reason and by any reason will be summarized for each of the six analysis groups, with number and percentage and by reason for discontinuation. Subjects may have multiple reasons for prematurely discontinuing during the Post-Treatment Follow-up Period, but will be counted no more than once for the total of "Any Reason."

10.0 Study Drug Exposure

For all subjects who enrolled into the extension study, the duration of study drug exposure in the extension study is defined as the difference between the dates of the first and last doses of the treatment in the extension study plus 1 day. For subjects randomized to active treatment in the pivotal Study M12-815 or Study M12-817, cumulative exposure to study drug in the pivotal studies and the extension study will be the treatment duration of pivotal studies and extension study by taking the sum of the exposure during pivotal studies and the exposure during the extension study. If a subject has a study drug end date in the pivotal study that is after the study drug start day in the extension study, the overlapping days will be only counted once.

Treatment exposure will be summaried for the extension study alone, for the pivotal studies and the extension study, and for the pivotal studies alone, separately. In each summary, the duration (days) of treatment will be summarized with the mean, standard deviation, median, minimum, and maximum for each of the four main analysis groups listed in Section 5.3 and overall. Additionally, the number and percentage of subjects exposed to study drug will be summarized by duration intervals.

In the cumulative summary of the pivotal and extension studies, for subjects randomized to placebo in the pivotal Study M12-815 or Study M12-817, exposure will be summarized only for the treatment duration of extension study. In the summary of exposure in the pivotal studies, for placebo subjects in the pivotal studies, the time on placebo will be summarized.

The duration intervals for the cumulative summary are: 1 – 28 days, > 28 – 56 days, > 56 – 84 days, > 84 – 112 days, > 112 – 140 days, > 140 – 168 days, > 168 – 196 days, > 196 – 224 days, > 224 – 252 days, > 252 – 280 days, > 280 – 308 days, > 308 – 336 days, and > 336 days. The duration intervals for the privotal Studies M12-815 and M12-817 alone, and for the extension Study M12-816 alone are: 1 – 28 days, > 28 – 56 days, > 56 – 84 days, > 84 – 112 days, > 112 – 140 days, > 140 – 168 days, > 168 – 196 days, > 196 – 224 days, > 224 – 252 days.

11.0 Efficacy Analysis

11.1 General Considerations

No statistical tests will be performed.

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 relevant visit will be summarized with the mean, SD, and 95% Confidence Intervals (CIs). 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.

Unless otherwise specified, summaries in efficacy analysis will be provided by the four main groups as in Section 5.3.

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

11.2 Primary Efficacy Analysis

11.2.1 Primary Efficacy Endpoint

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

- MBL volume < 80 mL at the Final Month (the last 28 days of treatment), and
- 50% or greater reduction in MBL volume from Baseline to the Final Month (the last 28 days of treatment)

Subjects who prematurely discontinue study drug due to "lack of efficacy," "requires surgery or invasive intervention for treatment of uterine fibroids," or adverse events will be considered as non-responders, regardless if the two conditions above are met or not. Only primary discontinuation reason is considered.

Baseline MBL volume is defined in Section 7.1.

Final Month is defined as in Section 7.2. If a subject has less than or equal to 28 days of treatment, then the Final Month will be the interval between Day 1 to the last dose date.

11.2.2 Primary Analysis of Primary Efficacy Endpoint

The primary analysis of the primary endpoint will be performed using the full analysis set (Section 5.1) in the extension study.

The responder status (yes/no) will be derived from the observed Final Month MBL volume (as in Section 7.6.1.1), using the criteria as described in Section 11.2.1.

For the primary efficacy analysis, the number and proportion of responders meeting the criteria for the primary efficacy endpoint will be summaried by analysis group as observed. The 95% CI using the normal approximation to the binomial distribution will also be provided for the proportion of responders in each analysis group.

11.2.3 Sensitivity Analyses of the Primary Efficacy Endpoint

The following sensitivity analyses for the primary endpoint will be performed using the full analysis set:

- The primary analysis will be repeated with all subjects categorized as responders/non-responders based on observed MBL volume data only (without taking into account their reasons for premature discontinuation of study drug).
- 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., subjects who prematurely discontinue study drug due to "lack of

efficacy," "requires surgery or invasive intervention for treatment of uterine fibroids," or adverse events will be considered as non-responders) with exception that all AH data are based on the total MBL volume collected from validated products only.

11.2.4 Adenomyosis Subset

The proportion of responders will be summarized by analysis group for subjects with Baseline adenomyosis present, and repeated for the subjects with adenomyosis at any time during the Baseline or Treatment Period (see definition in Section 11.3.5.3).

11.3 Secondary and Other Efficacy Analyses

Secondary efficacy endpoints during the Treatment Period include the following:

- Change and percent change from Baseline in MBL volume to each month and to the Final Month;
- Percentage of subjects with suppression of bleeding (no bleeding allowed, spotting allowed) at the Final Month;
- Percentage of subjects with Baseline hemoglobin ≤ 10.5 g/dL who have an increase in hemoglobin > 2 g/dL at Month 6.

Other efficacy endpoints during the Treatment Period include the following:

- Percentage of subjects with MBL volume < 80 mL and 50% or greater reduction in MBL volume from Baseline during 28-day intervals throughout the Treatment Period;
- Percentage of subjects with amenorrhea;
- Percentage of subjects with control of bleeding;
- The number of bleeding days;
- Change and percent change from Baseline in hemoglobin concentration;
- Patient Global Impression of Change (PGIC) questionnaire for Menstrual Bleeding and Non-Bleeding Uterine Fibroid Symptoms;

- Change and percent change from baseline in fibroid and uterine volume;
- Change from baseline for the UFS-QoL;
- Change from baseline for the EuroQoL-5D (EQ-5D-5L);
- The Health Care Resource Utilization (HCRU) questionnaire;
- Change from baseline for the WPAI.

11.3.1 Reduction of Bleeding

The change and percent change from Baseline to the observed Final Month in MBL volume will be summarized by analysis group.

The change and percent change from Baseline to each 28-day window starting at Study Day 1 up to Study Day 168 in observed MBL volume in the Treatment Period (i.e., Month 1 (Study Days 1 – 28), Month 2 (Study Days 29 – 56), Month 3 (Study Days 57 – 84), Month 4 (Study Days 85 – 112), Month 5 (Study Days 113 – 140), and Month 6 (Study Days 141 – 168)) will be summarized by analysis group. For 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.

Each of the two response criteria that are components 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 summarized separately in the same way as for the primary analysis. The number and percentage of subjects meeting each criterion at the Final Month will be summarized by analysis group.

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, and in the analysis for subjects identified in the adenomyosis subset of interest.

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 analysis group based on observed MBL volume.

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

- Cumulative distribution function for MBL volume at Final Month;
- Cumulative distribution function for percent change from Baseline to Final Month in MBL volume;
- Proportion of responders at Final Month;
- Cumulative distribution function for change from baseline to Final Month in MBL volume.
- Cumulative distribution function for change from baseline in MBL volume to each month;
- Mean change in MBL volume over time.

11.3.2 Amenorrhea, Suppression and Control of Bleeding

11.3.2.1 Amenorrhea

The number and percentage of subjects who achieved amenorrhea will be calculated for each analysis group. Amenorrhea analysis will include subjects on study drug for at least 38 days in this extension study. For each subject, amenorrhea is defined as having 0 days of bleeding or spotting during the last 28 days of treatment with the interval starting from Study Day 11. If a subject did 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" was indicated on UBQ, then she is considered amenorrheic.

For subjects initially randomized to placebo in the pivotal Study M12-815 or Study M12-817, time to amenorrhea is defined as the number of days from the first study drug dose date to the day a subject achieving cumulative amenorrhea (i.e., the day a subject achieving amenorrhea – the first study drug dose date + 1). For a subject who achieved amenorrhea, the day she achieved 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 using Kaplan-Meier method.

In the monthly analysis of bleeding for amenorrhea, the numerator is the number of subjects on treatment who did not bleed during the specified time window (but may have begun bleeding thereafter) and the denominator includes subjects on drug for the full window. For the final visit, the denominator includes subjects with at least 38 days on study drug.

Table 8. Monthly Analysis of Bleeding for Amenorrhea

Time Interval	Amenorrheic Status
Month 1 (Days 1 – 28)	Subject did not bleed during Study Days 1 – 28
Month 2 (Days 29 – 56)	Subject did not bleed during Study Days 29 – 56
Month 3 (Days 57 – 84)	Subject did not bleed during Study Days 57 – 84
Month 4 (Days 85 – 112)	Subject did not bleed during Study Days 85 – 112
Month 5 (Days 113 – 140)	Subject did not bleed during Study Days 113 – 140
Month 6 (Days 141 – 168)	Subject did not bleed during Study Days 141 – 168
Final Month (last 28 days of treatment)	Subject did not bleed during the last 28 days of treatment

Incidence and cumulative incidence of amenorrhea by last bleeding/spotting day at each month will be reported using the categories specified in in Table 9.

The summary will be provided in the following 2 ways: (1) including only subjects who have at least 168 days on study drug in the extension study. The denominator includes subjects who have at least 168 days on study drug in the extension study for all windows. (2) LOCF: including all subjects who have at least 38 days on study drug in the extension study. If a subject meets the criteria for amenorrhea but discontinues, this subject's

amenorrheic status was carried forward to the time points after the subject has discontinued study drug. The denominator includes subjects who have at least 38 days on study drug in the extension study.

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

Table 9. Categorical Summary of Incidence and Cumulative Incidence of Amenorrhea by Last Bleeding/Spotting Day

Analysis	Time Interval	Numerator Calculation
Incidence	Month 2 (≤ Study Day 56)	Last bleeding/spotting day occurred during Study Days 1 – 28
	Month 3 (Study Day 57 – 84)	Last bleeding/spotting day occurred during Study Days 29 – 56
	Month 4 (Study Day 85 – 112)	Last bleeding/spotting day occurred during Study Days 57 – 84
	Month 5 (Study Day 113 – 140)	Last bleeding/spotting day occurred during Study Days 85 – 112
	Month 6 (Study Day 141 – 168)	Last bleeding/spotting day occurred during Study Days 113 – 140
Cumulative	Month 2 (≤ Study Day 56)	Last bleeding/spotting day occurred before Study Day 29
	Month 3 (Study Day 57 – 84)	Last bleeding/spotting day occurred before Study Day 57
	Month 4 (Study Day 85 – 112)	Last bleeding/spotting day occurred before Study Day 85
	Month 5 (Study Day 113 – 140)	Last bleeding/spotting day occurred before Study Day 113
	Month 6 (Study Day 141 – 168)	Last bleeding/spotting day occurred before Study Day 141

11.3.2.2 Suppression of Bleeding

Suppression of bleeding is defined similarly to amenorrhea as in Section 11.3.2 except that spotting is allowed. For each subject, achieving suppression of bleeding is defined as having 0 days of bleeding during the last 28 days of treatment with the interval starting

from Study Day 11. The suppression of bleeding analysis will include subjects with at least 38 days on study drug in the extension study. If a subject did not have evaluable AH data during the last 28 days and no bleeding was indicated on UBQ ("Subject only has spotting that did not require the use of sanitary products" or "There was no visible blood on the sanitary products" was allowed), then the subject is considered having achieved suppression of bleeding. The number and percentage of subjects achieving suppression of bleeding will be summarized by analysis group.

For subjects initially randomized to placebo in the pivotal Study M12-815 or Study M12-817, time to suppression of bleeding is defined as the number of days from the first study drug dose date to the day a subject achieving suppression of bleeding (i.e., the day a subject achieving suppression of bleeding – the first study drug dose date + 1). For a subject who achieved suppression of bleeding, the day she achieved 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 using Kaplan-Meier method.

Monthly analysis of bleeding for suppression of bleeding, incidence and cumulative incidence of suppression of bleeding by last bleeding day will be provided in the same way as was done for amenorrhea in Table 8 and Table 9. Plots will be provided by analysis group for percentage of subjects with incidence and cumulative incidence of suppression of bleeding during the Treatment Period, respectively.

11.3.2.3 Control of Bleeding

For each subject, achieving control of bleeding is defined as having 0 days of bleeding and up to 1 day of spotting during the last 28 days of treatment with the interval starting from Study Day 11. The control of bleeding analysis will include subjects with at least 38 days on study drug in the extension study. If a subject did 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" was indicated on UBQ, then the subject is considered having achieved

control of bleeding. If the subject indicates bleeding or spotting on the UBQ, the subject does not meet the criteria for control of bleeding.

The number and percentage of subjects achieving control of bleeding will be summarized by analysis group.

11.3.3 Bleeding Days

The numbers of bleeding days by intensity categories (Section 7.6.2) will be calculated for each 28-day interval starting at Study Day 1 up to Study Day 168 in the Treatment Period (i.e., Study Days 1 – 28, Study Days 29 – 56, Study Days 57 – 84, Study Days 85 – 112, Study Days 113 – 140, and Study Days 141 – 168) and the first 28-day interval in the Post-Treatment Follow-up Period (i.e., Study End Days 1 – 28) for each analysis group. For each 28-day interval, each analysis includes subjects who did not premature discontinue in or before this 28-day interval. For example, for Study Days 29 – 56, it includes subjects who were on treatment for at least 56 days in the extension study. For Final Month analysis, it includes subjects who were on treatment for at least 38 days in the extension study. The summary of average number of bleeding days in a 28 day window during the Treatment Period will be provided for each analysis group by intensity categories.

If a subject did not have evaluable AH data during an interval and no bleeding was indicated on UBQ ("Subject only has spotting that did not require the use of sanitary products" or "There was no visible blood on the sanitary products" was allowed), then the number of bleeding days over the interval is 0. The number and percentage for each analysis group and overall will be provided for subjects who did not have evaluable AH data during an interval and bleeding was indicated on UBQ.

The change and percent change from Baseline in the numbers of bleeding/spotting days and bleeding days to each 28-day interval and Final Month in the Treatment Period and the first 28-day interval in the Post-Treatment Follow-up Period will be summarized for each analysis group, respectively.

11.3.4 Hemoglobin Concentration

Hemoglobin (Hgb) concentration data will be summarized as observed.

The change and percent change from Baseline in Hgb concentration to Months 1 – 6 during the Treatment Period will be summarized by analysis group.

The number and percentage of subjects with changes in Hgb concentration from Baseline to Months 1 – 6 during the Treatment Period in each of the following categories will be summarized by analysis group.

- Change from Baseline in Hgb ≤ –1.0 g/dL
- –1.0 g/dL < Change from Baseline in Hgb ≤ –0.5 g/dL
- –0.5 g/dL < Change from Baseline in Hgb < 0.5 g/dL
- 0.5 g/dL ≤ Change from Baseline in Hgb < 1.0 g/dL
- 1.0 g/dL ≤ Change from Baseline in Hgb < 1.5 g/dL
- 1.5 g/dL ≤ Change from Baseline in Hgb < 2.0 g/dL
- Change from Baseline in Hgb ≥ 2.0 g/dL

Also, shift tables from Baseline to Month 3 and Month 6 will be summarized by the following categories.

- Hgb ≤ 10.5 g/dL
- 10.5 g/dL < Hgb ≤ 12 g/dL
- Hgb > 12 g/dL

The number and percentage of subjects who had Hgb Baseline ≤ 10.5 g/dL and have an increase in Hgb concentration > 2 g/dL from Baseline will be summarized for each month by analysis group.

The number and percentage of subjects who had Hgb Baseline ≤ 10.5 g/dL and have an increase in Hgb concentration > 1 g/dL from Baseline will be summarized for each month by analysis group.

In addition, the number and percentage of subjects who have an increase in Hgb concentration > 1 g/dL from Baseline will be summarized for each month by analysis group and Baseline Hgb concentration categories as follows. Same summary will be provided for increase in Hgb concentration > 2 g/dL from Baseline to each month.

- Hgb ≤ 10.5 g/dL
- 10.5 g/dL < Hgb ≤ 12 g/dL
- Hgb > 12 g/dL

11.3.5 Fibroid and Uterine Volume, FIGO Classification, Adenomyosis, and Post-Treatment Menses

11.3.5.1 Fibroid and Uterine Volume

Fibroid and uterine volume data will be summarized as observed. Analyses will be conducted separately for results obtained from TAU/TVU and MRI.

The change and percent change in the volume of the largest (primary) fibroid, the total fibroid volume (3 largest fibroids), and the uterine volume from Baseline to Month 3 (if applicable) and Month 6 during the Treatment Period will be summarized for each analysis group. The change and percent change at Post-Treatment Month 3 and Post-Treatment Month 6 (if applicable) during the Post-Treatment Follow-up Period will also be summarized for each analysis group.

The number and percentage of subjects with ≥ 25% reduction from Baseline in total fibroid volume at Month 3 (if applicable) and Month 6 each during the Treatment Period will be summarized for each analysis group. The same summary will be provided for largest (primary) fibroid volume and uterine volume, respectively.

11.3.5.2 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[3] will be summarized for each analysis group at Baseline, Month 3 (if

applicable) and Month 6 during the Treatment Period and Post-Treatment Month 3 and Post-Treatment Month 6 (if applicable) during the Post-Treatment Follow-up Period.

11.3.5.3 Adenomyosis

The number and percentage of subjects with presence of adenomyosis will be summarized for each analysis group at Baseline and each relevant post-baseline visit in the Treatment and Post-Treatment Follow-up Periods. Analyses above will be conducted separately for results obtained from TAU/TVU and MRI.

For subjects initially randomized to placebo in the pivotal studies, baseline presence of adenomyosis is defined by considering available TAU/TVU or MRI results at Baseline of this extension study (prior to or on Study Day 14). If a subject has adenomyosis results only from either TAU/TVU or an MRI at Baseline, then the Baseline adenomyosis is determined by that result; if a subject has adenomyosis results from both TAU/TVU and MRI at Baseline and they differ, then the MRI adenomyosis result is used. If a subject has multiple assessments of the same type (TAU/TVU, MRI), the results closest and prior to Study Day 1 will be used.

Additionally, the subset of subjects with adenomyosis identified at any point during the Baseline or Treatment Period in the pivotal studies and extension study will be defined as follows. If a subject has adenomyosis identified via TAU/TVU or MRI at any time during the Baseline or Treatment Period in the pivotal studies and extension study, the subject will be included in this set. This will include subjects who have differing results during the Treatment Period. Post-Treatment results will not be considered in the definition of this subset.

11.3.5.4 Post-Treatment Menses

The time to first post-treatment menses is defined as the number of days between the last dose date of study drug and the first day of a subject's first post-treatment menses in the Post-Treatment Follow-up Period of the extension study. Menses are defined as having at least 1 day when uterine bleeding categories of bleeding and/or spotting were reported.

Truncated on-treatment menses at the last dose date, i.e., menses that are ongoing at the last dose date will be excluded, and the next occurrence of menses in the Post-Treatment Follow-up Period will be considered as the first post-treatment menses and will be used for reporting the time to first post-treatment menses. If the gap between two consecutive post-treatment menstrual cycles (or between truncated on-treatment menstrual cycle and the next post-treatment menstrual cycle) is 1 day apart or less (i.e., out of the two consecutive cycles, the start date of the second cycle – the end date of the first cycle ≤ 2), these two menstrual cycles will be combined as one menstrual cycle in determining the first post-treatment menses.

AH data will be used for post-treatment menses. If post-treatment menses could not be identified from available AH data, UBQ will be used. For the first UBQ that indicates bleeding ("Subject only has spotting that did not require the use of sanitary products" or "There was no visible blood on the sanitary products" not allowed) after the last dose date, the date of that UBQ will be considered as the first day of the subject's first post-treatment menses.

Volume of bleeding in a subject's first post-treatment menses will be calculated and summarized. Only AH data will be used for volume of bleeding in first post-treatment menses. Time to first post-treatment menses and volume of bleeding in first post-treatment menses will be summarized and analyzed separately for the following groups of subjects:

- All subjects who entered the Post-Treatment Follow-up Period of extension study.
- Subjects who entered the Post-Treatment Follow-up Period and were amenorrhiec at the Final Month in the Treatment Period of extension study. Amenorrhea is defined as in Section 11.3.2.
- Subjects who entered the Post-Treatment Follow-up Period and achieved suppression of bleeding at the Final Month in the Treatment Period of extension study. Suppression of bleeding is defined as in Section 11.3.3.

Based on the time to first post-treatment menses during the Post-Treatment Follow-up Period, each subject will be categorized into one of the following non-overlapping categories: Study End Days 1 – 28, Study End Days 29 – 56, Study End Days 57 – 84, Study End Days 85 – 112, Study End Days 113 – 140, Study End Days 141 – 168, and Study End Days ≥ 168. Summary tables will present the number and percentage of subjects in the specified categories.

11.3.6 Quality of Life Questionnaire

11.3.6.1 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 Month 3, Month 6, and Final Visit during the Treatment Period will be calculated and summarized by analysis group for each of the UFS-QoL subscales (symptom severity, concern, activities, energy/mood, control, self-conscious, and sexual function) and the HRQL total. Missing scores will be dealt with using the methods recommended by the UFS-QoL Scoring Manual. The UFS-QoL questionnaire is presented in Appendix A, and the relevant UFS-QoL Scoring Manual is presented in Appendix B.

11.3.6.2 EQ-5D-5L

The EQ-5D-5L questions and relevant scoring rules are presented in Appendix E.

The number and percentage of subjects with answers in each category of the EQ-5D-5L (Mobility, Self-care, Usual activities, Pain/Discomfort, and Anxiety/Depression) domains will be summarized at Baseline, each planned assessment, and Final Visit during the Treatment Period by analysis group.

Subject's responses to the EQ-5D-5L will be combined into a unique health state using a 5-digit code with 1 digit from each of the 5 dimensions at Baseline, each planned assessment, and Final Visit during the Treatment Period. The EQ-5D-5L states will be converted into a single preference-weighted health utility index score by applying

country-specific weights if available or US weights if country-specific weights are unavailable.[6,7]

The EQ VAS scale is numbered from 0 – 100, with 100 indicating the best health that a subject can imagine, and 0 indicating the worst health that a subject can imagine.

11.3.6.3 Patient Global Impression of Change (PGIC)

11.3.6.3.1 PGIC on Menstrual Bleeding Questionnaire

For PGIC on menstrual bleeding (PGIC – MB, Appendix C), the number and percentage of subjects in each response category will be summarized at Month 1, Month 3, Month 6, and Final Visit by analysis group.

For PGIC – MB, the response categories of "Very Much Improved" and "Much Improved" will be combined together. 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 Month 1, Month 3, Month 6, and Final Visit by analysis group.

11.3.6.3.2 PGIC on Non-Bleeding Uterine Fibroid Symptoms Questionnaire

For each of the PGIC questions on non-bleeding uterine fibroids symptoms (PGIC – NBUFS, Appendix D), the number and percentage of subjects in each response category will be summarized at Month 1, Month 3, Month 6, and Final Visit by analysis group.

For each of the PGIC – NBUFS questions, the response categories of "Very Much Improved" and "Much Improved" will be combined together. 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 Month 1, Month 3, Month 6, and Final Visit by analysis group.

11.3.6.4 Work Productivity and Activity Questionnaire: Uterine Fibroids (WPAI:UF)

For each of the measures as presented in the scoring guide Appendix I based on the data collected via the WPAI:UF, the change from Baseline to Month 6 and Final Visit during the Treatment Period will be summarized by analysis group. The WPAI:UF is presented in Appendix H, and the WPAI:UF scoring guide is presented in Appendix I.

11.3.6.5 Health Care Resource Utilization Questionnaire (HCRU)

HCRU data will be summarized as observed; missing data will not be imputed. HCRU data will be summarized by analysis group at Months 1 – 6 and Final Visit. Descriptive statistics will be presented for the total number of Non-Study Health Care Practitioner Visits, overall and by the type of facility subjects were seen at. The number and percentage of subjects will also be presented by the type of Non-Study Health Care Practitioner who administered care to the subject.

The number and percentage of subjects will be presented by the type of diagnostic or therapeutic procedures performed based on HCRU at Months 1 – 6 and Final Visit during the Treatment Period.

11.3.6.6 Number of Days in Hospital

Hospitalization related data will be summarized as observed; missing data will not be imputed. Hospitalization data will be summarized by analysis group in the Treatment Period based on the Adverse Events eCRF.

The number and percentage of subjects who were hospitalized or had prolonged hospitalization will be summarized by analysis 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 analysis group. The number of days in hospital (i.e., discharge date – admission date + 1) will be summarized by analysis

group. Only hospitalizations with an 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) are 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.

11.4 Handling of Multiplicity

No multiplicity adjustment is needed since only summary statistics will be provided for the efficacy analysis.

12.0 Safety Analysis

12.1 General Considerations

Safety analyses will be carried out using the Safety Analysis Set, which includes all subjects who received at least one dose of study drug. Safety data will be summarized by actual treatment received. If a subject receives 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.

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 relevant visit will be summarized by analysis group with the mean, SD (or SE), and 95% CIs. The between-group differences of the two elagoix dose groups with at least 12 month of total exposure to elagolix in the Treatment Period 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 analysis group.

For qualitative categorical variables, Fisher's exact test will be used to analyze between-group differences when applicable. Categorical data will be summarized by number and percentage of subjects by analysis group.

Pregnancies and outcomes will also be summarized by analysis group.

Statistical comparisons will be performed for BMD between the two elagoix dose groups with at least 12 month of total exposure to elagolix in the Treatment Period, as specified in Section 12.4.1. There will be no testing for other safety endpoints in the Treatment Period. There will be no testing for safety 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 specified in the relevant sections.

12.2 Adverse Events

12.2.1 Analysis of Adverse Events in the Extension Study

Adverse events (AEs) in the extenstion study will be summarized using system organ classes (SOCs) and preferred terms (PTs) from the Medical Dictionary for Regulatory Activities (MedDRA) Version 19.0 or higher by the six analysis groups as described in Section 5.3.

For each AE summary, the number and percentage of subjects experiencing at least one AE will be presented. Subjects reporting more than one AE within a SOC will be counted only once for that SOC. Subjects reporting more than one AE for a PT will be counted only once for that PT.

Treatment-emergent AEs are defined as AEs with a start date on or after the first dose of study drug in the extension study. AEs starting more than 30 days following discontinuation of study drug will not be included in summaries of treatment-emergent

AEs. Post-Treatment AEs are defined as AEs starting more than 30 days following discontinuation of study drug in the Treatment Period of this study.

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.

Treatment Period

The number and percentage of subjects with treatment-emergent AEs will be calculated by analysis group as follows:

- Any AEs
- Any AEs by primary MedDRA system organ class (SOC) and preferred term (PT)
- Any AEs occurring in ≥ 5% of subjects by primary MedDRA system organ class (SOC) and preferred term (PT)
- Any AEs by PT in descending frequency of 12 month elagolix 300 mg BID plus E2/NETA treatment group
- Any AEs by maximum severity (by primary MedDRA SOC and PT)
- Any AEs by maximum relationship (by primary MedDRA SOC and PT)
- Any AEs reasonably possibly related to study drug (by primary MedDRA SOC and PT, by PT)
- Any AEs leading to study drug discontinuation (by primary MedDRA SOC and PT)
- Any serious AEs (SAEs) (by primary MedDRA SOC and PT, by PT)
- Any severe AEs (by primary MedDRA SOC and PT)

- Any SAEs reasonably possibly related to study drug (by primary MedDRA SOC and PT, by PT)
- Any SAEs leading to study drug discontinuation (by primary MedDRA SOC and PT)
- Any AEs of special interest as specified in Appendix M
- Any AEs leading to death (by primary MedDRA SOC and PT)

A listing by the four main analysis group as in Section 5.3 of treatment-emergent AEs with subject IDs will be generated. Listings of all treatment-emergent SAEs, AEs leading to death, and AEs leading to study drug discontinuation will be generated.

Summary of total number of hot flush and night sweat by subject-reported severity categories, summary of total number of hot flush and night sweat by maximum subject-reported severity, and summary of time to first onset of hot flush and night sweat will be provided. Listings of subjects associated with hot flush and night sweat will also be provided.

Post-Treatment Follow-Up Period

The Post-Treatment AEs will be summarized as follows:

- Any AEs by primary MedDRA SOC and PT
- Any SAEs by primary MedDRA SOC and PT.

12.3 Laboratory Variables

Hematology, clinical chemistry, urinalysis, and endocrine panel variables collected in this study are specified in the protocol. An overall summary will be provided for all laboratory variables. For all other analysis, analysis will be performed for the following laboratory variables: 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, triglycerides (TG), 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.

12.3.1 Analysis of Laboratory Variables

Treatment Period

All laboratory variables will be summarized with mean, median, standard deviation, minimum, and maximum by the four main analysis groups as in Section 5.3.

For continuous laboratory variables, the analyses of change (for all laboratory variables of interest) and percent change (only for lipid variables) from Baseline to each relevant visit in the Treatment Period will be summarized by the four main analysis groups as in Section 5.3.

In the Treatment Period, the laboratory data will be categorized as low, normal, or high based on the normal ranges of the laboratory used in this study. The low or high laboratory values will be flagged in the data listings.

In the Treatment Period, shift tables for change from Baseline according to the normal range will be provided for laboratory parameters of interest mentioned above by the six analysis groups as in Section 5.3. The shift tables will tabulate the number and percentage of subjects with baseline values of high or normal to post-baseline low, baseline values of low or normal to post-baseline high, baseline values of high or normal to final post-baseline low, baseline values of low or normal to final post-baseline high. The final value in the Treatment Period refers to the last non-missing value collected within 3 days following the last dose of study drug.

In the Treatment Period, shift tables for change from Baseline to the minimum, maximum, and final value will also be presented based on the Common Terminology Criteria for

Adverse Events (CTCAE) v 4.0 grades[5] for the laboratory tests. For those analytes that are not reflected in the CTCAE, Exceptions to Standard CTCAE Lab Grading Criteria for Elagolix Studies will be used as specified in Appendix N. The final value in the Treatment Period is as defined above. The number and percentage of subjects meeting the CTCAE will be summarized by the four main analysis groups as in Section 5.3 in the Treatment Period.

The number and percentage of subjects meeting the following criteria will be summarized at Baseline and each relevant visit by the six analysis groups as in Section 5.3 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.

In addition, the number and percentage of subjects who have potentially clinically significant (PCS) lipid values meeting the following criteria any time during the Treatment Period will be summarized by the six analysis groups as in Section 5.3:

- 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 the four main analysis groups as in Section 5.3 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 CTCAE Grade 3 or 4;
- Maximum post-baseline lab values vs. Baseline lab values;
- Maximum post-baseline lab values vs. Baseline lab values for subjects with NCI CTCAE Grade 3 or 4.

Plots will be provided by the four main analysis groups as in Section 5.3 for HDL-C, LDL-C, and triglycerides during the Treatment Period:

- Mean percent change from Baseline in lipid over month.

Post-Treatment Follow-Up Period

All laboratory variables will be summarized with mean, median, standard deviation, minimum, and maximum by the four main analysis groups as in Section 5.3.

Laboratory data collected more than 3 days after the last dose of study drug in the Treatment Period will be included in the summary of data from the Post-Treatment Follow-up Period. Baseline for summaries/analyses in the Post-Treatment Follow-up Period is the same as Baseline for summaries/analyses in the Treatment Period.

In the Post-Treatment Follow-up Period, the following summaries will be presented:

- During the Post-Treatment Follow-up Period, shift tables for change from Baseline to relevant Post-Treatment Follow-up visit(s) according to the normal ranges will be provided for LDL-C, HDL-C, TG, and total cholesterol by the six analysis groups as in Section 5.3.
- During the Post-Treatment Follow-up Period, shift tables for change from Baseline to relevant Post-Treatment Follow-up visit(s) will also be presented

based on the CTCAE grades for LDL-C, HDL-C, TG, and total cholesterol by the four main analysis groups as in Section 5.3.

- Change in liver enzymes (SGPT/ALT, SGOT/AST, total bilirubin, and alkaline phosphatase) and lipid parameters (LDL-C, HDL-C, TG, and total cholesterol) from Baseline to relevant Post-Treatment Follow-up visit(s) will be summarized by the four main analysis groups as in Section 5.3 with descriptive statistics.

12.3.2 Assessment of Hepatotoxicity

The number and percentage of subjects in each analysis group with maximum on-treatment laboratory values meeting the following criteria compared to the upper limit of normal (ULN) will be summarized by the six analysis groups as in Section 5.3 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, ≥ 2.0 × 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 met any of the criteria defined above.

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

12.4 Bone Mineral Density

All analyses and summaries of bone mineral density (BMD) will be performed for each segment, 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 were included in the analysis with the data for the left femoral neck and left total hip, respectively (available for the majority of subjects), and an additional analysis was performed using the left side only. If more than one scan is reported for an anatomical segment within an analysis window, the worse (the lower value) of the multiple measurements were used for analysis for each anatomical segment.

Unless otherwise specified, summaries will be provided by the six analysis groups as in Section 5.3.

12.4.1 BMD in the Treatment Period

BMD

A continuous summary of BMD at Baseline, Between Month 3 and Month 6, Month 6, and Final Visit will be provided by analysis group. This summary will include the mean, SD, median, minimum, maximum, and first and third quartiles. This analysis will be repeated excluding subjects who switched machine manufacturer type (Lunar or Hologic), subjects with scans only on Lunar machines, and subjects with scans only on Hologic machines.

The analysis of percent change in BMD from Baseline to each relevant visit in the Treatment Period will be performed. The mean percent change in BMD from baseline to Month 6 and Final Visit during the Treatment Period will be compared between the two elagoix dose groups with at least 12 month of total exposure to elagolix, using ANCOVA with treatment as the main effect and Baseline value of corresponding parameter as a covariate. This analysis will be repeated excluding subjects who switched machine manufacturer type (Lunar or Hologic), subjects with scans only on Lunar machines, and subjects with scans only on Hologic machines. Statistical tests will be two-sided and a significance level of 0.05 will be used.

The number and percentage of subjects with percent change from Baseline to Between Month 3 and Month 6, Month 6, and Final Visit in the Treatment Period in mutually exclusive categories of BMD decrease: $\leq 0\%$, $> 0\% - \leq 1.5\%$, $> 1.5\% - \leq 3\%$, $> 3\% - \leq 5\%$, $> 5\% - < 8\%$, and $\geq 8\%$ will be summarized by analysis group and comparison between the two elagoix dose groups with at least 12 month of total exposure to elagolix will be made using Fisher's exact test.

Z-Score and T-Score

A continuous summary of the Z-score at Baseline, Between Month 3 and Month 6, Month 6, and Final Visit in the extension study will be provided by analysis group. This summary will include the mean, SD, median, minimum, maximum, and first and third quartiles. A categorical summary of Z-score will be summarized at Baseline, Between Month 3 and Month 6, Month 6, and Final Visit for the following categories: ≤ -2.0, > -2.0 to ≤ -1.5, > -1.5 to ≤ -1.0, and > -1.0 for all subjects, subjects with BMD decrease $\geq 3\%$ at Month 6, and subjects with BMD decrease $< 3\%$ at Month 6, respectively. The categorical summary of worst Z-score at any time in the Treatment Period will be produced by analysis group for each anatomical region.

A continuous summary of the T-score at Baseline, Between Month 3 and Month 6, Month 6, and Final Visit will be provided by analysis group. A categorical summary of T-score will be summarized at Baseline, Between Month 3 and Month 6, Month 6, and

Final Visit for the following categories: ≤ –2.5, > –2.5 to < –1.0, and ≥ –1.0 for all subjects, subjects with BMD decrease ≥ 3% at Month 6, and subjects with BMD decrease < 3% at Month 6, respectively. The categorical summary of worst T-score at any time in the Treatment Period will be produced by analysis group for each anatomical region.

Plots will be provided by the four main analysis groups as in Section 5.3 for:

- Percent change from Baseline to Month 6 in BMD vs. Baseline BMD values;
- Percent change from Baseline to Month 6 in BMD vs. Baseline BMD values for subjects with greater than 3% BMD decrease;
- Percent change from Baseline in BMD at Month 6 vs. Baseline Z-score;
- Maximum percent change from Baseline to Post-Baseline in BMD vs. Baseline Z-score;
- Categorical summary of lumbar spine BMD percent change from Baseline to Month 6.

A plot by the four main analysis groups as in Section 5.3 of the cumulative distribution function of percent change from baseline in BMD at Month 6 and the Z-scores at Month 6 will be provided. Additionally, the boxplots of Z-scores at baseline, Month 6, Post-Treatment Month 6, and Post-treatment Month 12 will be provided by the four main analysis groups as in Section 5.3. This will be repeated for T-scores.

Listings of subjects meeting the following thresholds during the Treatment Period in the extension study will be provided:

- Listing of subjects with bone decrease ≥ 8%
- Listing of subjects with Z-score ≤ –1.5
- Listing of subjects with T-score ≤ –1.5
- Listing of subjects with Z-score ≤ –1.5 and bone decrease ≥ 8% in the same region
- Listing of subjects with T-score ≤ –1.5 and bone decrease ≥ 8% in the same region.

12.4.2 BMD in the Post-Treatment Follow-Up Period

Percent change from Baseline to Post-Treatment Follow-up Month 6, Post-Treatment Follow-up Months 7 – 11, and Post-Treatment Follow-up Month 12 will be summarized for each analysis group with descriptive statistics. There will be no statistical testing of Post-Treatment Follow-up BMD values.

The analysis of percent change in BMD will be repeated to include only the subset of subjects who had a Treatment Month 6 and a Post-Treatment Follow-up Month 6 scan. Similarly, this analysis will be repeated to include only the subset of subjects who had a Treatment Month 6 and a Post-Treatment Follow-up Month 12 scan.

Percent change from Baseline in BMD to each of the Post-Treatment visits will be summarized for each analysis group with frequencies and percentage for categories of BMD decrease: $\leq 0\%$, $> 0\% - \leq 1.5\%$, $> 1.5\% - \leq 3\%$, $> 3\% - \leq 5\%$, $> 5\% - < 8\%$, and $\geq 8\%$.

Cross-tabulations of the number and percentage of subjects in each of the previously noted categories of percent change from Baseline for the following:

- Treatment Month 6 to Post-Treatment Follow-up Month 6
- Treatment Month 6 to Post-Treatment Follow-up Month 12
- Final Treatment to Post-Treatment Follow-up Month 6
- Final Treatment to Post-Treatment Follow-up Month 12
- Post-Treatment Follow-up Month 6 to Post-Treatment Follow-up Month 12.

The subjects included in these analyses will be limited to those who had values at these time points.

As described above for the Treatment Period, continuous and categorical summaries of the Z-score and T-score at each time point during the Post-Treatment Follow-up Period will be provided.

Listings of subjects meeting the following thresholds during the Post-Treatment Follow-up period will be provided:

- Listing of subjects with bone decrease ≥ 3% and T-score < –1.0 in the same region
- Listing of subjects with T-score ≤ –1.5.

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

Recovery at post-treatment month X = 100 × ((% change from baseline to final) – (% change from baseline to post-treatment month X))/(% change from baseline to final) i.e., it is the proportion of BMD decrease at the final treatment scan recovered at Post-Treatment Month X. For example, a subject who has a change of –2% at the end of treatment and has a change of –1% at Post-Treatment Month 6 will have a recovery value of 50% at Post-Treatment Month 6. This statistic is only defined for subjects who experience a decrease from baseline at final treatment.

A continuous summary of recovery at each post-treatment visit will be provided including mean, standard deviation, median and a within-group 95% confidence interval. Additionally, the number and percentage of subjects in each of the following categories will be provided: < 0%, 0 – 25%, > 25 – 50%, > 50 – 75%, > 75 – 100%, > 100%.

A plot of the cumulative distribution function of recovery at Post-Treatment Month 6 and Post-Treatment Month 12 will be provided.

12.5 Vital Signs and Body Weight

Vital sign variables include pulse rate, sitting systolic blood pressure, sitting diastolic blood pressure, and oral body temperature.

Vital sign variables will be summarized with mean, median, standard deviation, minimum, and maximum by the four main groups as in Section 5.3.

Analyses of mean change from Baseline to each relevant visit during the Treatment Period in vital sign variables and weight will be performed by the four main groups as in Section 5.3.

The number and percentage of subjects who have PCS vital sign and weight values meeting the following criteria will be summarized by the six analysis groups as in Section 5.3. All increase/decrease is 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 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
  - ≤ 50 mmHg and ≥ 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
- Weight
  - ≥ 5% decrease
  - ≥ 7% increase.

Change in vital signs from Baseline to relevant Post-Treatment Follow-up visit(s) (Post-Treatment Month 1, Month 3, Month 6, Month 9, Month 12, and Final Visit) will be summarized by the four main groups as in Section 5.3 with descriptive statistics.

12.6 Endometrial Biopsy

The number and percentage of subjects in each category of endometrial biopsy results will be summarized at Baseline, Prior to Month 6, and Month 6 for each of the six analysis groups as in Section 5.3. If multiple assessments exist in a specific time window, all assessments will be displayed.

12.7 Pelvic Ultrasound and MRI

Analyses will be conducted separately for results obtained from TAU/TVU and MRI. The actual values and the change from Baseline to Month 3 (TAU/TVU only), Month 6, Follow-up Month 3 (TAU/TVU only), and Follow-up Month 6 will be summarized by the six analysis groups as in Section 5.3 with descriptive statistics.

12.7.1 Ovarian Cysts

The presence of ovarian cysts will be noted at Baseline and each relevant post-baseline visit in the Treatment Period and Post-Treatment Follow-up Period. Significant ovarian findings include complex ovarian cyst > 3.5 cm or simple ovarian cyst > 5 cm. The number and percentage of subjects with the pre-defined complex ovarian cysts and simple ovarian cysts will be summarized for each analysis group at each relevant visit in the Treatment Period.

For each TVU/TAU or MRI 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). For such assessments, the worst assessment based on the greatest of the three dimensions across the multiple cysts at the left and/or right ovary location within the same cyst type (i.e., simple or complex) will be included in the analyses/summary. Additionally, the subject with multiple cyst findings will be 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 will include all results across multiple cysts findings from multiple assessments (where available) from subjects who have significant ovarian findings.

12.7.2 Endometrial Thickness

Analysis of change in endometrial thickness from Baseline to Month 3 and Month 6 in the Treatment Period will be performed.

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 in the Treatment and Post-Treatment Follow-up Periods will be summarized by analysis group at each relevant visit in the Treatment Period.

Change in endometrial thickness from Baseline to relevant Post-Treatment Follow-up visit(s) will be summarized by analysis group with descriptive statistics.

12.8 Columbia Suicide Severity Rating Scale (C-SSRS)

The C-SSRS data will be summarized as observed by the six analysis groups as in Section 5.3. The C-SSRS – Since Last Visit questionnaires and corresponding scoring rules are presented in Appendix J and Appendix K.

The number of subjects with suicidal ideation, suicidal behavior, and self-injurious behavior without suicidal intent at each scheduled assessment during the Treatment Period (including Day 1) will be summarized.

The number of subjects with suicide-related treatment-emergent events based on the C-SSRS during the Treatment Period (including Day 1) will be summarized.

A listing of subjects with suicidal ideation, suicidal behavior, or self-injurious behavior without suicidal intent based on the C-SSRS during the Treatment Period (including Day 1) will be provided.

12.9 Pregnancy Results

Pregnancies and outcomes will be summarized by the four main groups as in Section 5.3. 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.

13.0 Summary of Changes from SAP Version 1.0

1. The analysis windows for parameters collected every 3 months in the Treatment Period (Table 2) was updated; normial days were added in Table 5, Table 6, and Table 7;
2. The analysis conventions to handle multiple assessments in a specific time window were updated;
3. The analysis set for prior, concomitant, and post-treatment medications was updated to be full analysis set;
4. The change and percent change from Baseline to each month in the number of bleeding days was added;
5. The general considerations for mean (percent) change in safety analysis were updated to include SE for within-group change;
6. Typo was fixed in one hepatotoxicity criteria;
7. Added one clarification for the analysis of endometrial biopsy.

14.0 References

1. Schumacher U, Schumacher J, Mellinger U, et al. Estimation of menstrual blood loss volume based on menstrual diary and laboratory data. BMC Womens Health. 2012;12:24.
2. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med. 2002;21(16):2409-19.

3. 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.
4. Belsey EM, Machin D, d'Arcangues C. The analysis of vaginal bleeding patterns induced by fertility regulating methods. World Health Organization Special Programme of Research, Development and Research Training in Human Reproduction. Contraception. 1986;34(3):253-60.
5. Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. Published: May 28, 2009 (v4.03: June 14, 2010). US Department of Health and Human Services, National Institutes of Health, National Cancer Institute.
6. Szende A, Williams A, editors. Measuring self-reported population health: an international perspective based on EQ-5D-5L. EuroQol Group Monographs Volume 1. SpringMed. 2004.
7. Rabin R, Oemar M, Oppe M, et al. EQ-5D-5L user guide: basic information on how to use the EQ-5D-5L instrument Version 1.0. April 2011.

Appendix A. UFS-QoL

Pt. Initials: _____ Pt. ID: _____

Date: _____

UTERINE FIBROID SYMPTOM AND HEALTH-RELATED QUALITY OF LIFE QUESTIONNAIRE (UFS-QOL)

Listed below are symptoms experienced by women who have uterine fibroids. Please consider each symptom as it relates to your uterine fibroids or menstrual cycle. Each question asks how much distress you have experienced from each symptom during the previous 4 weeks.[1]

There are no right or wrong answers. Please be sure to answer every question by checking (✓) the most appropriate box. If a question does not apply to you, please mark "not at all" as a response.

During the previous 4 weeks[1], how distressed were you by…	Not at all	A little bit	Some-what	A great deal	A very great deal
1. Heavy bleeding during your menstrual period	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
2. Passing blood clots during your menstrual period	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
3. Fluctuation in the duration of your menstrual period compared to your previous cycles	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
4. Fluctuation in the length of your monthly cycle compared to your previous cycles	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
5. Feeling tightness or pressure in your pelvic area	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
6. Frequent urination during the daytime hours	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
7. Frequent nighttime urination	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
8. Feeling fatigued	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

[1] This questionnaire has been modified by Abbott with the permission of the SIR Foundation. Specifically, rather than asking how much distress you have experienced from various symptoms during the past 3 months, this questionnaires focuses on the past 4 weeks. SIR Foundation has not tested and is not responsible for the validity of this modification. Abbott plans to test the validity of this instrutment using phase 2a trial data. Your use of the questionnaire constitutes your agreement to release SIR Foundation from any responsibility for Abbott's changes to the document.

The following questions ask about your feelings and experiences regarding the impact of uterine fibroid symptoms on your life. Please consider each question as it relates to your experiences with uterine fibroids during the previous 4 weeks.[1]

There are no right or wrong answers. Please be sure to answer every question by checking (✓) the most appropriate box. If the question does not apply to you, please check "none of the time" as your option.

During the previous 4 weeks[1], how often have your symptoms related to uterine fibroids...	None of the time	A little of the time	Some of the time	Most of the time	All of the time
9. Made you feel anxious about the unpredictable onset or duration of your periods?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
10. Made you anxious about traveling?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
11. Interfered with your physical activities?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
12. Caused you to feel tired or worn out?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
13. Made you decrease the amount of time you spent on exercise or other physical activities?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
14. Made you feel as if you are not in control of your life?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
15. Made you concerned about soiling underclothes?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
16. Made you feel less productive?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
17. Caused you to feel drowsy or sleepy during the day?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
18. Made you feel self-conscious of weight gain?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
19. Made you feel that it was difficult to carry out your usual activities?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
20. Interfered with your social activities?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
21. Made you feel conscious about the size and appearance of your stomach?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
22. Made you concerned about soiling bed linen?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

During the previous 4 weeks[1], how often have your symptoms related to uterine fibroids...	None of the time	A little of the time	Some of the time	Most of the time	All of the time
23. Made you feel sad, discouraged, or hopeless?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
24. Made you feel down hearted and blue?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
25. Made you feel wiped out?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
26. Caused you to be concerned or worried about your health?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
27. Caused you to plan activities more carefully?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
28. Made you feel inconvenienced about always carrying extra pads, tampons, and clothing to avoid accidents?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
29. Caused you embarrassment?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
30. Made you feel uncertain about your future?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
31. Made you feel irritable?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
32. Made you concerned about soiling outer clothes?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
33. Affected the size of clothing you wear during your periods?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
34. Made you feel that you are not in control of your health?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
35. Made you feel weak as if energy was drained from your body?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
36. Diminished your sexual desire?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
37. Caused you to avoid sexual relations?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

This questionnaire is being used at Abbott's independent election pursuant to a license from the SIR Foundation. Abbott is solely responsible for the administration of this questionnaire and any related findings, conclusions or recommendations arising from such use. SIR Foundation is not responsible for any such use, findings, conclusions, or recommendations.

3
I:\institu\outadap\project\hc1617\question\original\final\uterq.doc-21/12/2001

Appendix B. UFS-QoL Scoring Manual

To calculate a symptom score for symptom severity, create a summed score from the items listed below and then use the formula below the table to transform the value. This will provide symptom scores where higher score values are indicative of greater symptom severity or bother and lower scores will indicate minimal symptom severity (high scores = bad).

Scale	Sum Item Values	Lowest and Highest Possible Raw Scores (Assuming All Related Scale Items are Answered)	Possible Raw Score Range (Assuming All Related Scale Items are Answered)
Symptom Severity	Sum 1 – 8	8, 40	32

Transformation for Symptom Severity Raw Score ONLY:

Transformed Score = (Actual raw score – lowest possible raw score)/possible raw score range × 100

For the UFS-QoL subscale (concern, activities, energy/mood, control, self-conscious, and sexual function), create summed scores of the items listed below for each individual subscale. To calculate the HRQL total score, sum the value of each individual subscale (do not sum individual items). Use the formula below to transform all values. Higher scores will be indicative of better UFS-QoL (high = good).

Scale	Sum Item Values	Lowest and Highest Possible Raw Scores (Assuming All Related Scale Items are Answered)	Possible Raw Score Range (Assuming All Related Scale Items are Answered)
Concern	9 + 15 + 22 + 28 + 32	5, 25	20
Activities	10 + 11 + 13 + 19 + 20 + 27 + 29	7, 35	28
Energy/Mood	12 + 17 + 23 + 24 + 25 + 31 + 35	7, 35	28
Control	14 + 16 + 26 + 30 + 34	5, 25	20
Self-conscious	18 + 21 + 33	3, 15	12
Sexual Function	36 + 37	2, 10	8
HRQL Total	Sum of 6 Subscale Scores	29, 145	116

Formula for transformation of UFS-QoL raw scores ONLY:

Transformed Score = (Highest possible score – Actual raw score)/Possible raw score range × 100

HRQL total = Sum of 6 Subscale Scores.

Missing Items:

- For each subscale listed in the tables above, including the scale of Symptom Severity, if some of the answers for related scale items are missing (not to exceed 50% of the scale items, details refer to Missing Items rule 2 below), then the lowest possible raw score = 1 × (total number of related scale items answered) and the highest possible raw score = 5 × (total number of related scale items answered). And the possible raw score range = the highest possible raw score – the lowest possible raw score.
- For the subscale analyses, if < 50% of the scale items are missing, the scale should be retained with the mean scale score of the items present used to impute a score for the missing items. If ≥ 50% of the items are missing, no scale score should be calculated, the subscale score should be considered missing. If a subscale score is missing, the HRQL total cannot be calculated.

Appendix C. Patient Global Impression of Change (PGIC-MB) – SAMPLE

Subjects will complete the PGIC-MB to assess the change in the severity of their menstrual bleeding (from very much improved to very much worse) since initiation of study drug by choosing one of seven responses at each scheduled monthly visit, starting with the Month 1 Visit in the Treatment Period or the Premature Discontinuation Visit, if applicable.

Visits to be completed: Treatment Period Month 1, Month 2, Month 3, Month 4, Month 5, Month 6/PD

Question:

Please answer the following question regarding your **menstrual bleeding:**

Since I started taking study medication, my **menstrual bleeding** has:

- ☐ Very much improved
- ☐ Much improved
- ☐ Minimally improved
- ☐ Not changed
- ☐ Minimally worse
- ☐ Much worse
- ☐ Very much worse

Appendix D. Patient Global Impression of Change, Non-Bleeding Uterine Fibroids Symptoms (PGIC-NBUFS) – SAMPLE

Subjects will complete the PGIC-NBUFS to document the presence of and to assess the change in the overall severity of non-bleeding symptoms and the severity of specific non-bleeding uterine fibroid symptoms (from very much improved to very much worse, or the subject will indicate whether she did not have a particular symptom or cannot remember if she had a particular symptom) at Treatment Period Month 1, Month 3 and Month 6, or the Premature Discontinuation Visit, if applicable.

Visits to be completed: Treatment Period Month 1, Month 3, Month 6/PD.

Questions:

Please answer the following questions regarding your non-bleeding uterine fibroid symptoms, that is, any symptom(s) that is present, whether or not you are having your period:

- Since I started taking study medication, my **abdominal or pelvic pain** has/is
  - ☐ Very much improved
  - ☐ Much improved
  - ☐ Minimally improved
  - ☐ Not changed
  - ☐ Minimally worse
  - ☐ Much worse
  - ☐ Very much worse
- Since I started taking study medication, my **abdominal or pelvic pressure** has/is
  - ☐ Very much improved
  - ☐ Much improved
  - ☐ Minimally improved
  - ☐ Not changed
  - ☐ Minimally worse

- ☐ Much worse
- ☐ Very much worse

- Since I started taking study medication, my **abdominal or pelvic cramping** has/is
  - ☐ Very much improved
  - ☐ Much improved
  - ☐ Minimally improved
  - ☐ Not changed
  - ☐ Minimally worse
  - ☐ Much worse
  - ☐ Very much worse
- Since I started taking study medication, my **back pain** has/is
  - ☐ Very much improved
  - ☐ Much improved
  - ☐ Minimally improved
  - ☐ Not changed
  - ☐ Minimally worse
  - ☐ Much worse
  - ☐ Very much worse
- Since I started taking study medication, my **abdominal bloating** has/is
  - ☐ Very much improved
  - ☐ Much improved
  - ☐ Minimally improved
  - ☐ Not changed
  - ☐ Minimally worse
  - ☐ Much worse
  - ☐ Very much worse

- Since I started taking study medication, my **urinary problems (urinating too frequently or feeling a sudden need to urinate)** has/is

  ☐ Very much improved
  ☐ Much improved
  ☐ Minimally improved
  ☐ Not changed
  ☐ Minimally worse
  ☐ Much worse
  ☐ Very much worse

- **Overall** since I started taking study medication, my **non-bleeding symptoms** have/are

  ☐ Very much improved
  ☐ Much improved
  ☐ Minimally improved
  ☐ Not changed
  ☐ Minimally worse
  ☐ Much worse
  ☐ Very much worse

Appendix E. EurolQol (EQ-5D-5L) – SAMPLE

Figure 1: EQ-5D-5L (UK English sample version)

Under each heading, please tick the **ONE** box that best describes your health **TODAY**

MOBILITY

I have no problems in walking about	❑
I have slight problems in walking about	❑
I have moderate problems in walking about	❑
I have severe problems in walking about	❑
I am unable to walk about	❑

SELF-CARE

I have no problems washing or dressing myself	❑
I have slight problems washing or dressing myself	❑
I have moderate problems washing or dressing myself	❑
I have severe problems washing or dressing myself	❑
I am unable to wash or dress myself	❑

USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*

I have no problems doing my usual activities	❑
I have slight problems doing my usual activities	❑
I have moderate problems doing my usual activities	❑
I have severe problems doing my usual activities	❑
I am unable to do my usual activities	❑

PAIN / DISCOMFORT

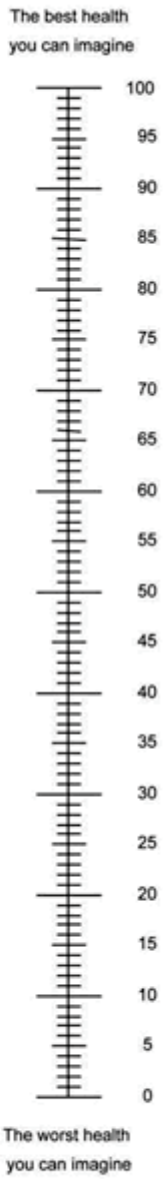
I have no pain or discomfort	❑
I have slight pain or discomfort	❑
I have moderate pain or discomfort	❑
I have severe pain or discomfort	❑
I have extreme pain or discomfort	❑

ANXIETY / DEPRESSION

I am not anxious or depressed	❑
I am slightly anxious or depressed	❑
I am moderately anxious or depressed	❑
I am severely anxious or depressed	❑
I am extremely anxious or depressed	❑

- We would like to know how good or bad your health is **TODAY**.
- This scale is numbered from **0** to **100**.
- **100** means the best health you can imagine.
  0 means the worst health you can imagine.
- Mark an **X** on the scale to indicate how your health is **TODAY**.
- Now, please write the number you marked on the scale in the box below.

YOUR HEALTH TODAY= ☐



Appendix F. Uterine Bleeding Questionnaire (UBQ) – Treatment Period – SAMPLE

Version 2.0

During site visits in the Treatment Period, subjects who **did not** return a Sanitary Product Collection Keg (for menstrual blood loss analysis) will be asked whether they had any uterine bleeding or spotting since their last study visit.

- **Did the subject have any bleeding or spotting since her last study visit?**

  ☐ No ☐ Yes

If yes, why were sanitary products not collected/returned? (Please select one response)

☐ Subject only had spotting that did not require the use of sanitary products*

☐ There was no visible blood on sanitary products*

☐ Subject forgot to/did not collect*

☐ Subject/Site discarded the sanitary products*

☐ Subject is still bleeding/spotting; will return when bleeding/spotting complete

☐ Subject collected sanitary products and did not bring them to this visit; will return sanitary products at a later date

☐ Other

* If this response is checked, remind subject to collect and return all used or worn sanitary products with or without visible blood.

Appendix G. Uterine Bleeding Questionnaire (UBQ) – Post-Treatment Follow-Up Period – SAMPLE

Version 2.0

Subjects who have not returned sanitary products for their first full menses in the Post-Treatment Follow-up Period, will be asked at the Post-Treatment Phone and Site Visits whether they had any bleeding or spotting since their last site or phone visit. Once a subject returns sanitary products for a full menses in the Post-Treatment Follow-up Period, this questionnaire no longer needs to be completed.

- **Did the subject have any bleeding or spotting since her last study visit (Site visit or phone visit)?**

  ☐ No ☐ Yes

If yes, why were sanitary products not collected/returned? (Please select one response)

☐ Subject only had spotting that did not require the use of sanitary products*

☐ There was no visible blood on sanitary products*

☐ Subject forgot to/did not collect*

☐ Subject/Site discarded the sanitary products*

☐ Subject is still bleeding/spotting; will return when bleeding/spotting complete

☐ Subject collected sanitary products and did not bring them in yet; will return sanitary products at a later date

☐ Other

* Subject will be required to collect sanitary products for her next full menses.

Appendix H. Work Productivity and Activity Impairment Questionnaire: Uterine Fibroids V2.0 (WPAI:UF) – SAMPLE

The following questions ask about the effect of uterine fibroid symptoms on your ability to work and perform regular activities. Please fill in the blanks or circle a number, as indicated.

- Are you currently employed (working for pay)? _____ NO ___ YES

  If *NO*, check *"NO"* and skip to Question 6.

  The next questions are about the **past seven days**, not including today.
- During the past seven days, how many hours did you miss from work because of problems associated with your uterine fibroid symptoms? *Include hours you missed on sick days, times you went in late, left early, etc., because of your uterine fibroid symptoms. Do not include time you missed to participate in this study.*

  _____ HOURS
- During the past seven days, how many hours did you miss from work because of any other reason, such as vacation, holidays, time off to participate in this study?

  _____ HOURS
- During the past seven days, how many hours did you actually work?

  _____ HOURS (If "0," skip to question 6.)
- During the past seven days, how much did your uterine fibroid symptoms affect your productivity while you were working?

  Think about days you were limited in the amount or kind of work you could do, days you accomplished less than you would like, or days you could not do your work as carefully as usual. If uterine fibroid symptoms affected your work only a little, choose a low number. Choose a high number if uterine fibroid symptoms affected your work a great deal.

Consider only how much uterine fibroids symptoms affected productivity while you were working.

Uterine fibroid symptoms had no effect on my work	0	1	2	3	4	5	6	7	8	9	10	Uterine fibroid symptoms completely prevented me from working

CIRCLE A NUMBER

- During the past seven days, how much did your uterine fibroid symptoms affect your ability to do your regular daily activities, other than work at a job?

  By regular activities, we mean the usual activities you do, such as work around the house, shopping, childcare, exercising, studying, etc. Think about times you were limited in the amount or kind of activities you could do and times you accomplished less than you would like. If uterine fibroid symptoms affected your activities only a little, choose a low number. Choose a high number if uterine fibroid symptoms affected your activities a great deal.

Consider only how much uterine fibroid symptoms affected your ability to do your regular daily activities, other than work at a job.

Uterine fibroid symptoms had no effect on my daily activities	0	1	2	3	4	5	6	7	8	9	10	Uterine fibroid symptoms completely prevented me from doing my daily activities

CIRCLE A NUMBER

WPAI:UF V2.0 (US English)

Appendix I. WPAI:UF Scoring Rules

The WPAI yields four types of scores:

- Absenteeism (work time missed)
- Presenteesism (impairment at work/reduced on-the-job effectiveness)
- Work productivity loss (overall work impairment/absenteeism plus presenteeism)
- Activity Impairment

WPAI:UF

WPAI outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity, i.e., worse outcomes, as follows:

Questions:

1 = currently employed

2 = hours missed due to specified problem

3 = hours missed other reasons

4 = hours actually worked

5 = degree problem affected productivity while working

6 = degree problem affected regular activities

Scores:

Multiply scores by 100 to express in percentages.

Percent work time missed due to problem: Q2/(Q2 + Q4)

Percent impairment while working due to problem: Q5/10

Percent overall work impairment due to problem:

$Q2/(Q2 + Q4) + [(1-(Q2/(Q2 + Q4))) \times (Q5/10)]$

Percent activity impairment due to problem: Q6/10

http://www.reillyassociates.net/WPAI_Scoring.html

Appendix J. Columbia-Suicide Severity Rating Scale (C-SSRS) – Since Last Visit – SAMPLE

COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS)

Since Last Visit

Version 1/14/09

Posner, K.; Brent, D.; Lucas, C.; Gould, M.; Stanley, B.; Brown, G.; Fisher, P.; Zelazny, J.; Burke, A.; Oquendo, M.; Mann, J.

Disclaimer:

This scale is intended to be used by individuals who have received training in its administration. The questions contained in the Columbia-Suicide Severity Rating Scale are suggested probes. Ultimately, the determination of the presence of suicidal ideation or behavior depends on the judgment of the individual administering the scale.

Definitions of behavioral suicidal events in this scale are based on those used in ***The Columbia Suicide History Form****, developed by John Mann, MD and Maria Oquendo, MD, Conte Center for the Neuroscience of Mental Disorders (CCNMD), New York State Psychiatric Institute, 1051 Riverside Drive, New York, NY, 10032. (Oquendo M. A., Halberstam B. & Mann J. J., Risk factors for suicidal behavior: utility and limitations of research instruments. In M.B. First [Ed.]* Standardized Evaluation in Clinical Practice*, pp. 103 - 130, 2003.)*

For reprints of the C-SSRS contact Kelly Posner, Ph.D., New York State Psychiatric Institute, 1051 Riverside Drive, New York, New York, 10032; inquiries and training requirements contact posnerk@nyspi.columbia.edu

SUICIDAL IDEATION

Ask questions 1 and 2. If both are negative, proceed to "Suicidal Behavior" section. If the answer to question 2 is "yes", ask questions 3, 4 and 5. If the answer to question 1 and/or 2 is "yes", complete "Intensity of Ideation" section below.

Since Last Visit

1. Wish to be Dead
Subject endorses thoughts about a wish to be dead or not alive anymore, or wish to fall asleep and not wake up.
Have you wished you were dead or wished you could go to sleep and not wake up?
If yes, describe:
Yes ☐ **No** ☐

2. Non-Specific Active Suicidal Thoughts
General, non-specific thoughts of wanting to end one's life/commit suicide (e.g., *"I've thought about killing myself"*) without thoughts of ways to kill oneself/associated methods, intent, or plan during the assessment period.
Have you actually had any thoughts of killing yourself?
If yes, describe:
Yes ☐ **No** ☐

3. Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
Subject endorses thoughts of suicide and has thought of at least one method during the assessment period. This is different than a specific plan with time, place or method details worked out (e.g., thought of method to kill self but not a specific plan). Includes person who would say, *"I thought about taking an overdose but I never made a specific plan as to when, where or how I would actually do it...and I would never go through with it."*
Have you been thinking about how you might do this?
If yes, describe:
Yes ☐ **No** ☐

4. Active Suicidal Ideation with Some Intent to Act, without Specific Plan
Active suicidal thoughts of killing oneself and subject reports having some intent to act on such thoughts, as opposed to *"I have the thoughts but I definitely will not do anything about them."*
Have you had these thoughts and had some intention of acting on them?
If yes, describe:
Yes ☐ **No** ☐

5. Active Suicidal Ideation with Specific Plan and Intent
Thoughts of killing oneself with details of plan fully or partially worked out and subject has some intent to carry it out.
Have you started to work out or worked out the details of how to kill yourself? Do you intend to carry out this plan?
If yes, describe:
Yes ☐ **No** ☐

INTENSITY OF IDEATION

The following features should be rated with respect to the most severe type of ideation (i.e., 1-5 from above, with 1 being the least severe and 5 being the most severe).

Most Severe Ideation: _____ _____
Type # (1-5) ***Description of Ideation***

Most Severe

Frequency
How many times have you had these thoughts?
(1) Less than once a week (2) Once a week (3) 2-5 times in week (4) Daily or almost daily (5) Many times each day
____

Duration
When you have the thoughts, how long do they last?
(1) Fleeting - few seconds or minutes
(2) Less than 1 hour/some of the time
(3) 1-4 hours/a lot of time
(4) 4-8 hours/most of day
(5) More than 8 hours/persistent or continuous
____

Controllability
Could/can you stop thinking about killing yourself or wanting to die if you want to?
(1) Easily able to control thoughts
(2) Can control thoughts with little difficulty
(3) Can control thoughts with some difficulty
(4) Can control thoughts with a lot of difficulty
(5) Unable to control thoughts
(0) Does not attempt to control thoughts
____

Deterrents
Are there things - anyone or anything (e.g., family, religion, pain of death) - that stopped you from wanting to die or acting on thoughts of committing suicide?
(1) Deterrents definitely stopped you from attempting suicide
(2) Deterrents probably stopped you
(3) Uncertain that deterrents stopped you
(4) Deterrents most likely did not stop you
(5) Deterrents definitely did not stop you
(0) Does not apply
____

Reasons for Ideation
What sort of reasons did you have for thinking about wanting to die or killing yourself? Was it to end the pain or stop the way you were feeling (in other words you couldn't go on living with this pain or how you were feeling) or was it to get attention, revenge or a reaction from others? Or both?
(1) Completely to get attention, revenge or a reaction from others
(2) Mostly to get attention, revenge or a reaction from others
(3) Equally to get attention, revenge or a reaction from others and to end/stop the pain
(4) Mostly to end or stop the pain (you couldn't go on living with the pain or how you were feeling)
(5) Completely to end or stop the pain (you couldn't go on living with the pain or how you were feeling)
(0) Does not apply
____

SUICIDAL BEHAVIOR *(Check all that apply, so long as these are separate events; must ask about all types)*	**Since Last Visit**
Actual Attempt: A potentially self-injurious act committed with at least some wish to die, *as a result of act.* Behavior was in part thought of as method to kill oneself. Intent does not have to be 100%. If there is ***any*** intent/desire to die associated with the act, then it can be considered an actual suicide attempt. ***There does not have to be any injury or harm,*** just the potential for injury or harm. If person pulls trigger while gun is in mouth but gun is broken so no injury results, this is considered an attempt. Inferring Intent: Even if an individual denies intent/wish to die, it may be inferred clinically from the behavior or circumstances. For example, a highly lethal act that is clearly not an accident so no other intent but suicide can be inferred (e.g., gunshot to head, jumping from window of a high floor/story). Also, if someone denies intent to die, but they thought that what they did could be lethal, intent may be inferred. ***Have you made a suicide attempt?*** ***Have you done anything to harm yourself?*** ***Have you done anything dangerous where you could have died?*** ***What did you do?*** ***Did you_____ as a way to end your life?*** ***Did you want to die (even a little) when you_____?*** ***Were you trying to end your life when you_____?*** ***Or did you think it was possible you could have died from_____?*** ***Or did you do it purely for other reasons / without ANY intention of killing yourself (like to relieve stress, feel better, get sympathy, or get something else to happen)?*** (Self-Injurious Behavior without suicidal intent) If yes, describe: **Has subject engaged in Non-Suicidal Self-Injurious Behavior?**	Yes ☐ No ☐ Total # of Attempts _____ Yes ☐ No ☐
Interrupted Attempt: When the person is interrupted (by an outside circumstance) from starting the potentially self-injurious act *(if not for that, actual attempt would have occurred).* Overdose: Person has pills in hand but is stopped from ingesting. Once they ingest any pills, this becomes an attempt rather than an interrupted attempt. Shooting: Person has gun pointed toward self, gun is taken away by someone else, or is somehow prevented from pulling trigger. Once they pull the trigger, even if the gun fails to fire, it is an attempt. Jumping: Person is poised to jump, is grabbed and taken down from ledge. Hanging: Person has noose around neck but has not yet started to hang - is stopped from doing so. ***Has there been a time when you started to do something to end your life but someone or something stopped you before you actually did anything?*** If yes, describe:	Yes ☐ No ☐ Total # of interrupted _____
Aborted Attempt: When person begins to take steps toward making a suicide attempt, but stops themselves before they actually have engaged in any self-destructive behavior. Examples are similar to interrupted attempts, except that the individual stops him/herself, instead of being stopped by something else. ***Has there been a time when you started to do something to try to end your life but you stopped yourself before you actually did anything?*** If yes, describe:	Yes ☐ No ☐ Total # of aborted _____
Preparatory Acts or Behavior: Acts or preparation towards imminently making a suicide attempt. This can include anything beyond a verbalization or thought, such as assembling a specific method (e.g., buying pills, purchasing a gun) or preparing for one's death by suicide (e.g., giving things away, writing a suicide note). ***Have you taken any steps towards making a suicide attempt or preparing to kill yourself (such as collecting pills, getting a gun, giving valuables away or writing a suicide note)?*** If yes, describe:	Yes ☐ No ☐
Suicidal Behavior: Suicidal behavior was present during the assessment period?	Yes ☐ No ☐
Suicide:	Yes ☐ No ☐
Answer for Actual Attempts Only	Most Lethal Attempt Date:
Actual Lethality/Medical Damage: 0. No physical damage or very minor physical damage (e.g., surface scratches). 1. Minor physical damage (e.g., lethargic speech; first-degree burns; mild bleeding; sprains). 2. Moderate physical damage; medical attention needed (e.g., conscious but sleepy, somewhat responsive; second-degree burns; bleeding of major vessel). 3. Moderately severe physical damage; *medical* hospitalization and likely intensive care required (e.g., comatose with reflexes intact; third-degree burns less than 20% of body; extensive blood loss but can recover; major fractures). 4. Severe physical damage; *medical* hospitalization with intensive care required (e.g., comatose without reflexes; third-degree burns over 20% of body; extensive blood loss with unstable vital signs; major damage to a vital area). 5. Death	*Enter Code* _____
Potential Lethality: Only Answer if Actual Lethality=0 Likely lethality of actual attempt if no medical damage (the following examples, while having no actual medical damage, had potential for very serious lethality: put gun in mouth and pulled the trigger but gun fails to fire so no medical damage; laying on train tracks with oncoming train but pulled away before run over). 0 = Behavior not likely to result in injury 1 = Behavior likely to result in injury but not likely to cause death 2 = Behavior likely to result in death despite available medical care	*Enter Code* _____

 C-SSRS—Since Last Visit (Version 1/14/09) Page 2 of 2

Appendix K. C-SSRS Scoring Rules

The following outcomes are C-SSRS categories and have binary responses (yes/no). The categories have been re-ordered from the actual scale to facilitate the definitions of the composite and comparative endpoints, and to enable clarity in the presentation of the results.

Category 1 – Wish to be Dead

Category 2 – Non-specific Active Suicidal Thoughts

Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act

Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan

Category 5 – Active Suicidal Ideation with Specific Plan and Intent

Category 6 – Preparatory Acts or Behavior

Category 7 – Aborted Attempt

Category 8 – Interrupted Attempt

Category 9 – Actual Attempt (non-fatal)

Category 10 – Completed Suicide

Self-injurious behavior without suicidal intent is also a C-SSRS outcome (although not suicide-related) and has a binary response (yes/no).

The following outcome is a numerical score derived from the C-SSRS categories.

- Suicidal Ideation Score: The maximum suicidal ideation category (1 – 5 on the C-SSRS) present at the assessment. Assign a score of 0 if no ideation is present.

Composite endpoints based on the above categories are defined below.

- Suicidal ideation: A "yes" answer at any time during treatment to any one of the five suicidal ideation questions (Categories 1 – 5) on the C-SSRS.
- Suicidal behavior: A "yes" answer at any time during treatment to any one of the five suicidal behavior questions (Categories 6 – 10) on the C-SSRS.
- Suicidal behavior ideation or behavior: A "yes" answer at any time during treatment to any one of the ten suicidal ideation and behavior questions (Categories 1 – 10) on the C-SSRS.

Comparative endpoints of interest are defined below. "Treatment emergence" is used for outcomes that include events that first emerge or worsen. "Emergence" is used for outcomes that include events that first emerge.

- Treatment-emergent suicidal ideation compared to recent history: An increase in the maximum suicidal ideation score during treatment from the maximum suicidal ideation category during a specified pre-treatment period (C-SSRS scales taken during the specified pre-treatment period; excludes "lifetime" scores from the Baseline C-SSRS scale or Baseline/Screening C-SSRS scale).
- Treatment-emergent serious suicidal ideation compared to recent history: An increase in the maximum suicidal ideation score to 4 or 5 on the C-SSRS during treatment from not having serious suicidal ideation (scores of 0 – 3) during a specified pre-treatment period (C-SSRS scales taken during the specified pretreatment period; excludes "lifetime" scores from the Baseline C-SSRS scale or Baseline/Screening C-SSRS scale).
- Emergence of serious suicidal ideation compared to recent history: An increase in the maximum suicidal ideation score to 4 or 5 on the C-SSRS during treatment from no suicidal ideation (scores of 0) during a specified pre-treatment period (CSSRS scales taken during the specified pre-treatment period; excludes "lifetime" scores from the Baseline C-SSRS scale or Baseline/Screening C-SSRS scale).
- Emergence of suicidal behavior compared to all prior history: The occurrence of suicidal behavior (Categories 6 – 10) during treatment from not having

suicidal behavior (Categories 6 – 10) prior to treatment (includes "lifetime" and/or "screening" scores from the Baseline C-SSRS scale, Screening C-SSRS scale, or Baseline/Screening C-SSRS scale, and any "Since Last Visit" from the Since Last Visit C-SSRS scales taken prior to treatment)

- Improvement in suicidal ideation at a time point of interest compared to baseline: An improvement in this endpoint can be considered as a decrease in suicidal ideation score at the time point of interest (e.g., the last measurement during treatment) from the baseline measurement (e.g., the measurement taken just prior to treatment. This analysis should only be performed for studies in which a baseline C-SSRS can be defined (i.e., having improvement from the worse event over a lifetime is not clinically meaningful).

Appendix L. Health Care Resource Utilization Questionnaire (HCRU) Version 2.0 – SAMPLE

(To be completed by Site Staff)

Non-Study Visits for Routine/General Health Care

Version 2.0

Instructions to Site Staff: At each scheduled monthly visit (Month 1 through Month 6) in the Treatment Period, please ask if the subject saw a ***Non-Study*** Health Care Practitioner (HCP) since her last scheduled monthly visit for a ***routine/general health care visit that is not associated with an adverse event***.

Only record routine/general health care visits with Non-Study HCPs below.

Record Non-Study HCP Visits associated with an adverse event on the Adverse Event form only. Do not record below.

- Since the Subject's last scheduled monthly study visit, has she seen a non-study Health Care Practitioner (e.g., Physician, Nurse Practitioner, Physician Assistant, Dentist, Physical Therapist) for a **routine/general health care visit that is not associated with an adverse event**? ☐ **No** ☐ **Yes**

If **Yes**, please complete the questions below.

Record Non-Adverse Event Related Visits Below Any Visits Associated with Adverse Events should be Recorded on the Adverse Event eCRF only			
2. What type of facility was the subject seen at?	**3. How many times was the subject seen by each facility?**	**4. What type(s) of *Non-Study* Health Care Practitioner was the Subject seen by? (Check all that apply)**	**5. How many times was the subject seen by each *Non-Study* Health Care Practitioner?**
☐ Office		☐ AUDIOLOGIST	
		☐ ALLERGIST	
		☐ CARDIOLOGIST	
		☐ DENTIST	
		☐ DERMATOLOGIST	
		☐ ENDOCRINOLOGIST	
		☐ ENT	
		☐ FAMILY PHYSICIAN	
		☐ GASTROENTEROLOGIST	
		☐ GYNECOLOGIST	
		☐ HEMATOLOGIST	
		☐ HEPATOLOGIST	
		☐ IMMUNOLOGIST	
		☐ INFECTIOUS DISEASE SPECIALIST	
		☐ INTERNAL MEDICINE SPECIALIST	
		☐ INTERNIST	
		☐ MEDICAL GENETICIST	
		☐ NEPHROLOGIST	
		☐ NEUROSURGEON	
		☐ NURSE	
		☐ NURSE PRACTITIONER	
		☐ OCCUPATIONAL THERAPIST	
		☐ OPHTHALMOLOGIST	

		☐ ORTHOPEDIC SURGEON	
		☐ OPTOMETRIST	
		☐ PHYSIATRIST	
		☐ PHYSICAL THERAPIST	
		☐ PLASTIC SURGEON	
		☐ PODIATRIST	
		☐ PSYCHOLOGIST	
		☐ PULMONOLOGIST	
		☐ RADIOLOGIST	
		☐ REPRODUCTIVE ENDOCRINOLOGIST	
		☐ RHEUMATOLOGIST	
		☐ SURGEON	
		☐ UROLOGIST	
		☐ UNKNOWN	
		☐ OTHER HEALTH CARE PRACTITIONER (specify type):	
☐ Urgent Care		☐ AUDIOLOGIST	
		☐ ALLERGIST	
		☐ CARDIOLOGIST	
		☐ DENTIST	
		☐ DERMATOLOGIST	
		☐ ENDOCRINOLOGIST	
		☐ ENT	
		☐ FAMILY PHYSICIAN	
		☐ GASTROENTEROLOGIST	
		☐ GYNECOLOGIST	
		☐ HEMATOLOGIST	
		☐ HEPATOLOGIST	

		☐ IMMUNOLOGIST	
		☐ INFECTIOUS DISEASE SPECIALIST	
		☐ INTERNAL MEDICINE SPECIALIST	
		☐ INTERNIST	
		☐ MEDICAL GENETICIST	
		☐ NEPHROLOGIST	
		☐ NEUROSURGEON	
		☐ NURSE	
		☐ NURSE PRACTITIONER	
		☐ OCCUPATIONAL THERAPIST	
		☐ OPHTHALMOLOGIST	
		☐ ORTHOPEDIC SURGEON	
		☐ OPTOMETRIST	
		☐ PHYSIATRIST	
		☐ PHYSICAL THERAPIST	
		☐ PLASTIC SURGEON	
		☐ PODIATRIST	
		☐ PSYCHOLOGIST	
		☐ PULMONOLOGIST	
		☐ RADIOLOGIST	
		☐ REPRODUCTIVE ENDOCRINOLOGIST	
		☐ RHEUMATOLOGIST	
		☐ SURGEON	
		☐ UROLOGIST	
		☐ UNKNOWN	
		☐ OTHER HEALTH CARE PRACTITIONER (specify type):	
☐ Emergency Room		☐ AUDIOLOGIST	

		☐ ALLERGIST	
		☐ CARDIOLOGIST	
		☐ DENTIST	
		☐ DERMATOLOGIST	
		☐ ENDOCRINOLOGIST	
		☐ ENT	
		☐ FAMILY PHYSICIAN	
		☐ GASTROENTEROLOGIST	
		☐ GYNECOLOGIST	
		☐ HEMATOLOGIST	
		☐ HEPATOLOGIST	
		☐ IMMUNOLOGIST	
		☐ INFECTIOUS DISEASE SPECIALIST	
		☐ INTERNAL MEDICINE SPECIALIST	
		☐ INTERNIST	
		☐ MEDICAL GENETICIST	
		☐ NEPHROLOGIST	
		☐ NEUROSURGEON	
		☐ NURSE	
		☐ NURSE PRACTITIONER	
		☐ OCCUPATIONAL THERAPIST	
		☐ OPHTHALMOLOGIST	
		☐ ORTHOPEDIC SURGEON	
		☐ OPTOMETRIST	
		☐ PHYSIATRIST	
		☐ PHYSICAL THERAPIST	
		☐ PLASTIC SURGEON	

		☐ PODIATRIST	
		☐ PSYCHOLOGIST	
		☐ PULMONOLOGIST	
		☐ RADIOLOGIST	
		☐ REPRODUCTIVE ENDOCRINOLOGIST	
		☐ RHEUMATOLOGIST	
		☐ SURGEON	
		☐ UROLOGIST	
		☐ UNKNOWN	
		☐ OTHER HEALTH CARE PRACTITIONER (specify type):	

6. Did the Subject have any diagnostic or therapeutic procedures performed since the last scheduled monthly study visit? ☐ No ☐ Yes (If Yes, complete questions 7 and 8 below)

7. Diagnostic/Therapeutic Procedure (Check all that apply)		**8. How many times was the procedure performed?**
Ultrasound Scan	☐	
Physical Examination	☐	
Vital Signs	☐	
MRI	☐	
CT Scan	☐	
X-Ray	☐	
Biopsy and Histologic Examination	☐	
Pelvic Exam	☐	
Urine Test	☐	
Blood Test	☐	
Other (specify):_____	☐	

Appendix M. Adverse Events of Special Interest

Item of Safety Interest	Method of Surveillance
Hot flashes	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
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

Appendix N. Exceptions to Standard CTCAE 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 (CTCAE) is applied to all hematology and chemistry analytes across the elagolix program where quantitative criteria are available. In the CTCAE 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 CTAE criteria are not provided for applicable lab parameters and alternative references are applied.