Sponsor:	**ObsEva**		
Protocol:	**18-OBE2109-003**		
Document Version No.:	V5.0	**Document Date:**	13-JUN-2022

Protocol 18-OBE2109-003

A Phase 3 multicenter, randomized, double-blind, placebo-controlled, clinical study to assess the efficacy and safety of linzagolix in subjects with moderate to severe endometriosis-associated pain.

Protocol Number: **(Version Date)**	18-OBE2109-003 Version 4.0, 27 July 2020
Name of Test Drug:	Linzagolix (OBE2109)
Phase:	3
Methodology:	Randomized, double-blind, placebo-controlled
Sponsor:	ObsEva S.A. [redacted]
Sponsor Representative:	[redacted]
Document Date:	13 June 2022
Document Version:	Final V5.0

Sponsor:	**ObsEva**		
Protocol:	**18-OBE2109-003**		
Document Version No.:	V5.0	**Document Date:**	13-JUN-2022

SIGNATURE PAGE

Protocol Title: A Phase 3 multicenter, randomized, double-blind, placebo-controlled, clinical study to assess the efficacy and safety of linzagolix in subjects with moderate to severe endometriosis-associated pain.

Sponsor: ObsEva S.A.

[redacted]

Protocol Number: 18-OBE2109-003

Document Date/Version: 13 June 2022/V5.0

Cytel, Inc. Author:

[redacted]

Sponsor Approval

By signing this document, I acknowledge that I have read the document and approve of the planned statistical analyses described herein. I agree that the planned statistical analyses are appropriate for this study, are in accordance with the study objectives, and are consistent with the statistical methodology described in the protocol, clinical development plan, and all applicable regulatory guidance's and guidelines.

I have discussed any questions I have regarding the contents of this document with the biostatistical author.

I also understand that any subsequent changes to the planned statistical analyses, as described herein, may have a regulatory impact and/or result in timeline adjustments. All changes to the planned analyses will be described in the clinical study report (CSR).

Sponsor:	**ObsEva**		
Protocol:	**18-OBE2109-003**		
Document Version No.:	V5.0	**Document Date:**	13-JUN-2022

Sponsor Signatory:

[redacted]

Sponsor Signatory:

[redacted]

Sponsor:	**ObsEva**		
Protocol:	**18-OBE2109-003**		
Document Version No.:	V5.0	**Document Date:**	13-JUN-2022

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

Table **Page**

Sponsor:	**ObsEva**		
Protocol:	**18-OBE2109-003**		
Document Version No.:	V5.0	**Document Date:**	13-JUN-2022

ABBREVIATIONS

Abbreviation	Definition
ABT	Add-back therapy
AE	Adverse events
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AMH	Anti-Mullerian hormone
AST	Aspartate Aminotransferase
ATC	Anatomic Therapeutic class
AUC	Area under the concentration versus time curve
AUC_{0-24}	Area under the concentration versus time curve from time 0 to the end of the dosing interval 24 hours later, calculated using linear trapezoid rule
BCP	Best Cut Point
BMD	Bone Mineral Density
BMI	Body Mass Index
CDF	Cumulative Distribution Function
CI	Confidence Interval
ClinRO	Clinician Reported Outcome
CM	Concomitant Medication
C_{max}	Maximum plasma concentration
C_{min}	Trough plasma concentration, taken 24 hours after dose and prior to subsequent dose
CSR	Clinical study report
C-SSRS	Columbia-Suicide Severity Rating Scale
CV	Coefficient of Variation
DXA	Dual-energy X-ray absorptiometry
DYS	Dysmenorrhea
E2	Estradiol
EAP	Endometriosis Associated Pain

Abbreviation	Definition
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EHP-30	30-Item Endometriosis Health Profile
ePRO	Electronic Patient Reported Outcome
EQ-5D-5L	EuroQOL 5-Dimension 5-Level
FAS	Full Analysis Set
FSH	Follicle-Stimulating Hormone
FU	Follow-up
GCP	Good Clinical Practice
GGT	Gamma Glutamyl Transferase
GnRH	Gonadotropin releasing hormone
HDL	High-density lipoprotein
HRPQ	Health Related Productivity Questionnaire
HRQoL	Health Related Quality of Life
HRUQ	HealthCare Resource Utilization Questionnaire
ICH	International Conference on Harmonisation
IDMC	Independent Data Monitoring Committee
IMP	Investigational Medicinal Product
INR	International Normalized Ratio
IRB	Investigational Review Board
ITT	Intent-to-treat
IWRS	Interactive Web Response System
LDL	Low density lipoprotein
LH	Luteinizing hormone
LOQ	Limit of Quantification
MAR	Missing At Random
mB&B	Biberoglu & Behrman
MCT	Meaningful change threshold
MedDRA	Medical Dictionary for Regulatory Activities
MME	Morphine Milligram Equivalent

Abbreviation	Definition
mmHg	Millimeters of mercury
MNAR	Missing Not At Random
mPGIS	Monthly Patient Global Impression of Severity
Ms	Millisecond
NETA	Norethisterone acetate
NMPP	Non-Menstrual Pelvic Pain
NRS	Numeric Rating Scale
NSAIDS	Non-steroidal anti-inflammatory drugs
OR	Odds-ratio
P4	Progesterone
PD	Pharmacodynamic
PDF	Probability Density Function
PGIS	Patient Global Impression of Severity
PGIC	Patient Global Impression of Change
PK	Pharmacokinetic
PP	Per Protocol Set
PPGIC	Post-treatment Global Impression of Change
PPV	Pelvic Pain Verbal Rating Scale
PROMIS	Patient Reported Outcomes Measurement Information System
PSIQ	Physician Surgery Intention Question
QT	Qt interval
QTC	Corrected Qt interval
QTCf	Corrected Qt interval Fridericia
REB	Research Ethics Board
ROC	Receiver operating characteristic
SAP	Statistical analysis plan
SHBG	Sex hormone-binding globulin
SOC	System Organ Class
SSIQ	Subject Surgery Intention Question
TBL	Total Bilirubin

Abbreviation	Definition
T_{max}	Time to maximum plasma concentration
TTO	Time Trade Off
TVUS	TransVaginal UltraSound
ULN	Upper Limit of Normal
VRS	Verbal Rating Scale

1. INTRODUCTION AND OBJECTIVES OF ANALYSIS

1.1. Introduction

This is a prospective, randomized, double-blind, placebo-controlled study to demonstrate the efficacy and safety of linzagolix administered orally once daily at doses of 75 mg alone and 200 mg in combination with add-back therapy (ABT; E2 1 mg/NETA 0.5 mg) versus placebo in the management of moderate to severe endometriosis associated pain (EAP) in women with surgically confirmed endometriosis.

1.2. Objectives of Statistical Analysis

The primary objective of this study is to demonstrate the efficacy and safety of linzagolix administered orally once daily for up to 3 months at a dose of 75 mg alone or of 200 mg in combination with ABT (E2 1 mg / NETA 0.5 mg) versus placebo, while under randomized treatment, in the management of moderate to severe EAP in women with surgically confirmed endometriosis. The two co-primary efficacy endpoints will be measured as clinically meaningful reduction in pain score over the last 28 days of randomized treatment up to the Month 3 visit, along with a stable or decreased use of rescue analgesics for EAP for 1) dysmenorrhea (DYS) and for 2) non-menstrual pelvic pain (NMPP).

Meaningful changes for the primary endpoints and for the additional supportive responder analyses for the ranked secondary endpoints will be estimated based on blinded Month 3 data for the primary endpoints and blinded Month 6 data for ranked secondary endpoints and will be applied on the responder threshold analyses.

Secondary objectives include evaluation of persistence of efficacy over the last 28 days of randomized treatment up to the Month 6 visit, evaluation of pain associated with sexual intercourse (dyspareunia) and defecation (dyschezia), difficulty of doing daily activities, analgesic use, assessment of subject perception of severity, change in uterine bleeding, Health Related Quality of Life (HRQoL) questionnaires, pharmacoeconomic burden of endometriosis by assessing changes in patient productivity, assessment of endometriosis related number of non-study health visits, number of days in hospital and type of medical procedures performed during the treatment period.

Safety and tolerability objectives include assessment of bone mineral density (BMD), endometrial health, cardiac safety including QT interval prolongation, standard laboratory safety parameters, gynecological assessments and adverse event (AE) frequency including specific hypoestrogenic AEs.

Exploratory objectives include assessment of bone turnover markers and collection of pharmacokinetic (PK) and pharmacodynamic (PD) data of linzagolix for a separate modelling exercise.

This statistical analysis plan (SAP) is designed to outline the methods to be used in the analysis of study data to answer the study objective(s). Populations for analysis, data handling rules, statistical methods, and formats for data presentation are provided. The statistical analyses and summary tabulations

described in this SAP will provide the basis for the results sections of the clinical study report (CSR) for this trial.

This SAP also will outline any differences in the currently planned analytical objectives relative to those planned in the study protocol.

2. STUDY DESIGN

2.1. Synopsis of Study Design

This is a prospective, randomized, double-blind, placebo-controlled study to demonstrate the efficacy and safety of linzagolix administered orally once daily at doses of 75 mg alone and 200 mg in combination with low dose ABT (E2 1 mg/NETA 0.5 mg) versus placebo in the management of moderate to severe EAP in women with surgically confirmed endometriosis.

The study starts with an approximately 3-month screening period during which subjects will receive no study drug. Eligibility will be confirmed based on data collected during the screening period. After randomization, a 6-month Treatment Period will follow. Subjects will be randomized to one of the three treatment groups:

1. Linzagolix 75 mg

2. Linzagolix 200 mg combined with ABT

3. Placebo

Randomization will be conducted via an interactive web response system (IWRS) in a 1:1:1 ratio. There will be no stratification.

At 6 months, BMD change will be assessed via DXA measurement. Eligible subjects who have completed the 6-month treatment period may enter a separate extension study for 6 additional months of active treatment (no placebo control). In this extension study, subjects who previously received placebo will be randomly switched to one of the two active treatments (75 mg alone or 200 mg + ABT). Subjects who received active treatment will continue with the same treatment.

Subjects who decline to participate to or do not qualify for the extension study and who were exposed to at least 3 months of treatment will enter a 6-month drug-free follow-up. At the end of the 6-month follow-up period, subjects with a BMD decrease from baseline of >1.5% for lumbar spine and >2.5% for total hip will have an additional DXA scan 6 months later. Subjects who discontinue treatment prior to Month 3 will not enter the follow-up period.

The main analysis will be performed after 6 months of treatment and will include all of the subjects' data up to Month 6. The primary endpoint analysis will be based on subjects' data up to Month 3.

The study will last on average 15 months for each patient; 3-month Screening Period, 6-month Treatment Period and 6-month Follow-up (with 1 month being defined as 28 days/4 weeks). This estimated duration excludes any washout period.

A schematic of the study design is shown in Figure 1.

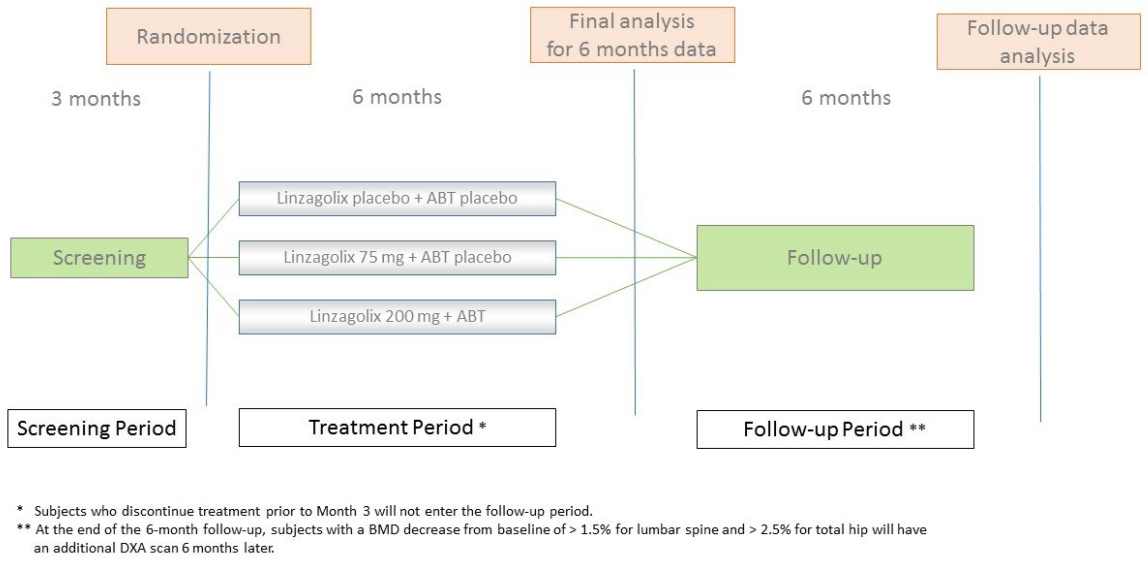

Figure 1: Study design

All subjects will receive once daily either linzagolix 75 mg alone (with ABT placebo) or linzagolix 200 mg combined with ABT, or placebo (linzagolix placebo with ABT placebo) for 6 months. ABT is a combination of estradiol (E2) 1 mg and norethisterone acetate (NETA) 0.5 mg. Linzagolix or its corresponding placebo will be supplied as tablets for oral administration. ABT or its corresponding placebo will be supplied as capsules for oral administration.

Linzagolix/placebo treatments will be provided in monthly treatment kits, packaged, labelled and administered in the same manner to protect the blinded nature of the trial.

ABT/placebo treatments will be supplied as 3-monthly kits, packaged, labelled and administered in the same manner to protect the blinded nature of the trial.

Blinded treatment kit numbers corresponding to each subject's randomized treatment will be provided through an IWRS.

2.2. Randomization Methodology

Randomization will be performed via a centralized IWRS. Subjects will be randomized on Day 1 to one of three treatment groups in a 1:1:1 ratio for linzagolix 75 mg, linzagolix 200 mg with ABT and placebo.

Subjects will be randomized into permutated blocks of a pre-determined length.

Subjects participating in the extension phase and who are in the placebo treatment group will be randomized at the Month 6 visit to either linzagolix 75 mg or linzagolix 200 mg with ABT in a 1:1 ratio. Subjects participating in the extension phase and who are in the linzagolix treatment groups will remain on the same dose that they were taking at the end of Month 6 and will undergo a dummy randomization in order to maintain the blind.

There will be no stratification.

2.3. Stopping Rules and Unblinding

Subjects will be informed that they have the right to withdraw from the study at any time, without prejudice to their medical care, and that they are not obliged to state the reason(s). Any withdrawal must be fully documented in the eCRF exit form.

The Investigator may withdraw a subject at any time if this is considered to be in the subject's best interest.

In addition, the sponsor could make a decision to temporarily or permanently discontinue a patient or the study for safety, ethical, compliance or other reasons. In this case, the subject's participation may be ended prematurely without asking for her consent.

In addition, an Independent Data Monitoring Committee (IDMC) will regularly review safety data and make recommendations concerning the continuation, modification or termination of the study.

After all subjects have completed the Treatment Period (Month 6 or Withdrawal visit), a complete analysis with unblinded treatments will be performed. A database lock will be performed prior to unblinding of treatment.

Discontinuation criteria

During the course of the study, a subject may be discontinued for the following reasons:

- Adverse Event
- Subject's request
- Protocol Violation
- Lost to Follow-up

- Pregnancy

- Other

Details are provided in the Protocol.

Discontinuation Rules at Day 1: Subjects presenting with alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma glutamyl transferase (GGT) or total bilirubin (TBL) ≥ 2 upper limit of normal (ULN) should be discussed with the Sponsor and may have to discontinue study treatment if these results are indicative of liver involvement.

Subjects presenting with a clinically significant finding in the biopsy sample collected at screening and who have completed Day 1, will have to discontinue study treatment.

Discontinuation Rules during the Treatment Period:

Endometrial biopsies: in case of endometrial biopsy diagnosis being an endometrial hyperplasia of any type or worse, the subject will have to discontinue the treatment (and will not be eligible to enter the extension study) and will be advised to consult for a gynecological evaluation and treatment.

Serum calcium: Subjects who have a serum calcium level on treatment above 2.9 mmol/L should have calcium supplements interrupted. If serum calcium level on treatment is above 3.1 mmol/L, calcium supplements and study treatment are to be interrupted. A repeat test of this parameter within 2 weeks under fasting conditions is to be performed. If the results of the repeat remain above 2.9 mmol/L, study treatment should definitively be discontinued and the subject should be advised to consult an endocrinologist for further evaluation.

Bone mineral density loss: subjects who experience more than 8% BMD loss or a Z-score ≤ -2.5 at any site (femoral neck, hip or spine) will be discontinued from study treatment and will enter the follow-up period (they will not be eligible to enter the extension study).

Liver function tests: following the Food and Drug Administration (FDA) guidance on drug-induced liver injury, subjects who have an elevation of hepatic enzymes are to be withdrawn immediately from treatment if:

- ALT or AST >8xULN
- ALT or AST >5xULN for more than 2 weeks
- ALT or AST >3xULN and (TBL >2xULN or international normalized ratio INR >1.5)
- ALT or AST >3xULN with the appearance of fatigue, nausea, vomiting, right upper quadrant pain or tenderness, fever, rash, and/or eosinophilia (>5%)

Withdrawn subjects will be followed up until hepatic parameters return to normal.

ECG : subjects who have a QTcF > 500 ms or increase > 60 ms from the highest value prior to first dose will be withdrawn from study treatment and followed up until return to QTcF < 480 ms or – if not reached after 3 months of treatment cessation- be referred to a cardiologist.

2.4. Study Procedures

The schedule of assessments, as outlined in the study protocol, is provided in Table 1 and Table 2.

Table 1 Schedule of Assessments – Screening and Treatment Periods

Schedule of study assessments – Screening and Treatment Periods								
Timing[1]	**Screening Period (up to 3 months**[2]**)**	**Treatment Period**						
		Day 1	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
Informed Consent	x							
Inclusion-Exclusion criteria	x	x						
Demography, height, weight, medical history	x							x[3]

[1] All post-baseline visits should take place at the end of the defined period (i.e., M1 visit should be scheduled at the end of Month 1, M2 visit should be scheduled at the end of Month 2, etc.). Post-baseline visits dates are calculated from Day 1 visit date. A month is defined as 28 days/4 weeks. Visits should be scheduled ± 3 days from the calculated date.

[2] If required for washout of oral contraceptives, other sex hormones, or GnRH antagonists/agonists (provided that the reason for discontinuing the previous GnRH-antagonist is not lack of efficacy), a period of up to 3 months is allowed between signing the informed consent and the screening visit.

[3] Only weight will be recorded.

Schedule of study assessments – Screening and Treatment Periods

Timing[1]	Screening Period (up to 3 months[2])	Treatment Period						
		Day 1	M1	M2	M3	M4	M5	M6
mB&B	x							
Columbia-Suicide Severity Rating Scale	x	x	x	x	x	x	x	x
ECG	x	x[4]	x	x	x	x	x	x
Physical examination	x				x			x
Vital signs	x	x	x	x	x	x	x	x
Urine pregnancy test	x	x	x	x	x	x	x	x
TVUS of uterus	x				x			x
Gynecological examination	x				x			x

[4] ECG on Day 1 should be performed twice: once before treatment and once after treatment (just before PK sampling).

Sponsor:	**ObsEva**		
Protocol:	**18-OBE2109-003**		
Document Version No.:	V5.0	**Document Date:**	13-JUN-2022

Schedule of study assessments – Screening and Treatment Periods

Timing[1]	Screening Period (up to 3 months[2])	Treatment Period						
		Day 1	M1	M2	M3	M4	M5	M6
Endometrial biopsy	x							x[5]
Pap test	x							
Breast examination (mammogram if required)	x							x
Clinical laboratory & urinary protein dipstick	x	x[6]	x	x	x[6]	x	x	x[6]
Blood sample for PK[7]		x	x	x	x	x	x	x
BMD by DXA	x							x
Adverse events	x	x	x	x	x	x	x	x

[5] If endometrium thickness in TVUS is ≤ 5 mm, no endometrial biopsy will be necessary.

[6] Overnight fasting is required.

[7] PK samples should be taken after the ECG and before the daily dose of IMP, except on Day 1 when the PK sample should be taken at least 1.5 h after the IMP administration.

Schedule of study assessments – Screening and Treatment Periods

Timing[1]	Screening Period (up to 3 months[2])	Treatment Period						
		Day 1	M1	M2	M3	M4	M5	M6
Previous/concomitant medication	x	x	x	x	x	x	x	x
Contraceptive dispensing and counselling	x	x	x	x	x	x	x	x
Permitted analgesic prescribing/dispensing	x	x	x	x	x	x	x	x
Vitamin D and calcium dispensing		x	x	x	x	x	x	x
Subject eDiary completion training/check	x	x	x	x	x	x	x	x
IMP accountability			x	x	x	x	x	x
Dispense study drug		x	x	x	x	x	x	
Dispense ABT		x			x			

Schedule of study assessments – Screening and Treatment Periods

Timing[1]	Screening Period (up to 3 months[2])	Treatment Period						
		Day 1	M1	M2	M3	M4	M5	M6
EHP-30, EQ-5D-5L and PROMIS[8]		x	x		x			x
mPGIS, PGIC, HRUQ and HRPQ		x[9]	x	x	x	x	x	x
Specific monthly severity questions[8]		x			x			x
SSIQ and PSIQ[8]		x						x
AMH		x						
E2, LH, P4		x	x	x	x	x	x	x
SHBG		x			x			x

[8] ClinROs (Monthly dyspareunia question, PSIQ and HRUQ) will be administered to the subject by the site staff and the responses will be filled in the eDiary. ePROs (EHP-30, EQ-5D-5L, PROMIS, PGIS, PGIC, HRPQ and SSIQ, and specific monthly severity questions) will be filled in by the subject in the eDiary.

[9] PGIC not done at Day 1.

Schedule of study assessments – Screening and Treatment Periods

Timing[1]	**Screening Period (up to 3 months**[2]**)**	**Treatment Period**						
		Day 1	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
Bone biomarkers		x			x			x
Question regarding treatment received								x

Table 2 Schedule of study assessments - Follow-up Period

Schedule of study assessments – Follow-up Period						
Timing[1]	**M1 FU**	**M2 FU**	**M3 FU**	**M4 FU**	**M5 FU**	**M6 FU**
Previous/concomitant medication	x	x	x	x	x	x
Adverse events	x	x	x	x	x	x
Columbia-Suicide Severity Rating Scale	x	x	x	x	x	x
ECG	x		x			x
Physical examination			x			x
Weight						x
Vital signs	x		x			

[1] All follow-up study visits should take place at the end of the defined period (i.e. M1 FU visit should be scheduled at the end of the first month of follow-up, M2 FU visit should be scheduled at the end of the second follow-up visit, etc.). A month is defined as 28 days/4 weeks. Visits should be scheduled ± 7 days from the calculated date.

Schedule of study assessments – Follow-up Period						
Timing[1]	**M1 FU**	**M2 FU**	**M3 FU**	**M4 FU**	**M5 FU**	**M6 FU**
Gynecological examination			x			
Breast manual examination			x			
Endometrium TVUS			x			x
Endometrial biopsy	x[2]	x[2]	x[2]	x[2]	x[2]	x[3]
Clinical laboratory & urinary protein dipstick	x		x[4]			
Subject eDiary completion check	x	x	x	x	x	x
Subject eDiary collection and deactivation						x

[2] Endometrial biopsy will be taken only if diagnosis at preceding month was different than "benign endometrium" or if no endometrial biopsy was done at Month 6 nor at any visit since Month 6.

[3] An end-of-study biopsy is mandatory if no endometrial biopsy was obtained at M6 nor at any of the subsequent monthly visits.

[4] Overnight fasting is required.

Schedule of study assessments – Follow-up Period

Timing[1]	M1 FU	M2 FU	M3 FU	M4 FU	M5 FU	M6 FU
EHP-30, HRUQ, HRPQ, EQ-5D-5L and PROMIS			x			x
mPGIS and PPGIC	x	x	x	x	x	x
BMD by DXA						x[5]
E2, LH, P4	x		x			
FSH at local laboratory for subjects that do not resume menses at M3 FU visit			x			
Bone biomarkers			x			
Permitted analgesic prescribing/dispensing	x	x	x	x	x	
Vitamin D and calcium dispensing	x	x	x	x	x	

[5] Subjects with a BMD decrease from baseline of > 1.5% for lumbar spine and/or > 2.5% for total hip at Month 6 FU visit will have an additional DXA scan 6 months later.

Schedule of study assessments – Follow-up Period						
Timing[1]	**M1 FU**	**M2 FU**	**M3 FU**	**M4 FU**	**M5 FU**	**M6 FU**
Urine pregnancy test and contraceptive dispensing and counselling	x	x	x			x

2.5. Efficacy, Pharmacokinetic, and Safety Variables

2.5.1. Efficacy Variables

2.5.1.1. Primary efficacy endpoint

The two co-primary, composite, efficacy endpoints are clinically meaningful reduction from baseline to the last 28 days preceding the Month 3 visit (the 4-week period preceding Month 3 visit) or, for subjects who discontinue randomized treatment prior to the Month 3 visit, to the last 28 days of randomized treatment, along with a stable or decreased use of analgesics for EAP, in the mean daily assessment of 1) DYS and of 2) NMPP measured on a Verbal Rating Scale (VRS) using an electronic diary (eDiary).

Additional details on the primary endpoint definition and primary analysis are provided in Section 4.6.

2.5.1.2. Ranked Secondary efficacy endpoints

Ranked secondary efficacy endpoints (in the order of the endpoints to be tested) are:

- Change from baseline to Month 6 in DYS (VRS)
- Change from baseline to Month 6 in NMPP (VRS)
- Change from baseline to Month 6 in dyschezia (Numeric Rating Scale - NRS)
- Change from baseline to Month 6 in overall pelvic pain (NRS)
- Change from baseline to Month 6 in the interference of pain with the ability to perform daily activities, measured using the pain dimension of the Endometriosis Health Profile-30 (EHP-30)
- Change from baseline to Month 6 in dyspareunia (VRS)
- No analgesics use for EAP during the preceding 4-week period at Month 6
- No opiate use for EAP during the preceding 4-week period at Month 6

Additional details on ranked secondary endpoints and analyses are provided in Section 4.6.

2.5.1.3. Additional secondary efficacy endpoints

Additional secondary efficacy endpoints include:

- Clinically meaningful reduction at scheduled visits other than Month 3 for DYS and NMPP
- Change from baseline to each scheduled assessment in the mean pelvic pain scores for DYS, NMPP and overall pelvic pain, during the previous 4-week period assessed on the NRS and VRS

- Change from baseline to each scheduled assessment in the number of days with moderate to severe pelvic pain during the previous 4-week period assessed on the VRS
- Change from baseline to each scheduled assessment in the mean worst pelvic pain score defined as the mean of the 5 highest daily pain scores reported during the previous 4-week period assessed on the NRS
- Change from baseline to each scheduled assessment in the mean of daily dyspareunia scores reported during the previous 4-week period on the dyspareunia VRS
- Change from baseline to each scheduled assessment in the mean of daily dyschezia scores reported during the previous 4-week period assessed on the dyschezia NRS
- Change from baseline to each scheduled assessment in non-opioid, opioid and combined analgesic use for EAP during the previous 4-week period based on pill count in the eDiary
- Change from baseline to each scheduled assessment in opioid analgesic use for EAP as reported in the eDiary during the previous 4-week period based on morphine milligram equivalent (MME)
- Change from baseline to each scheduled assessment in the number of days of analgesic use (including any class) for EAP during the previous 4-week period as assessed in the eDiary
- Change from baseline to each scheduled assessment in the number of days of opioid analgesic use for EAP during the previous 4-week period as assessed in the eDiary
- Change from baseline to each scheduled assessment in the number of pelvic pain-free days (assessed on the VRS) during the previous 4-week period
- Change from baseline to each scheduled assessment in ability to perform daily activities during the previous 4-week period, as assessed in the eDiary (daily function NRS)
- Change from baseline to each scheduled assessment in the number of days with no difficulty in doing daily activities due to EAP during the previous 4-week period as assessed in the eDiary (daily function NRS)
- Change from baseline to each scheduled assessment in the number of days when dyspareunia was a problem during the previous 4-week period (including days when sexual intercourse was avoided because of anticipation of pain) as assessed on the dyspareunia VRS
- Change from baseline to each scheduled assessment in the number of days when sexual intercourse was avoided because of anticipation of pain during the previous 4-week period as assessed on the dyspareunia VRS
- Change from baseline to each scheduled assessment in the number of days with uterine bleeding (including spotting) during the previous 4-week period measured by eDiary

- Change from baseline to each scheduled assessment in the number of days when school or work was missed due to EAP in the previous 4-week period as reported in the eDiary
- Change from baseline to each scheduled assessment in the number of days when the subject had to go to bed or lie down due to EAP in the previous 4-week period as reported in the eDiary
- Change from baseline to each scheduled assessment in the Pain, Control and powerlessness, Emotional well-being, Social support, Self-image dimensions and the Modular sexual relationship questionnaire of EHP-30 scores
- Change from baseline to each scheduled assessment in the Health-Related Productivity Questionnaire (HRPQ) scores
- Number of non-study endometriosis related health visits, number of days in hospital and type of procedures performed based on Health Resource Utilization Questionnaire (HRUQ) at each scheduled assessment
- Change from baseline to each scheduled assessment in the Physician/Subject Surgery Intention Question (PSIQ/SSIQ)
- Change from baseline to each scheduled assessment in the PROMIS Fatigue – Short Form 6a
- Change from baseline to each scheduled assessment in the EuroQoL 5-Dimension 5-Level (EQ-5D-5L) questionnaire
- Response at each scheduled assessment according to Patient Global Impression of Change (PGIC) (and Post-treatment Patient Global Impression of Change, PPGIC)
- Change from baseline to each scheduled assessment in the monthly PGIS (mPGIS) score

Additional details on secondary endpoints and analyses are provided in Section 4.6.

2.5.2. Pharmacokinetic Variables

PK blood samples will be collected from each subject for determining linzagolix and KP017 plasma levels.

On Day 1, blood samples for PK assessment will be taken at least 1.5 h post-first dose.

During the treatment period, on days of site visits (Months 1, 2, 3, 4, 5 and 6), the subject will be asked to take the dose of IMP at site, after a pre-dose PK sampling. The approximate time of dose administration on the four previous days and time of PK sampling will be recorded in the eCRF.

Pharmacokinetic analyses will be considered exploratory. Analyses are described in Section 4.7.

2.5.3. Safety Variables

Safety endpoints include:

- Change from baseline to each scheduled assessment in BMD measured by DXA of lumbar spine (L1-L4), femoral neck, and total hip
- Incidence and severity of treatment emergent adverse events (TEAEs)
- Incidence and severity of hypoestrogenic TEAEs (hot flush)
- Time to the first post-treatment menses
- Changes in clinical laboratory assessments (hematology, biochemistry, coagulation parameters, hormones, lipids and urinalysis) from baseline to each scheduled assessment
- Any pathological changes from baseline in the endometrium as assessed by histology from endometrial biopsies
- Changes from baseline to each scheduled assessment in any other safety parameter including weight, vital signs, ECG, gynecological assessments and endometrial thickness

Safety analyses are described in Section 4.8.

2.5.4. Exploratory Variables

Exploratory endpoints include:

- Change from baseline in bone turnover markers at each scheduled assessment.

Exploratory analyses are detailed in Section 4.9.

3. SUBJECT POPULATIONS

3.1. Analyses Sets Definitions

The following data sets will be used for the statistical analysis:

Screened Set: All screened subjects.

Randomized Set: All randomized subjects.

Safety Set (SAF): All randomized subjects who received at least one dose of double-blind study drug irrespective of the treatment received. Subjects will be analyzed according to treatment received.

Full Analysis Set (FAS): All randomized subjects who received at least one dose of double-blind study drug irrespective of the treatment received. This is the same as the Safety Set, but for the FAS, subjects will be analyzed according to randomized treatment.

Threshold Analysis Set: All subjects from the FAS with any anchor data at both Baseline and Month 3 or early discontinuation visit, or available Month 3 or early discontinuation visit PGIC data only.

Anchor data for the Threshold Analysis Set includes monthly Patient Global Impression of Severity (mPGIS), mPGIC, monthly DYS, monthly NMPP.

M6 Threshold Analysis Set: All subjects from the FAS with any anchor data at both Baseline and Month 6 or early discontinuation visit, or only available Month 6 or early discontinuation visit PGIC data.

Anchor data for Month 6 include monthly Patient Global Impression of Severity (mPGIS), mPGIC, monthly DYS, monthly NMPP, monthly dyschezia, monthly overall pelvic pain, monthly difficulty in daily activities, and monthly dyspareunia.

Per Protocol (PP) Set: All randomized subjects who completed the six months of treatment excluding those identified as having major protocol deviations that could potentially affect the efficacy assessments up to Month 6. Subjects will be analyzed according to randomized treatment.

Follow-up (FU) Set: All randomized subjects who completed the six months of treatment or discontinued treatment between Month 3 and Month 6 and entered the drug free follow-up period. Subjects will be analyzed according to randomized treatment.

Subjects who entered the drug free follow-up period are subjects eligible for the 6-month treatment free follow-up and with any data post month 6 or post treatment discontinuation.

Follow-up Safety Set (FU SAF): All randomized subjects who completed the six months of treatment or discontinued treatment between Month 3 and Month 6, and entered the drug free follow-up period. Subjects will be analyzed according to treatment received.

Pharmacokinetic (PK) Set: all subjects who received active study medication, had no major protocol deviations impacting PK evaluation and with available PK data.

In general, analyses of efficacy will be conducted using the FAS, PP and FU sets, analyses of safety will be conducted using the Safety Set and FU SAF, and analyses of PK will be conducted using the PK Set.

The actual treatment will be derived as follows, based on daily e-diary IMP intake data:

Subjects received a different kit of active linzagolix or placebo for each month, and a different kit for add-back or placebo for each 3-month period.

The number of days with one pink tablet, one grey tablet and one red capsule taken based on e-diary IMP intake data for each treatment will be computed. Data from “Today” will be used primarily. If data from today is missing, data from “Yesterday” of the following day will be considered. If the e-diary is not completed for a day (neither “Today” nor “Yesterday” of following day), it will be assumed that no intake of drug was taken on that day.

Further rules are as follows:

- Treatment administration errors:
    - If a subject received placebo with placebo ABT in more than 50% of days, then the actual treatment group will be placebo with placebo ABT.
    - Otherwise, the active treatment group (combined with add-back or placebo add-back) with the most days taken will be assigned.
    - If the number of days is equal for two treatments groups containing active treatment then LGX 200 mg with ABT will be assigned.
- If a subject discontinued, only data received up to discontinuation will be used.
- It is possible that the derived actual treatment group does not correspond to any of the planned treatment groups. In that case actual treatment group will be assigned to the planned treatment group with same dose of active treatment. Example: LGX 75mg with ABT will be assigned to LGX 75mg with placebo ABT; LGX 200mg with placebo ABT will be assigned to LGX 200mg with ABT. A listing with such cases will be provided.

3.2. Protocol Violations

A protocol deviation is any noncompliance with the clinical trial protocol or International Conference on Harmonization Good Clinical Practice (ICH GCP) requirements. It is the responsibility of the site investigator to use continuous vigilance to identify and report deviations. All deviations must be addressed in study source documents and reported to the reviewing IRB/IEC per their policies. The site investigator is responsible for knowing and adhering to the reviewing IRB/IEC requirements.

All Protocol Deviations will be reported to the Sponsor and documented in the monitoring report. These will be classified as minor or major based on their effect on the right, safety or well-being of the subjects and/or the quality and integrity of the data, and the final rating of all deviations will be confirmed prior to database lock. The final protocol deviation file (formatted as a Microsoft Excel file) will be provided to Cytel.

Major protocol deviations will be summarized by category (if available) and by treatment group for the FAS population.

Major deviations that could potentially affect the efficacy assessments up to Month 6 and thus may exclude subjects from the PP Set may include:

- Non-compliance with inclusion criteria 3; 6; 8
- Non-compliance with exclusion criteria 2 to 10 inclusive; 12; 13; 17; 19; 20; 21; 23; 24; 25
- Non-compliance with study treatment: threshold to be defined during blinded data review meeting (e.g. IMP compliance in tablets <60% for Day1-Month 6)
- IP Dispensing error;
- Unauthorised prior or concomitant therapy;
- M3 and/or M6 Visits performed outside the +/-14 days of the theoretical visit date;
- Randomisation code broken;

All protocol deviations will be presented in the data listings.

4. STATISTICAL METHODS

4.1. Sample Size Justification

The planned sample size for this study is 150 subjects per treatment group (450 subjects in total).

An overall two-sided type I error of 0.05 will be used. As there are two linzagolix versus placebo comparisons, Bonferroni corrected p-values will be produced (raw p-values will be multiplied by two prior to comparing to 0.05). The planned sample size considers the hierarchical, fixed sequence testing of the ranked secondary endpoints as well as the co-primary endpoints. The assumptions used for the sample size calculations are based on analyses of clinically meaningful reduction in pain with a stable or decreased use of analgesics from the Phase 2b Edelweiss study. Calculations were performed using East® 6.5 software.

One hundred and fifty (150) subjects per treatment group will provide a power greater than 95% to reject the null hypothesis for both co-primary endpoints for either treatment group, assuming a placebo response rate of 14.6% and an active treatment response rate of 48.6% (75 mg, Edelweiss Phase II study results) or 64.7% (200 mg, Edelweiss Phase II study result) for DYS, and a placebo response rate of 18.8% and an active treatment response rate of 42.1% (75 mg, Edelweiss Phase II study result; the response rate for 200 mg in Edelweiss was lower but was inconsistent with the other doses and other timepoints for the same dose and so was not used) for NMPP. In addition, 150 subjects per treatment group provides 85% power to reject the null hypotheses for all the ranked secondary endpoints based on the observed results from the placebo and 200 mg treatment group in the Edelweiss study.

The values used for the sample size calculations assume that these responses rates are what would be seen on average when under treatment including taking into account any subjects who might withdraw from treatment early; therefore the calculations do not need to be further adjusted for dropouts.

4.2. General Statistical Methods and Data Handling

4.2.1. General Methods

All output will be incorporated into Microsoft Excel or Word files, sorted and labeled according to the International Conference on Harmonisation (ICH) recommendations, and formatted to the appropriate page size(s).

Tabulations will be produced for appropriate demographic, baseline, efficacy, pharmacokinetic and safety parameters.

For continuous data and for ordered categorical data, if appropriate, the number of non-missing observations, mean, standard deviation, median, first and third quartiles, minimum and maximum will be calculated, including for change from baseline when applicable. The baseline mean will be calculated

for all subjects based on the FAS. For efficacy endpoints and BMD, a baseline mean will also be calculated for each visit using the baseline data for the subset of subjects who attended that visit, such that the same subjects contribute to the mean for the visit and the mean for the corresponding baseline values. Summary statistics will be based primarily on non-missing values. For ordered categorical data and nominal data, absolute counts and relative frequencies (in %) will be calculated.

This study will be conducted at approximately 100 study sites, resulting in only a few subjects per site on average. Subjects from all sites will be pooled and no adjustment for sites will be carried out in the statistical analyses.

Raw and derived data will be listed.

All statistical hypothesis tests and confidence intervals will be two sided.

4.2.2. Computing Environment

All descriptive statistical analyses will be performed using SAS statistical software (Version 9.4), unless otherwise noted. Medical History and adverse events will be coded using MedDRA available version. Concomitant medications will be coded using World Health Organization (WHO) Drug Dictionary available version.

4.2.3. Methods of Pooling Data

Not applicable to the present study.

4.2.4. Adjustments for Covariates

Adjustments for possible covariates will be described in the appropriate statistical analyses' sections.

4.2.5. Multiple Comparisons/Multiplicity

All statistical hypothesis tests and confidence intervals will be two-sided. To maintain an overall type I error rate of 0.05, as there are two linzagolix versus placebo comparisons, Bonferroni corrected p-values will be produced (raw p-values will be multiplied by two prior to comparing to 0.05), along with corresponding 97.5% confidence intervals.

The two co-primary efficacy endpoints will be clinically meaningful reduction from baseline over the last 28 days of randomized treatment up to the Month 3 visit (i.e., the 4-week period preceding the Month 3 visit) in the mean daily assessment of DYS and of NMPP measured on a VRS, along with a stable or decreased use of analgesics for EAP.

The primary objective will be assessed by testing the following two-sided hypotheses for each co-primary endpoint, for each linzagolix group versus placebo separately:

Null hypothesis (H0): There is no difference in the odds of subjects meeting the primary endpoint for the active treatment group compared to placebo, that is, the odds ratio (OR) equals 1.

H0: OR =1

Alternative hypothesis (H1): There is a difference in the odds of subjects meeting the primary endpoint for the active treatment group compared to placebo, that is, the OR is not equal to 1.

H1: OR ≠ 1

Where odds are equal to the proportion of subjects meeting the endpoint divided by proportion of subjects not meeting the endpoint.

Each linzagolix group will need to demonstrate a statistically significant difference for both co-primary endpoints for the group to be considered more efficacious than placebo, thus maintaining an overall type I error rate of 0.05.

For each linzagolix group that is statistically significantly more efficacious than placebo for the co-primary endpoints, a fixed-sequence testing strategy shall be used within the group to test the ranked secondary endpoints as ordered in Section 2.5.1.2, to maintain the family-wise type I error rate. That is, the comparison for each linzagolix group versus placebo for each ranked endpoint will only be declared statistically significant different if the raw p-value multiplied by two is less than or equal to 0.05 for that endpoint and for all preceding endpoints for that dose versus placebo.

Response rates for DYS and NMPP at each visit will be analyzed to evaluate the changes in pain over time, with nominal Bonferroni corrected p-values used in the comparisons of treatment groups at each visit. These are supportive analyses and will primarily be considered as descriptive. The analyses do not form part of the fixed-sequence strategy being used for the ranked secondary endpoints and will not be fully controlled for an overall type I error rate. However, to maintain consistency with the primary and secondary analyses, Bonferroni corrections will be used due to there being two linzagolix versus placebo comparisons for each endpoint / time point.

4.2.6. **Subpopulations**

Subgroup analyses will be performed for each co-primary endpoint and for BMD.

Subgroups include:

- Race: Black or African American, Other
- Ethnicity: Hispanic or Latino, Not Hispanic or Latino
- Age group 1 (<median value at baseline, ≥ median value at baseline) (based on FAS population)
- Age group 2 (<25, ≥ 25 to < 35, ≥ 35 to < 45, ≥ 45)
- Weight: < median value at baseline, ≥ median value at baseline (based on FAS population)
- BMI group 1 (< median value at baseline, ≥ median value at baseline) (based on FAS population)
- BMI group 2 (< 20, ≥ 20 to < 25, ≥ 25 to < 30, and ≥ 30)

- Baseline analgesic use: Subjects who took no rescue medication for endometriosis pain at baseline, subjects who took any NSAIDS and Narcotics at baseline
- Baseline pain VRS scores for dysmenorrhea (DYS), non-menstrual pelvic pain (NMPP), and dyspareunia: < median value at baseline, ≥ median value at baseline (based on FAS population)
- Time since endometriosis diagnosis (time in years): < 2 years, ≥ 2 – < 5 years, and ≥ 5 years
- History of pregnancy at baseline – Yes, No

A subgroup analysis will be performed for the secondary endpoint of dyspareunia on subjects who have a mean dyspareunia score >1 at baseline. Detailed analysis is described in Section 4.6.

4.2.7. Withdrawals, Dropouts, Loss to Follow-up

In general, there will be no substitutions made to accommodate missing data points due to withdrawals, dropouts and loss to follow-up. All data recorded on the CRF will be included in data listings that will accompany the clinical study report.

In case of early discontinuation, the subject will be instructed to complete the relevant eDiary questionnaires as soon as possible upon treatment discontinuation.

Subjects who did not receive the study drug will be withdrawn from the study and no further study procedures will be performed.

Withdrawal during treatment period:

Subjects who discontinue between Day 1 and Month 3 should undergo the procedures required at Month 6 (except the DXA scan and the endometrium biopsy if a biopsy was obtained at screening) and will not enter the 6-month follow-up period.

Subjects discontinuing the study treatment between Month 3 and Month 6 should undergo the procedures required at Month 6. These subjects will enter a 6-month follow-up period and will continue daily eDiary recording for 6 months and up to Month 6 FU visit in order to continue to collect efficacy data.

PK sampling is not required if the subject had no IMP administration the day prior to the withdrawal visit.

Withdrawal during follow-up period:

Subjects discontinuing participation in the study during the follow-up period should undergo the procedures required at Month 6 FU visit, which includes completion of all ClinRO and ePRO questionnaires, except the DXA in case of discontinuation before Month 3 FU visit.

Subject Replacement:

Discontinued subjects who did not receive the study drug may be replaced.

Additional subjects may be recruited due to subjects who were discontinued due to ALT, AST, GGT or total bilirubin ≥ 2 times the upper limit of normal at Day 1, or due to a clinically significant biopsy finding at Day 1 (treatment start), or due to high calcium level at Day 1.

4.2.8. Missing, Unused, and Spurious Data

4.2.8.1. Efficacy Endpoints Missing Data

Summary statistics will be based primarily on non-missing values. The planned statistical analyses aim to estimate the efficacy of linzagolix versus placebo while under randomized treatment and therefore lack of data after treatment discontinuation will not be considered as missing for these analyses.

Subjects who cannot be assessed as responders for the co-primary efficacy endpoints, e.g. due to lack of on treatment pain data (i.e. subjects who received less than 28 days of treatment), will be considered as non-responders. In addition, for endpoints assessing the numeric change from baseline, such subjects will be assigned a change from baseline of zero. It is expected that very few subjects will be affected by these rules.

There may be missing data during the screening period and whilst under treatment, for example, missed days of completing the eDiary. The primary analyses will use observed data only in such cases provided that a minimum number of completed daily eDiary entries are available. This effectively assumes that the individual values such as daily eDiary assessments (observed or not) are independent of being missing.

Further details on the handling of missing values and the methods and planned sensitivity analyses to check the robustness of the analysis results under alternative assumptions with regards to missing data are provided in the analyses sections. As additional analyses, in order to estimate the effect of treatment policy, a reference based multiple imputation approach will be used for the co-primary endpoints for subjects who discontinue the study early under the assumption that the efficacy of the linzagolix treated subjects gradually transitions to that observed in the placebo subjects.

Missingness Assumptions:

Missing At Random (MAR): Missingness is independent of unobserved outcomes given observed data. Subjects with missing data can be modeled based on similar subjects with available data if we account in the model for relevant observed factors.

Missing Not At Random (MNAR): The probability of missingness depends on the unobserved outcome and cannot be predicted only based on observed data.

Multiple Imputation Methodology:

Multiple imputation inference involves three distinct steps:

1. The missing data are filled in m times to generate m complete data sets, using defined assumptions on missingness.

2. The m complete data sets are analyzed using standard statistical analyses.

3. The results from the m complete data sets are combined to produce inferential results.

Reference-based multiple imputation:

Imputation model for the missing observations in the active treatment groups that is constructed not from the observed data in the active treatment groups but rather from the observed data in the control group. Missing observations from the control group are also imputed using observed data from the control group.

The fully conditional specification method will allow multivariate imputation of the different endpoints.

Application of multiple imputation for binary endpoints:

The two co-primary endpoints are binary endpoints derived from underlying continuous endpoints. The first step of multiple imputation will be performed on the continuous variables and transformation into a binary endpoint will be applied after imputation.

As the second step, the statistical analysis for treatment effect on binary endpoints (logistic regression) will be performed on each m imputed datasets.

Odds ratios of the contrast of interest will be presented. To obtain the combined estimates of odds ratios from each of the m models, the log of odds ratios estimates should be used in the MIANALYZE procedure, and the combined estimates should be back transformed afterwards. This is in order to fulfill the assumption that the statistics estimated from each imputed dataset are normally distributed (Ratitch, Lipkovich, & O'Kelly, 2013)

4.2.8.2. Adverse events and Concomitant Medication Missing Dates

Imputation of missing/partial AE and CM dates will be done only to identify treatment-emergent AEs.

AE onset dates:

- Partially missing AE onset dates will be imputed as follows:
    - When only Day is missing:
        - If Month & Year of the onset date are the same as Month & Year of the first administration date, the imputed onset date will be imputed as the minimum of the first administration date and the AE resolution date (imputed if needed).

- Else if the subject entered the Extension Study and Month & Year is the same as Month & Year of the first administration date in Extension, the imputed onset date will be imputed as the minimum of the first administration date in Extension and the AE resolution date (imputed if needed).
- Otherwise, the missing day will be replaced by "1"

- When Day & Month are missing:
  - If the subject entered the Extension Study and if Year of the onset date is the same as Year of the first administration date in Extension, the imputed onset date will be imputed as the minimum of the first administration date in Extension and the AE end date (imputed if needed).
  - Else if Year of the onset date is the same as Year of the first administration date, the imputed onset date will be imputed as the minimum of the first administration date and the AE end date (imputed if needed).
  - Otherwise, the missing Day & Month will be replaced by "01 JAN."

- Completely missing onset dates for AEs
  - If the subject entered the Extension, the AE onset date will be imputed by the first administration date in Extension and the AE will be considered as treatment-emergent, unless the end date of the AE (imputed if needed) or the end year of the AE (if day and month are missing) is entered and is before the first administration date in Extension.
  - If the subject did not enter the Extension, or if the end of the AE is before the first administration date in Extension, the AE onset date will be imputed by the first administration date and the AE will be considered as treatment-emergent, unless the end date of the AE (imputed if needed) or the end year of the AE (if day and month are missing) is entered and is before the first administration date. If the end date is before the first administration date, the AE will not be considered as treatment-emergent.

AE end dates

- If Day only is missing, incomplete end dates will be replaced by the last day of the month, if not resulting in a date later than the date of the subject's death or date of study discontinuation. In the latter case, the date of death/date of study discontinuation will be used to impute the incomplete end date.
- If Day & Month are missing, Day & Month will be replaced by 31DEC, if not resulting in a date later than the date of the subject's death or date of study discontinuation. In the latter case, the date of death/ date of study discontinuation will be used to impute the incomplete end date.
- In all other cases the incomplete end date will not be imputed.

Concomitant Medications dates

Partially missing dates for prior and concomitant medications and procedures will be imputed. Of note, imputation of missing/partial dates will be done only to identify the treatment period assignment.

- End date: Missing day will be imputed as the last day of the month, and missing month will be imputed by December, if not resulting in a date later than the date of the subject's death or date of study discontinuation. In the latter case, the date of death/date of study discontinuation will be used to impute the incomplete end date.
- Start date: Missing day will be imputed as the first day of the month, and missing month will be imputed by January.
- If the start date is completely missing, then:
    - If the end date is prior to the date of first administration of the study drug, then the medication is considered as prior
    - If the end date is prior to the date of last administration of the study drug, then the medication is considered as prior and concomitant
    - If the end date is completely missing or after the date of last administration of the study drug, then the medication is considered as prior, concomitant and post.

4.2.8.3. eDiary Devices Data Cleaning

To handle eDiary system issues, mapping rules are defined for data analysis, as described in Appendix E in Section 7.5.

In addition, the following rules are defined.

Handling of duplicate records

- A change of eDiary device vendor ([redacted] vendor to Signant Health vendor) occurred during the study.
- Some subjects may have entered data in the two different devices during the transition period. In case of duplicates results on the same day for daily diaries or on the same visit for monthly questionnaires, data from the new vendor (Signant Health) will be used for analysis.
- In case of duplicates within the same device (same date, same name), the first non-empty record (with earliest date/time) should be used for analysis.

Daily Records entered after 00:00

For daily diaries, records entered with a time between 00:00 (midnight) and 02:00 AM should be considered for the previous day.

Additional rules to specific eDiary domains:

Analgesic use data:

- Analgesic use data from “Today” will be used primarily. If data from ”Today” is missing, data from “Yesterday” of the following day will be considered

IMP Intake Data:

- IMP Intake data from “Today” will be used primarily. If data from ”Today” is missing, data from “Yesterday” of the following day will be considered

[redacted] device specific rules

NOTE: by design, when a subject is completing a questionnaire (daily or monthly) on a [redacted] device, the device automatically saves all answers provided by the subject as soon as recorded, unless the subject uses the “previous” button to come back to already answered question(s) within the questionnaire and changes her answer(s) prior to finishing the questionnaire.
There are several triggering questions in the study questionnaires: the next question(s) presented to the subject will depend on the answer chosen by the subject to the current question.

Due to device design issue, when an answer to a triggering question has been changed by the subject (using previous button), the device keeps subsequent answer(s) to the previously triggered question(s), even if this/these question(s) is/are no longer relevant considering the new answer to the triggering question.

The agreed rules consider the last entry made by the subject i.e. the entry that would have not allowed her to provide additional data if it had been her initial answer (in Signant Health device, the intermediate answers are not saved and will not be present in the database; the agreed rules are aligned with this functioning).

Analgesic use data:

- If answer to question “Did you take any provided/prescribed analgesic for your endometriosis related pain?” is “Yes” or empty but the number of ibuprofen intakes and narcotic analgesics intakes are both 0, it will be considered that no analgesic was used on that day.
- If answer to question “Did you take any provided/prescribed analgesic for your endometriosis related pain?” is “Yes” or empty, but question to Ibuprofen intake is “No” and the number of ibuprofen intakes is not zero, it will be considered that no ibuprofen was used on that day. The same rule will be applied for narcotic analgesics.

- If answer to question “Did you take any provided/prescribed analgesic for your endometriosis related pain?” is “No” and the number of ibuprofen intakes and narcotic analgesics intakes are not 0, it will be considered that no analgesic was used on that day and following answers will be ignored.
- If answer to “Did you take any provided/prescribed analgesic for your endometriosis related pain?” is “No” and answer to question Ibuprofen intake is “Yes”, but the number of Ibuprofen intakes is 0, it will be considered that no ibuprofen was used on that day. The same rule will be applied for narcotic analgesics.

Dyspareunia VRS Data:

- when question “Did you have any sexual intercourse?” is answered “No, I was not sexually active for reasons other than my endometriosis” or “No, I avoided sexual intercourse because of anticipation of pain” and there is an answer to question “How did your endometriosis related pain interfere with sexual intercourse?”, only the “No” answer will be considered.
- when question “Did you have any sexual intercourse?” is blank and there is an answer to question “How did your endometriosis related pain interfere with sexual intercourse?”, consider that “Yes” was answered to “Did you have any sexual intercourse”.

HRUQ Data:

- when answer to triggering question 1 is “No” and at least one answer to question 2 to 6 (any answer different from 0) is present, ignore answer 2-6 in the analysis
- when answer to triggering question 7 is “No” and at least one answer to questions 8 to 9 (any answer different from 0) is present, ignore answers 8-9 in the analysis
- For questions 8-9, consider the number of times whatever the answer provided for “procedures” (i.e., if it is NO include number of times if different than 0)

HRPQ Data:

- when answer to question 9 is different than “endometriosis symptoms or its treatment(s) forced me to work part-time when I wanted to work full-time” and question 9bis is different than 0, ignore answer 9bis in the analysis.
- when answer to question 9 is different than “endometriosis symptoms or its treatment(s) kept me from having a job when I wanted to work full-time” and question 9ter is different than 0, ignore answer 9ter in the analysis.
- when answer to question 9 is different than “endometriosis symptoms or its treatment(s) kept me from having a job when I wanted to work part-time” and question 9quater is different than 0, ignore answer 9quater in the analysis

4.2.8.4. Transvaginal Ultrasound

If the depth recorded for a TVUS of uterus is 0 mm, the depth and the volume should be considered as missing (i.e., this means that the assessment was done in 2 dimensions instead of 3 dimensions).

4.2.8.5. Partial and Missing Dates of diagnosis

For partial date of first medical diagnosis/treatment, first surgical diagnosis and most recent surgical diagnosis the following rules of imputation will be applied:

- If only Day is missing, Day will be imputed to 01.
- If Day and Month are missing:
    - if Year is the same as the screening date, Date will be imputed to Screening Date.
    - Otherwise, Day will be imputed to 01 and Month will be imputed to 07.
- If Date is completely missing, no imputation will be done.

4.2.9. Visit Windows

For all analyses except for the DXA assessments and the diary questionnaires, the visits as collected in the eCRF will be used.

E-diary questionnaires:

For diary questionnaires (EHP-30, EQ-5D-5L, HRPQ, HRUQ, PGIC, PPGIC, PGIS, PROMIS, SSIQ) except for the PSIQ, the e-diary dates of assessment of each visit will be checked versus the dates of same visit in eCRF to handle for e-diary visit errors. Only assessments within the window of the date of eCRF+/-14 days will be used in analyses. For Day 1 assessments, only data prior to the first dose will be used, regardless of the questionnaire visit name. In case of multiple assessments falling within the same window of an eCRF date, the non-empty record with completion date closest to the eCRF date will be used.

For data from discontinued visits, the discontinuation visit date should be compared to theoretical visit dates. Only assessments within the window of theoretical date +/-14 days will be used for analyses, if it is planned to have a such assessment at this visit and no other assessment is in the window. If a subject discontinued in the main treatment period, date will be compared to the theoretical visit dates in main treatment period as described in Table 3. If a subject discontinued in the follow-up period, date will be compared to the theoretical visit dates in Follow-up period from Table 4. Note that assessments from

discontinued visits in [redacted] eDiary device are entered under Month 6 (Month 6 FU in follow-up). They will be recoded as discontinued visits before applying the windowing rules.

Table 3 Theoretical Visit Dates in Main Treatment Period

Visit	Theoretical Date
Month 1	Day 29
Month 2	Day 57
Month 3	Day 85
Month 4	Day 113
Month 5	Day 141
Month 6	Day 169

Table 4 Theoretical Visit Dates in Follow-up

Visit	Theoretical Date
Month 1 FU	Month 6 date or treatment discontinuation date + 28
Month 2 FU	Month 6 date or treatment discontinuation date + 56
Month 3 FU	Month 6 date or treatment discontinuation date + 84
Month 4 FU	Month 6 date or treatment discontinuation date + 112
Month 5 FU	Month 6 date or treatment discontinuation date + 140
Month 6 FU	Month 6 date or treatment discontinuation date + 168

Pharmacokinetics data:

Visit windows rules for PK data will be the following:

- At day 1: PK assessment will be included in analysis if PK sampling is done >=1.5 hours after the dose and on the same day as Day 1.
- At other visits: PK assessments will be included in analysis if:
    - Assessment is Pre-dose (or PK time = dose time), or IMP intake is not done on the day of PK sampling (i.e. considered as pre-dose), and,
    - PK assessment date is within +/-10 days of the theoretical visit and up to 2 days after last IMP intake for Month 6 visit.

If the PK sampling time is missing, or the IMP intake on the same day is done and the IMP intake time is missing then assessments will not be included.

IMP intake status (done/not done) and IMP intake time from the Pharmacokinetics eCRF page will be used.

DXA data:

Visit windows rules for DXA assessments will be the following:

- The baseline DXA assessment will be derived as the latest non-missing DXA assessment with acceptable quality (accepted="Yes") prior to the baseline visit or the first non-missing assessment with acceptable quality (accepted="Yes") done up to 10 days after the baseline visit if there is no assessment prior to baseline..
- The DXA assessment at Month 6 will be derived as the first non-missing DXA assessment with acceptable quality (accepted="Yes") and dated between Month 6 theoretical date (=Day 169) - 28 days and Month 6 theoretical date + 28 days.
- The DXA assessment at Month 6 Follow-Up will be derived in the same way as the DXA assessment at Month 6. Theoretical date of Month 6 Follow-up is 168 days after Month 6 theoretical date or after treatment discontinuation.

Other Safety Data:

Endometrial biopsy Screening window will include any assessment done prior to the first dose.

For other safety data, data from early discontinuation should be compared to theoretical visits dates. Only assessments within the window of theoretical date +/-14 days will be used for analyses, if it is planned to have a such assessment at this visit and no other assessment is in the window. If a subject discontinued in the main treatment period, date will be compared to the theoretical visit dates in main treatment period as described in Table 3. If a subject discontinued in the follow-up period, date will be compared to the theoretical visit dates in Follow-up period from Table 4. For unscheduled safety assessments, if the unscheduled assessment occurs in a window of +/- 14 days of a planned eCRF visit for

which it was planned per protocol to have such assessment, and no assessment was done for that visit, the unscheduled assessment will be assigned to that visit. If an unscheduled assessment is equally distant from two eCRF visits with no corresponding safety assessment, the unscheduled assessment will be assigned to the next visit.

4.2.10. **Baseline Definition**

The primary endpoint and other efficacy endpoints based on daily e-diary data (DYS, NMPP, dyschezia, overall pelvic pain, dyspareunia, analgesic use, difficulty in daily activities, uterine bleeding, missing school or work, having to sleep or laying down) will base the baseline calculation on diary data from menstrual cycles of the screening period. The screening period must cover at least two full menstrual cycles. The following rules apply to select the two baseline menstrual cycles for assessment of eligibility and baseline:

- If only two full screening cycles are covered prior to Day 1: those two screening cycles should be used for both eligibility assessment and baseline calculation.
- If 3 full screening cycles are covered during the screening period, for logistical reason, as allowed per protocol: "If for logistical reasons beyond her control, the subject is unable to come to the clinic for Day 1 within the acceptable time window (1st to 7th day inclusive of the cycle) then the subject will be allowed to start Day 1 in the following cycle (again on the 1st to 7th day of the cycle). In this case, the eligibility assessments should be assessed based on the first two cycles only"; the two first screening cycles should be used for both eligibility assessment and baseline calculation
- If more than 3 full screening cycles are covered during the screening period: the last two menstrual cycles prior to the baseline visit will be considered for both eligibility assessment and baseline calculation.
- Due to possible technical issues with the [redacted] e-diary device, some data during screening menstrual cycle may not have been collected properly. Consequently the menstrual cycles may not be considered as valid for assessment of eligibility and baseline and cycles other than the first and/or second ones might be used.

ObsEva will provide an excel file listing all the menstrual cycles that were used for measurement of eligibility and baseline values.

Further details on baseline calculations are specified in the analysis sections for each endpoint.

The baseline DXA assessment will be derived as the latest non-missing DXA assessment with acceptable quality (accepted=”Yes”) prior to the baseline visit or the first non-missing assessment with acceptable quality (accepted=”Yes”) done up to 10 days after the baseline visit if there is no assessment prior to baseline.

For other data, unless otherwise specified, baseline will be defined as the data most recently collected prior to the first dose of IMP.

4.2.11. Last treatment administration date, End of Study date and Period definitions

The last treatment administration date will be defined as the date from the treatment termination of main period eCRF page. If this date is missing, the last date with a drug intake in eDiary strictly inferior to Month 6 date or the last date with a drug intake in eDiary available if the subject discontinued will be used.

For each 28-day period prior to Month X (except Month 6), the end of the period will be defined as the Month X visit date. If subject discontinued prior to Month X, the end of the period will be the date from the treatment termination of main period eCRF page or the last date with a drug intake in the eDiary if missing. If the subject missed the Month X visit (but did not discontinued), the subject might still have filled the diary during that time: the theoretical Month X date will be used to define the end of the 28-day period. The theoretical Month X is defined with respect to Day 1 as Month X= Day 1+ 28*X. Theoretical visits are presented in Table 3 of Section 4.2.9.

For the 28-day period prior to Month 6, the end of the period will be defined as the last treatment administration date.

The end of study date will be Month 6 date – 1 day for subjects entering the Extension, and the last assessment date defined as last eCRF (excluding AE and CM end dates) or eDiary date in main study or main Follow-up for subjects not entering the Extension. This will not necessarily be the same date as last assessment date as recorded in the eCRF.

Note: For summaries by visit (efficacy or safety), the Month 6 assessments will be included in the main study analyses. For Adverse Events, Concomitant Medications and Daily diaries, data collected on Month 6 date will be part of the extension for subjects entering the extension study, and be part of the Follow-up period for subjects entering the main study follow-up period.

The main treatment period will be defined as Day 1 up to (Month 6 date -1) or treatment discontinuation date.

The main follow-up period will be defined as the Month 6 date or (treatment discontinuation date+1) up to the end of study date (for subjects who entered the follow-up).

4.3. Interim Analyses

An interim analysis is not planned for this study. This SAP will however be updated twice prior to unblinding and analysis of the data up to Month 6, in order to incorporate the results of the blinded meaningful change estimation analyses at Month 3 and at Month 6.

At Month 6 Database Lock, the blinded meaningful change estimation for ranked secondary endpoints based on Month 6 data will be performed by a blinded team at Cytel and a blinded team at Clinical Outcome Solutions, while a Cytel restricted team will be unblinded to produce on the Month 6 CSR analysis. The blinded and unblinded teams will not be in contact, and any unblinded results will be provided to Sponsor only after the Month 6 meaningful change estimation has been completed and the SAP updated.

Results will be presented twice: after Month 6 completion and after Month 6 Follow-up completion.

4.4. Subject Disposition

A tabulation of subject disposition will be presented, including the number of subjects screened, randomized, who received at least one dose of study drug, the number in each subject population for analysis, the number of subjects who withdrew prior to completing the study, and reasons for withdrawal, by treatment group and overall, for randomized subjects and for safety population.

A subject will be considered to have completed study if the subject completed the treatment and follow-up periods, or completed the treatment period and entered the extension study.

The number of subjects who completed each visit will be summarized by treatment group and overall, for the FAS population.

The following listings will be presented:

- Study completion information, including the reason for premature study withdrawal, if applicable;
- Inclusion/exclusion criteria not met for Screened Subjects;
- Subject inclusion in each of the analysis sets (Randomized Set, Full Analysis Set, Per Protocol set, Threshold Analysis Set, M6 Threshold Analysis Set, PK Analysis Set, Safety Analysis Set, Follow-up Set, and Follow-up Safety Set).

4.5. Demographic and Baseline Characteristics

Baseline, demographic and medical history information will be analyzed for the (SAF), FAS and PP set. No formal statistical comparisons will be performed.

4.5.1. Demographics

Demographics and baseline characteristics will be summarized by treatment group and overall using descriptive statistics.

Demographics and baseline data include Age (years), Gender, Ethnicity, Race, Height (cm), Weight (kg), Body Mass Index (kg/m2), Systolic Blood Pressure (mmHg), Diastolic Blood Pressure (mmHg), Heart Rate (beats/minute), QTc interval (Fridericia) (ms).

BMI is auto-calculated, defined as: Weight (kg)/(Height*Height) (m^2).

Demographic and Baseline data will be provided in data listings.

4.5.2. Baseline Disease Characteristics

The definition of the two baseline menstrual cycles is described in section 4.2.10

Baseline Dysmenorrhea (DYS) and Non-Menstrual Pelvic Pain (NMPP) will be calculated by averaging over the two baseline menstrual cycles, the e-diary daily answers of Endometriosis Related Pelvic Pain (PP VRS Questionnaire). Responses of "None," "Mild," "Moderate," and "Severe" will be assigned a score of 0, 1, 2, and 3, respectively. Dysmenorrhea will be computed using days with uterine bleeding, defined as those days on which the subject records any uterine bleeding or spotting in the subject eDiary; Non-Menstrual Pelvic Pain will use days with no uterine bleeding.

The baseline analgesic use will be calculated as the mean of daily pill count of analgesics over the two baseline menstrual cycles for each class of rescue analgesics (endometriosis-associated: Ibuprofen, Narcotic Analgesic) collected in the eDiary, separately for bleeding days, non-bleeding days and overall.

Dysmenorrhea, Non-Menstrual Pelvic Pain and Analgesic use will be summarized as continuous variables

The average duration of the two baseline menstrual cycles (days) will be summarized. The duration (days) of each of the two individual eligible baseline menstrual cycle will be calculated using the (menstrual start dates provided in the excel file. The duration is computed as (start date of $Cycle_{x+1}$ – start date of $cycle_x$).

The average number of days with uterine bleeding over the two baseline menstrual cycles will be summarized as continuous variables.

The number and percentage of subjects with Normal, Abnormal, Abnormal clinically significant (or Not Assessable) results for Physical Examination, Gynecological Examination, Breast Examination, Mammography, Endometrial Biopsy and PAP Smear assessments will be presented by treatment group and overall. Abnormality details will be provided in data listings.

Transvaginal ultrasound data (i.e., presence of ovarian endometrioma with a diameter of 7 cm or greater, uterus length, width, and depth in mm and corresponding calculated uterine volume in cm3 (using the prolate ellipsoid formula [L x H x W x 0.523]), endometrium thickness in mm, and presence of any uterus, left ovary or right ovary abnormality) will be summarized.

Dysmenorrhea, Deep Dyspareunia, Non Menstrual Pelvic Pain, Total Pelvic Pain (None: 0, Mild: 1-3, Moderate: 4-6, Severe: 7-9), Pelvic Tenderness, Induration, Total Physical Sign Score (None: 0, Mild: 1-2, Moderate: 3-4, Severe: 5-6) and Composite Pelvic pain and Physical Sign Score (None: 0, Mild: 1-2, Moderate: 3-5, Severe: 6-10, Very Severe 11-15) from the Modified Biberoglu & Behrman symptom

severity scale assessment as collected in eCRF will be summarized at baseline as categorical variables with the following categories: None, Mild, Moderate, Severe (Very Severe).

Baseline Disease Characteristics will be reported in listings.

4.5.3. Medical History

Medical history will be summarized by System Organ Class and Preferred Term using the Medical Dictionary for Regulatory Activities (MedDRA), by treatment group and overall.

Medical history is summarized by subject incidence rates; therefore, a subject contributes only once to the count for a given medical history (SOC or preferred term).

Medical history will be reported in a listing.

4.5.4. Endometriosis History

Time since first medical diagnosis/treatment, first surgical diagnosis and most recent surgical diagnosis in years defined as (randomization date – date of diagnosis)/365.25 will be summarized.

Presence within 2 months before screening and ongoing symptoms of the following symptoms will be summarized, by treatment group and overall: Dyspareunia, Dyschezia, Dysuria, Adenomyosis, Rectovaginal endometriosis nodes.

Endometriosis History will be reported in a listing.

4.6. Efficacy Evaluation

Efficacy analysis will be conducted using the Full Analysis Set. Primary and ranked secondary analyses will be repeated on the Per Protocol Set.

4.6.1. Meaningful Change Threshold Estimation

The analyses to estimate the meaningful change thresholds (MCTs) using the blinded interim data will be produced by an independent programming team and statistician to the Sponsor. Analysis datasets will be provided by Cytel. The external programming team will execute the analyses and the statistician will review and interpret the results and provide recommendations to the Sponsor for threshold estimation. Analyses to determine the MCTs will be described in a separate document.

While the responder threshold analyses using these MCTs (see Section 4.6.2) will be conducted on all FAS subjects, the derivation of the MCTs for the co-primary endpoints was undertaken once patients had completed Month 3, using data up to Month 3 on the Threshold Analysis Set (FAS subjects with available anchor data). The derivation of the MCTs for the additional supportive responder analyses on the ranked secondary endpoints will be undertaken once patients have completed Month 6, using data up to Month 6 on the M6 Threshold Analysis Set (FAS subjects with available anchor data at Month 6). For the specific mPGIS anchors, all subjects enrolled after Protocol Amendment V3.0 and collecting data using updated

e-Diary (expected to be approximately 50 subjects) with available anchor data at Baseline and Month 3 (Month 6 for additional supportive responder analyses on the ranked secondary endpoints) are considered, as these specific mPGIS anchors were introduced as part of Amendment 3. For the generic mPGIS and PGIC anchors the analyses were (will be) performed on a random sample of 200 subjects with data for both anchors: mPGIS at Baseline and Month 3 (respectively Month 6 for ranked secondary endpoints) and mPGIC at Month 3 (respectively Month 6). The same random sample was (will be) used for PGIS and PGIC anchors. For the EHP-30 Pain Domain endpoint, the analyses will be repeated to confirm results using the remaining subjects (out of the selected sample of 200 subjects) from the study. As with the under treatment efficacy analysis, for subjects who discontinue treatment prior to Month 3 (prior to Month 6 for ranked secondary endpoints), the 28 calendar days immediately prior to and including the last dose date will be used, along with the closest time-matched monthly anchor measures.

SAS Code for Random Sample:

```
proc surveyselect data=dataset out=sample method=srs sampsize=200 seed=12345;

run;
```

4.6.1.1. Responder Threshold Analysis

Responder threshold analyses will be performed for the co-primary endpoints (See Section 4.6.2) and, as additional supportive analyses for the ranked secondary endpoints (see Sections 4.6.3.1 to 4.6.3.5) using the MCTs estimated from the blinded data.

The final responder thresholds will be included in the final SAP prior to breaking the blind to the sponsor and main study team after Month 6. Based on the results from the Phase 2b Edelweiss study, it is expected that the responder threshold will be in the range of a 0.7 to 1.25 point reduction in pain in the 4-point VRS for a 2-point improvement in anchor status and a 0.25 to 0.75 point reduction in pain in the 4-point VRS for a 1-point improvement in anchor status. The numerical thresholds for DYS are expected to be larger than those for NMPP. No MCTs were estimated in Phase 2 for the other Phase 3 ranked secondary endpoints (overall pelvic pain, dyschezia, dyspareunia, EHP-30 pain domain). In Phase 3, additional anchor measures are included to estimate MCTs for these endpoints.

4.6.1.1.1. Primary Efficacy Endpoints

The two co-primary, composite, efficacy endpoints are reaching the responder threshold in 1) DYS and 2) NMPP while having no accompanying increase in analgesics for EAP. These endpoints will be assessed at Month 3.

For each of the co-primary endpoints, the criterion for defining a subject as a responder over the last 28 days of randomized treatment up to the Month 3 visit will be a reduction of X (the MCT) or greater from baseline in pain. The threshold for response in the responder analysis (i.e., the value of X) will be chosen to represent a clinically meaningful reduction in pain and may not be the same for each endpoint (DYS

and NMPP). The obtained recommended values of X following meaningful change threshold estimation at Month 3 is reported in Section 4.6.1.3.

4.6.1.1.2. Ranked Secondary Endpoints

The six ranked secondary efficacy endpoints for which additional supportive responder analyses will be performed are reaching the MCT in 1) DYS, 2) NMPP, 3) dyschezia, 4) overall pelvic pain, 5) EHP-30 Pain domain, and 6) dyspareunia. These endpoints will be assessed at Month 6.

For each of the additional supportive responder analyses for the ranked secondary efficacy endpoints, the criterion for defining a subject as a responder over the last 28 days of randomized treatment up to the Month 6 visit will be a reduction of X (the MCT) or greater from baseline. The threshold for response in the responder analysis (i.e., the value of X) will be chosen using data at Month 6, to represent a clinically meaningful reduction and may not be the same for each endpoint (DYS, NMPP, dyschezia, overall pelvic pain, EHP-30 Pain domain, dyspareunia). The MCT of DYS and NMPP may be different than those obtained at the Month 3 MCT analysis. The obtained recommended values of X following meaningful change threshold estimation at Month 6 is reported in Section 4.6.1.4.

4.6.1.2. Estimation of Meaningful Change Threshold

Anchor-based methods will be used to estimate the MCT for each co-primary and ranked secondary efficacy endpoint. Anchor-based methods are where an external criterion (anchor) is used to provide an indication of the minimal change in the PRO that is of (clinical) relevance. The anchor-based methods will include estimation of mean within-group change, with 95% confidence intervals (CIs), and receiver operating characteristic (ROC) curves to identify specific cut points. The MCT determination will also consider supportive information from cumulative distribution function (CDF) and probability distribution function (PDF) curves. Lastly, shift tables will be used, where appropriate, to examine the change in the endpoint by baseline and Month 3 (respectively Month 6 for ranked secondary endpoints) anchor score.

Because linzagolix is expected to improve the symptoms of endometriosis, the MCT for each endpoint will indicate the amount of *improvement* in each endpoint which is perceived to indicate a meaningful change. The analyses will be conducted using blinded data from Edelweiss 3 that combine the data from all treatment arms including the Placebo arm and use change from Baseline to Month 3 for the co-primary endpoints and change from Baseline to Month 6 for the ranked secondary endpoints.

Anchor-based methods are the preferred methods for estimating MCTs. An anchor is a reference measure that is used to determine change. The purpose of an anchor measure is to identify change on the target clinical outcome assessment (COA) measure that represents patient (or clinician) perceived improvement or deterioration. Any potential anchor should be less complex than the measure it is being used to assess. Anchors are expected to have a correlation coefficient of $\gtrapprox 0.35$ with the target COA measure to be used in the anchor-based analysis (Revicki et al. 2008). Correlation coefficients for Baseline scores and change from Baseline scores will be assessed to inform the suitability of the proposed anchor.

Potential anchor measures

The following single-item measures were selected to be evaluated as possible anchors for the two co-primary endpoints DYS and NMPP: monthly patient global impression of severity (mPGIS) items of DYS and NMPP, together with the non-specific mPGIS and patient global impression of change (PGIC) items of endometrial symptom severity (Table 5). The following single-item measures were selected to be evaluated as possible anchors for the ranked secondary efficacy endpoints (DYS, NMPP, dyschezia, overall pelvic pain, EHP-30 Pain domain, dyspareunia): mPGIS items of DYS, NMPP, dyschezia, overall pelvic pain, difficulty in daily activities, and dyspareunia, together with the non-specific mPGIS and patient global impression of change (PGIC) items of endometriosis symptom severity (Table 5).

Table 5 Proposed anchors for MCT estimation for each endpoint

DYS	NMPP	Overall Pelvic Pain	Dyschezia	EHP-30 Pain domain	Dyspareunia
Endpoint (shaded = main study endpoints)					
Daily 0-3 VRS	**Daily 0-3 VRS**	**Daily 0-10 NRS**	**Daily 0-10 NRS**	**Monthly 0-100**	**Daily 0-3 VRS**
0 – No pain	0 – No pain	0 (No pelvic pain) to 10 (Worst pain) NRS	0 (No dyschezia pain) to 10 (Worst pain) NRS	EHP-30 Pain Domain score	0 - No pain during sexual intercourse.
1 – Mild	1 – Mild				1 - I was able to tolerate the pain during sexual intercourse.
2 – Moderate	2 – Moderate				2 - Intercourse was interrupted due to pain.
3 - Severe	3 - Severe				3 - I avoided sexual intercourse because of anticipation of pain
Bleeding days	*Non-Bl days*	*All days*	*All days*	*4-week recall*	*All days*
Monthly anchors					
a) Specific anchors					
mPGIS 0-4 VRS menstrual PP	**mPGIS 0-4 VRS non-menstrual PP**	**mPGIS 0-4 VRS Overall PP**	**mPGIS 0-4 VRS Dyschezia**	**mPGIS 0-4 VRS Difficulty activities**	**mPGIS 0-4 VRS Dyspareunia**
0 – No pain	0 – No pain	0 – No pain	0 – No pain	0 – No difficulty	0 – No symptoms
1 - Very Mild	1 - Very Mild	1 - Very Mild	1 - Very Mild	1 – Slight	1 - Tolerated discomfort during intercourse
2 – Mild	2 – Mild	2 – Mild	2 – Mild	2 – Mild	2 – Interference of usual frequency of sexual intercourse due to pain
3 – Moderate	3 – Moderate	3 – Moderate	3 – Moderate	3 – Moderate	3 – Avoids, or wishes to avoid, intercourse because of pain
4 - Severe	4 - Severe	4 - Severe	4 - Severe	4 - Severe	
b) Generic anchors					
mPGIS 0-4 VRS Endometriosis symptoms					
0 – No pain					

1 - Very Mild
2 – Mild
3 – Moderate
4 - Severe
PGIC 7-pt VRS Endometriosis Symptoms
-3 - Very much improved
-2 - Much improved
-1 - Minimally improved
0 - No change
1 - Minimally worse
2 - Much worse
3 - Very much worse

4.6.1.3. Results of Meaningful Change Threshold Analysis for the Co-Primary Endpoints

Following the Meaningful Change Threshold analysis performed on soft locked data at Month 3, the recommendation for the final MCT estimates of the co-primary endpoints is the following:

- Dysmenorrhea (VRS): X=-1.10
- Non-Menstrual Pelvic Pain (VRS): X=-0.80

The criterion for defining a subject as a responder over the last 28 days of randomized treatment up to Month 3 will be a reduction of 1.10 or greater from baseline pain for Dysmenorrhea; a reduction of 0.80 or greater from baseline pain for Non-Menstrual Pelvic Pain.

4.6.1.4. Results of Meaningful Change Threshold Analysis for the Ranked Secondary Endpoints

Following the Meaningful Change Threshold analysis performed on Month 6 database lock, the recommendation for the final MCT estimates of the ranked secondary endpoints are the following:

- Dysmenorrhea (VRS): X=-1.25
- Non-Menstrual Pelvic Pain (VRS): X=-0.85
- Dyschezia (Numeric Rating Scale - NRS): X=-1.5
- Overall pelvic pain (NRS): X=-2.7
- EHP-30 pain domain: X=-28
- Dyspareunia (VRS): X=-0.9

The criterion for defining a subject as a responder over the last 28 days of randomized treatment up to Month 6 will be a reduction of 1.25 or greater from baseline pain for Dysmenorrhea; a reduction of 0.85 or greater from baseline pain for Non-Menstrual Pelvic Pain; a reduction of 1.5 or greater from baseline pain for Dyschezia; a reduction of 2.7 or greater from baseline pain for Overall Pelvic Pain; a reduction of 0.9 or greater from baseline pain for Dyspareunia.

The criterion for defining a subject as a responder at Month 6 or discontinuation will be a reduction of 28 or greater from baseline for EHP-30 pain.

4.6.2. **Primary Efficacy Endpoint**

4.6.2.1. Primary Efficacy Analysis

4.6.2.1.1. **Primary Endpoint Specifications**

The primary efficacy analysis aims to estimate the effect of linzagolix versus placebo on DYS and NMPP over the last 28 days up to and including the Month 3 visit or, for subjects who discontinued randomized treatment prior to the Month 3 visit, over the last 28 days of randomized treatment. Use of analgesic medication for EAP will be included in the responder definitions. A responder for each of the co-primary endpoints needs to demonstrate both a reduction in pain (DYS or NMPP, depending on the endpoint) and a stable or decreased use of analgesics for EAP (assessed separately for DYS and NMPP). Any subject who does not meet both of these criteria will be defined as a non-responder.

As detailed below, in line with ICH E9 (R1) addendum: four attributes (population, endpoint, treatment and population level summary) have been specified to translate the primary objective into treatment effect that is to be estimated (estimands):

A. Population: Premenopausal women aged 18 to 49 years inclusive at screening, with surgically and, if available, histologically confirmed pelvic endometriosis and with moderate to severe EAP.

B. Endpoint: The two co-primary, composite efficacy endpoints are clinically meaningful reduction from baseline to the last 28 days on randomized treatment, preceding the Month 3 visit or discontinuation, along with a stable or decreased use of analgesics for EAP, in the mean daily assessment of 1) DYS and of 2) NMPP measured on a Verbal Rating Scale (VRS) using an electronic diary.

C. Treatment: linzagolix administered orally once daily for up to 3 months at a dose of 75 mg alone or of 200 mg in combination with ABT, versus placebo

D. Population-level summary: Odds-ratio of proportions of responders of linzagolix versus placebo from a logistic regression model for each co-primary endpoint, with treatment group as the main effect (three values) and including the baseline pain score as a covariate.

Intercurrent events: Increased use of EAP analgesia is a non-response in the endpoint definition. Treatment discontinuations are assessed up to the time of discontinuation. Randomized treatment will be considered, regardless of lack of compliance to treatment or treatment assignment errors. The clinical question of interest is based on a while-on-treatment strategy for discontinuation and composite strategy for analgesic use. Lack of compliance to treatment and treatment assignment errors are handled using a treatment policy strategy.

The “while-on-treatment strategy” was chosen for the intercurrent event of discontinuation of treatment for the co-primary endpoints based on the recommended strategies for addressing intercurrent events in the (then draft) ICH E9 (R1) guideline.

The response to treatment prior to the occurrence of the patient discontinuing treatment is of clinical interest for a symptomatic treatment of endometriosis-associated pain. Patients and prescribers are generally interested in what happens when a patient actually takes (and can take) the treatment. This corresponds most directly to the "while on treatment" strategy for treatment discontinuation. Of note, a comparison of patient disposition and discontinuation rates will also constitute an important aspect when interpreting the study results. In addition, a "while-on-treatment" strategy for efficacy aligns with the analysis of adverse events, which are also considered while on treatment. Moreover, as endometriosis-related pelvic pain is measured repeatedly each day, the variable can be readily calculated for the 28-day period up to the point of discontinuation and the value up to the time of discontinuation can be considered to account for the intercurrent event.

For each of the co-primary endpoints, the criterion for defining a subject as a responder over the last 28 days of randomized treatment up to the Month 3 visit will include a reduction of X or greater from baseline in pain, where X was determined by a meaningful change threshold estimation using appropriate anchors as described in Section 4.6.1, as well as stable or decreased use of analgesics for EAP. Results of the meaningful change threshold estimation for the primary endpoints are reported in Section 4.6.1.3.

Mean pelvic pain scores for DYS and for NMPP will be calculated by averaging over the corresponding 28 calendar days immediately prior to and including the day of the Month 3 visit. DYS will use days with uterine bleeding, defined as those days on which the subject records any uterine bleeding or spotting in the subject eDiary; NMPP will use days with no uterine bleeding. If a subject's mean score for DYS is undefined numerically because her daily eDiary reports indicate that she did not experience uterine bleeding on any days during the 28 calendar day period, then the mean score for DYS will be set equal to zero (which reflects the absence of any DYS during that reporting time period). Similarly, if a subject's diary reports indicate that she only experienced days with uterine bleeding during the period, then the mean score for NMPP will be set equal to zero. For subjects who discontinue treatment prior to Month 3, the last available observations prior to discontinuation will be used, in order to estimate the treatment effect while under treatment, that is the 28 calendar days immediately prior to and including the last dose date. The baseline mean overall pelvic pain scores for DYS and NMPP will be calculated by averaging over the two baseline menstrual cycles (as defined in section 4.2.10), which may not be exactly 28 days each. For both DYS and NMPP, responses of "None," "Mild," "Moderate," and "Severe" will be assigned a score of 0, 1, 2, and 3, respectively.

For the assessment of stable or decreased use of analgesics for EAP (assessed separately for DYS and NMPP), the analgesic use for any defined period will be based on the mean of the total pill count for each class of rescue analgesics (endometriosis-associated). The total pill count for each class of rescue analgesic is the sum of the pill count of the corresponding class of rescue analgesic, as reported in the eDiary during the time period of interest. Baseline analgesic use will be calculated as the mean of the daily pill count of analgesics over the two baseline menstrual cycles (as defined in section 4.5.2). For the purposes of determining a stable or decreased use of analgesics, the averaging will be done over the same calendar days as used for the pelvic pain scores (i.e. for DYS days with uterine bleeding, defined as

those days on which the subject records any uterine bleeding or spotting in the subject eDiary; for NMPP days with no uterine bleeding). The evaluation of stable or decreased use of analgesics will be done per the specification in 7.1 Appendix A.

4.6.2.1.2. Primary Efficacy Analysis

The primary efficacy analysis will be conducted using the FAS.

The analysis of each co-primary endpoint will be conducted using a logistic regression model, with treatment group as the main effect (three values) and including the baseline pain score as a covariate. Individual linzagolix versus placebo treatment group comparisons will be made using the same logistic regression model. Estimated odds-ratios and 97.5% intervals of proportions of responders of linzagolix treatment groups versus placebo will be presented, along with Bonferroni corrected p-values. Estimates of proportions of responders with 95% confidence intervals will be presented as well, estimated at the overall mean baseline pain score.

SAS Code (may be updated at analysis stage to fit the data):

```
PROC LOGISTIC data=dataset order=internal;
      Class treatment(ref=placebo) / param=glm ;
      Model resp(event= "Yes")= treatment base_pain /alpha=0.05;
      Oddsratio treatment ;
      Lsmeans treatment / ilink exp cl alpha=0.05 at means diffs=control('1')
    adjust=bon;
      Ods output OddsRatiosWald=or  ModelANOVA=ma ParameterEstimates=pe
       lsmeans=lsm diffs=diffs;
Run;
```

In addition, the mean pelvic pain scores for DYS and NMPP and change from baseline for the last 28 days prior to Month 3 or the last 28 days of treatment (or less) for subjects who discontinued will be summarized as continuous variables by treatment group. Cumulative distribution functions (CDF) for change from baseline at Month 3 for DYS and NMPP will be provided, with separate lines for each treatment group. The x-axis will present the different values for the DYS or NMPP change from baseline, and the y-axis will present the proportion of subjects reaching each value of change. Vertical lines of the relevant MCT thresholds obtained at the Month 3 responder threshold analysis will be displayed on the plots.

The number and percentage of subjects with:

- Reduction of X for DYS,
- Reduction of X for NMPP,

- Stable or decreased use of analgesics for EAP during bleeding days, during non-bleeding days, and overall,
- Reduction of X for DYS and stable or decreased use of analgesics for EAP (during bleeding days),
- Reduction of X for NMPP and stable or decreased use of analgesics for EAP (during non-bleeding days),

will be presented.

Subjects whose response cannot be assessed for the co-primary efficacy endpoints, e.g. due to lack of on treatment pain data (i.e. subjects who received less than 28 days of treatment), will be considered as non-responders. In addition, for outputs presenting the numeric change from baseline, such subject will be assigned a change from baseline of zero. It is expected that very few subjects will be affected by these rules.

There may be missing data during the screening period and whilst under treatment, for example, missed days of completing the eDiary. The primary analyses will use observed data only in such cases, provided that a minimum number of completed daily eDiary entries are available. This effectively assumes that the individual values such as daily eDiary assessments (observed or not) are independent of being missing. Subjects who have less than the minimum number of days available will be excluded from the analysis. This is valid under the assumption that such data is MAR, conditional on the baseline value.

The required minimum number of completed daily eDiary entries will be set to 12 days of each 28-day period (two Baseline menstrual cycles and the 28-day period prior to Month 3), i.e., 43% of non-missing assessments for the primary analysis, consistent with the use of 3 days per week as the minimum number of completed diary entries as suggested in recent research ((Griffiths & al, 2018), (Cappelleri & Bushmakin, Biopharmaceutical Report, 2020), (Cappelleri & Bushmakin, Biopharmaceutical Report, 2021)). The analysis will be repeated using subjects with at least 75% (21 days) and 35% (10 days) of completed daily eDiaries for DYS, NMPP and analgesic use.

4.6.2.2. Sensitivity Analyses

- A Cochran-Mantel-Haenszel (CMH) test will be used to test the null hypothesis of no treatment effect at Month 3 for each linzagolix group vs. placebo with regards to the proportion of subjects with a response for DYS and NMPP. Odds ratios will be estimated from the CMH test together with the associated 97.5% confidence intervals (CIs) and corresponding Bonferroni corrected p-values. The proportion per treatment arm will be displayed together with exact Clopper-Pearson 95% CIs.

PROC FREQ (SAS Institute) will be used for the statistical analysis.

Sample SAS code (may be updated at analysis stage to fit the data):

Cochran-Mantel-Haenszel (CMH) test (to be performed pairwise).

```
PROC FREQ data=dataset;
      Tables Treatment*Response /CMH RELRISK RISKDIFF alpha=0.025;
      Where treatment in ();
  ods output riskdiffcol2=diff2 cmh=cmh commonrelrisks=cmrr;Run;

data pvals_mult2 ;
      set cmh ;
      where ALTHYPOTHESIS="General Association" ;
      pval_vs_trt1=prob * 2 ;
      if pval_vs_trt1 > 1 then pval_vs_trt1 = 1;
run;
```

Note: Response is coded as 1 for responders and 0 for non-responders.

- A sensitivity analysis will be performed, including in the definition of a responder any analgesic medication also taken for non-endometriosis associated pain. Information will be based on the concomitant medications collected in the eCRF during the two baseline menstrual cycles and the same calendar days as used for the assessment of pelvic pain (DYS and NMPP) for the 28 days prior to Month 3 visit.

  Concomitant medications in the study identified as analgesics and opioids will be reviewed and confirmed by the Sponsor and be provided in an excel file prior to Database Lock. These may include any medication with one of the following ATC codes from the WHO drug dictionary version March 2016:

Table 6 List of ATC codes for analgesics and opioids

ATC code	Description
A03D	Antispasmodics in combination with analgesics
A03EA	Antispasmodics, psycholeptics and analgesics in combination

C03AH	Thiazides, combinations with psycholeptics and/or analgesics
M01	Anti-inflammatory and anti-rheumatic products
M02	Topical products for joint and muscular pain
M03	Muscle relaxants
N02	Analgesics
N07BC	Drugs used in opioid dependence
S02DA	Analgesics and anesthetics

Among those analgesic medications, all medications with Pharmacological Subgroup (3rd level of the ATC classification) equal to 'OPIOIDS' (ATC code = N02A) will be considered as opiate medications. Other analgesic medications will be considered as non-opiate medications.

The increase/stable/decrease status of analgesic use will be defined in a blinded fashion case by case after review of the blinded data, and prior to Database Lock.

Analyses described in Section 4.6.2.1.2 will be repeated using the new responder definition.

4.6.2.3. Per Protocol Analysis

The primary analysis as described in Section 4.6.2.1.2 will be repeated using the Per Protocol Set.

4.6.2.4. Subgroup analysis

The primary analysis model and descriptive statistics as described in Section 4.6.2.1.2 will be repeated for each co-primary endpoint by subgroups as specified in Section 4.2.6.

4.6.3. Ranked Secondary Efficacy Endpoints

All secondary efficacy endpoints will be summarized by descriptive statistics for each treatment group, for each time point, including summaries of change from baseline when applicable. As with the primary analysis, individual linzagolix versus placebo treatment groups comparisons will be made.

In general, between-group comparisons for continuous endpoints will be analyzed via analysis of (co)-variance. Between-group comparisons for binary endpoints will be analyzed via logistic regression. Between-group comparisons for ordinal categorical data will be analyzed using a Mantel-Haenszel test or using Koch's method when there is also a baseline covariate.

As described in section 4.2.5, the ranked secondary endpoints will follow a fixed-sequence testing strategy within the group, so as to maintain the family-wise type I error rate.

For endpoints based on daily ediary data, analyses will use observed data only in such cases, provided that a minimum number of 43% (at least 12 days of each 28-day period) completed daily eDiary entries are available. Subjects who have less than the minimum number of days available will be excluded from the analysis. This is valid under the assumption that such data is MAR, conditional on the baseline value.

4.6.3.1. Change from baseline to Month 6 in DYS (VRS) and in NMPP (VRS)

Mean pelvic pain scores for DYS and for NMPP for each subject will be calculated in the same way as for the primary endpoints. The change from baseline to Month 6 (the last 28 days of randomized treatment up to the Month 6 visit, defined as the last 28 days prior to and including the last treatment administration date as defined in Section 4.2.11, or the 28 calendar days immediately prior to and including the last dose date for subjects who discontinue treatment prior to Month 6) in DYS using the VRS, and the change from baseline to Month 6 (the last 28 days of randomized treatment up to the Month 6 visit, defined as the last 28 days prior to and including last treatment administration date as defined in Section 4.2.11 or the 28 calendar days immediately prior to and including the last dose date for subjects who discontinue treatment prior to Month 6) in NMPP using the VRS, will be analyzed via analysis of covariance with treatment group as the main effect (three values) and including the baseline pain score as a covariate. The mean pelvic pain scores for DYS and NMPP and change from baseline will also be descriptively summarized.

SAS Code (may be updated at analysis stage to fit the data): Analysis of covariance

```
proc GLM data=dataset order=internal ;
      class treatment;
      model chg = base_pain treatment;
      lsmeans treatment / pdiff = control('1') cl alpha=0.05 at means
adjust=bon;
      ods output lsmeans = lsm diff=pvals lsmeancl=lsmeancl
lsmeandiffcl=lsmeandiffcl;  * contains the adjusted means;
run;
```

The number and percentage of subjects with a reduction of X from baseline to Month 6 for DYS, with a reduction of X from baseline to Month 6 for NMPP will be presented. The reduction will be defined in a similar way as for the primary endpoint, but for the Month 6 timepoint. The value for the threshold X will be obtained from the MCT analysis based on blinded Month 6 data (X can be different for DYS and NMPP).

4.6.3.2. Change from baseline to Month 6 in dyschezia (Numeric Rating Scale - NRS)

Mean dyschezia scores for each subject will be calculated by averaging over the corresponding 28 calendar days immediately prior to and including the day of the last treatment administration as defined in Section 4.2.11 (or the 28 calendar days immediately prior to and including the last dose date for subjects who discontinue treatment prior to Month 6). The baseline mean dyschezia score will be calculated by averaging scores over the two baseline menstrual cycles. The change from baseline to Month 6 in dyschezia using the NRS will be analyzed via analysis of covariance with treatment group as the main effect (three values) and including the baseline pain score as a covariate. The mean dyschezia score and change from baseline will be descriptively summarized as well.

The number and percentage of subjects with a reduction of X from baseline to Month 6 for dyschezia will be presented. The reduction will be defined in a similar way as for the primary endpoint, but using dyschezia scores at the Month 6 timepoint. The value for the threshold X will be obtained from the MCT analysis based on blinded Month 6 data.

4.6.3.3. Change from baseline to Month 6 in overall pelvic pain (NRS)

The change from baseline to Month 6 (the last 28 days of randomized treatment up to the Month 6 visit, defined as the last 28 days prior to and including last treatment administration date as defined in Section 4.2.11 or the 28 calendar days immediately prior to and including the last dose date for subjects who discontinue treatment prior to Month 6) in overall pelvic pain using the NRS will be analyzed the same way as for dyschezia as described in Section 4.6.3.2. The baseline will be computed as the mean of the overall pelvic pain score over the two baseline menstrual cycles.

The number and percentage of subjects with a reduction of X from baseline to Month 6 for Overall Pelvic Pain will be presented. The reduction will be defined in a similar way as for the primary endpoint, but for Overall Pelvic Pain score at the Month 6 timepoint. The value for the threshold X will be obtained from the MCT analysis based on blinded Month 6 data.

4.6.3.4. Change from baseline to Month 6 in the interference of pain with the ability to perform daily activities, measured using the pain dimension of the Endometriosis Health Profile-30 (EHP-30)

The EHP-30 pain dimension consists of 11 items each addressing the effect of pain on various activities and each assessed on a 5-point scale (0=Never through to 4=Always). The EHP-30 Pain Impact Domain comprises the following items:

During the last 4 weeks, because of your endometriosis, how often have you...

1. Been unable to go to social events because of the pain?
2. Been unable to do jobs around the house because of the pain?
3. Found it difficult to stand because of the pain?
4. Found it difficult to sit because of the pain?
5. Found it difficult to walk because of the pain?
6. Found it difficult to exercise or do the leisure activities you would like to do because of the pain?
7. Lost your appetite and/or been unable to eat because of the pain?
8. Been unable to sleep properly because of the pain?
9. Had to go to bed/lie down because of the pain?
10. Been unable to do the things you want because of the pain?
11. Felt unable to cope with the pain?

This endpoint shall be scored as per the Appendix B in Section 7.2, resulting in a score on a scale from 0 (best possible health status) to 100 (worst possible health status). The score will be computed as the sum of raw scores of each item (ranging from 0 to 4) divided by the maximum possible raw score (11*4), multiplied by 100. The change from baseline to Month 6 (the last 28 days of randomized treatment up to the Month 6 visit) in ability to perform daily activities, measured using the pain dimension of the EHP-30 will be analyzed via analysis of covariance with treatment group as the main effect (three values) and including the baseline value as a covariate. EHP-30 score at baseline and Month 6 and change from baseline will also be descriptively summarized.

The number and percentage of subjects with a reduction of X from baseline to Month 6 for the EHP-30 Pain Dimension score will be presented. The value for the threshold X will be obtained from the MCT analysis based on blinded Month 6 data.

4.6.3.5. Change from baseline to Month 6 in dyspareunia (VRS)

Responses to the daily dyspareunia questions will be scored as follows. Responses on days where the subject was not sexually active for reasons other than their endometriosis will not be included in the analysis.

Response	Score
No pain during sexual intercourse.	0

I was able to tolerate the pain during sexual intercourse.	1
Intercourse was interrupted due to pain.	2
I avoided sexual intercourse because of anticipation of pain.	3

Mean dyspareunia scores for each subject will be calculated by averaging over the corresponding days during the 28 calendar days immediately prior to and including the last treatment administration date as defined in Section 4.2.11 (or the 28 calendar days immediately prior to and including the last dose date for subjects who discontinue treatment prior to Month 6). The baseline mean dyspareunia score will be calculated by averaging scores over the two baseline menstrual cycles.

Dyspareunia scores will be summarized using the FAS and for the subgroup of subjects who have a mean dyspareunia score > 1 at baseline. The statistical analysis of dyspareunia will only use the sub-group of subjects in order to estimate the effect of treatment on such subjects. It is anticipated that this will be around 60% to 70% of the randomized subjects. This is a baseline assessment and does not influence the blinded, randomized treatment group. As per section 4.2.5, the type I error is maintained due to using fixed-sequence testing strategy. The change from baseline to Month 6 in dyspareunia will be analyzed via analysis of covariance with treatment group as the main effect (three values) and including the baseline value as a covariate.

The number and percentage of subjects with a reduction of X from baseline to Month 6 for Dyspareunia will be presented. The reduction will be defined in a similar way as for the primary endpoint, but for Dyspareunia score and for the Month 6. The value for the threshold X will be obtained from the MCT analysis based on blinded Month 6 data.

4.6.3.6. No analgesic use / no opiate use for EAP during the preceding 4-week period at Month 6

The analysis of the proportion of subjects reporting no analgesic use for EAP, and the proportion of subjects reporting no opiate use for EAP during the preceding 4-week period at Month 6 (last 28 days prior to and including last treatment administration date) will be conducted using a logistic regression model, with treatment group as the main effect (three values).

In addition to analgesic use for EAP as collected in the eDiary, concomitant medications in the study will be reviewed and medications to be considered as analgesic use for endometriosis associated pain will be confirmed by the Sponsor and be provided in an excel file prior to Database Lock. They may include any

medication with the ATC codes from WHO drug dictionary version March 2016 as listed in Table 6 in Section 4.6.2.2.

In addition to analyzing the absence of analgesic use for EAP, the amount used for each 4-week period up to Month 6 will be summarized as a continuous variable, including the change from baseline. This analysis will be based on e-diary data only, as subjects did not record opioid intake in the concomitant medication page on a tablet-by-tablet basis.

As it is likely that more than one opioid will be used (opioid analgesic may be different depending on countries), the amount of opioid used will be converted to morphine milligram equivalents (MME) in order to better quantify the actual amount of opioids used by subjects.

For countries (Spain, Hungary, France, Czech Republic) where several opioids with different MME were allowed, the data collected in the e-diary is not sufficient to determine exactly which opioid was used and what the MME is. The MME will be derived as the average of possible MME of the allowed opioids in the country.
No opioids are allowed in Ukraine. Subjects from Ukraine will therefore be excluded from the summary in MME. If some opioids are nevertheless recorded for subjects from Ukraine, they will not be considered for the summary in MME, as the conversion in MME would be unknown. Those opioids records will however be considered for other analgesic use analyses not requiring conversion in MME.

The baseline amount of analgesic used will be computed by averaging the amount used during the two baseline menstrual cycles.

Conversion to MME by country is described in Table 7.

Table 7 Morphine Milligram Equivalents for Narcotic Analgesics

Country	Permitted rescue analgesic: Narcotic analgesic local equivalent	Conversion in morphine milligram equivalents
Austria	tramadol 37.5mg + paracetamol 325mg	3.75 MME
Bulgaria	tramadol 37.5mg + paracetamol 325mg	3.75 MME
Czech Republic	codeine 30mg	4.125 MME

	paracetamol 500 mg+ codeine 30 mg tramadol 37.5mg + paracetamol 325mg	
France	paracetamol 500 mg+ codeine 30 mg tramadol 37.5mg + paracetamol 325mg	4.125 MME
Hungary	paracetamol 500 mg+ codeine 30 mg codeine 30mg tramadol 37.5mg + paracetamol 325mg	4.125 MME
Poland	tramadol 37.5mg + paracetamol 325mg	3.75 MME
Romania	codeine 30mg paracetamol 500 mg+ codeine 30 mg	4.5 MME
Spain	paracetamol 500 mg+ codeine 30 mg tramadol 37.5mg + paracetamol 325mg	4.125 MME
Ukraine	No Narcotic Analgesic allowed	
US	5 mg hydrocodone + 300 mg acetaminophen	5 MME

4.6.4. **Additional Secondary Endpoints**

All additional efficacy endpoints will be summarized by descriptive statistics for each treatment group, for each time point, including summaries of change from baseline when applicable. Individual linzagolix versus placebo treatment group comparisons will be made.

In general, between group comparisons for continuous endpoints will be analyzed via analysis of (co)-variance. Between-group comparisons for binary endpoints will be analyzed via logistic regression. Between-group comparisons for ordinal categorical data will be analyzed using a Mantel-Haenszel test, or using Koch's method when there is also a baseline covariate.

As described in section 4.2.5 additional secondary endpoints will use Bonferroni correction for p-values of comparison of treatment groups at each visit. The analyses do not form part of the fixed-sequence strategy being used for the ranked secondary endpoints and will not be fully controlled for an overall type I error rate.

Values for all endpoints that are collected daily using the eDiary (from Section 4.6.4.1 to Section 4.6.4.18) will be based on the 28 calendar days immediately prior to and including the day of the corresponding visit or prior to and including the last treatment administration date for Month 6, or the 28 calendar days immediately prior to and including the last dose date for subjects who discontinue treatment prior to the visit in question (or fewer if treatment is stopped less than 4 weeks after first dose date), in order to estimate the treatment effect while under treatment. The last 28 days on treatment will be used at all subsequent scheduled visits for those discontinued subjects. Descriptive summaries for each 28-day period will be performed in two ways however:

- Once including only values of 28 days prior to Month X, i.e. using the last 28 days prior to Month X eCRF visit date for subjects who have not discontinued up to that month. If subjects discontinued prior to Month X the corresponding value will be missing from the summary at Month X;
- Once applying the while on treatment strategy, i.e. using the last 28 days prior to Month X eCRF visit date for subjects who have not discontinued up to that month, and for subjects who discontinued prior to Month X, the last 28 days prior to treatment discontinuation will be used. For all subsequent timepoints (Month X+1, etc.) the same last 28 days prior to treatment discontinuation will be used.

Corresponding baseline values will be calculated based on the two baseline menstrual cycles, which may not be exactly 28 days each.

The values for other endpoints (monthly and 3-monthly questionnaires from Section 4.6.4.19 to Section 4.6.4.26) are defined as the value at the corresponding visit or value falling into the corresponding visit window as described in section 4.2.9. For these other endpoints, descriptive summaries by visit will be

performed in two ways, once including values at the corresponding visit or value falling into the corresponding visit window as described in section 4.2.9, and once applying the while on treatment strategy, i.e., including values from the discontinued visits at each next scheduled visit.

For endpoints based on daily diary data, analyses will use observed data only in such cases, provided that a minimum number of 43% (at least 12 days of each 28-day period) completed daily eDiary entries are available.

4.6.4.1. Clinically meaningful reduction at scheduled visits other than Month 3 for DYS and NMPP

The proportion of responders for DYS and NMPP based on daily e-diary data will be reported at each visit (each 28 days period) along with odds-ratio and 97.5% confidence intervals from logistic regression as described in Section 4.6.2.1.2.

The same responder definition defined for primary endpoint in Section 4.6.2.1.1 will be used, but for other scheduled visits.

Due to the war situation in Ukraine, data from subjects in Ukraine sites may be incomplete, missing or uncleaned. The descriptive analysis of proportion of responders for DYS and NMPP at each visit will be repeated, excluding data collected after 24FEB2022 for subjects from Ukraine sites, for the Follow-up period only (Follow-up set), as the main treatment period was completed before the start of the war. Listings will contain country information to identify those subjects.

4.6.4.2. Change from baseline to each scheduled assessment in mean pelvic pain scores for DYS, NMPP and overall pelvic pain, during the previous 4-week period assessed on the NRS and VRS

The change from baseline to each scheduled assessment in mean pelvic pain scores for DYS, NMPP and overall pelvic pain will be analyzed as described for the ranked secondary endpoint in Section 4.6.3.1 (continuous summary and analysis of covariance at each visit).

4.6.4.3. Change from baseline to each scheduled assessment in the number of days with moderate to severe pelvic pain during the previous 4-week period assessed on the VRS

The actual value and change from baseline of number of days with moderate or severe pelvic pain collected in the e-diary on the VRS will be summarized for each 28-day period. The baseline value will be computed by averaging the number of days with moderate or severe pelvic pain of the two baseline menstrual cycles. The number of days will be analyzed with a negative binomial model including baseline

as a covariate. The estimated least square means treatment differences along with 97.5% CI will be provided.

SAS Code for negative binomial model with repeated measurements

```
proc genmod data=dataset order=internal ;
      class treatment;
      model nbdays =  baseline treatment / dist=negbin type3 ;

      lsmeans treatment / diff = control('1') exp at means cl alpha=0.05
adjust=bon;
estimate 'TRT Placebo' intercept 1 base %sysevalf(&meanbase.) trt01pn 1 0 0 /
exp E;
estimate 'TRT LGX 75'  intercept 1 base %sysevalf(&meanbase.) trt01pn 0 1 0  /
exp E;
estimate 'TRT LGX 200'  intercept 1 base %sysevalf(&meanbase.) trt01pn 0 0 1
/ exp E;
      ods output lsmeans=lsm diffs=lsmd Estimates=estimate   ;
run;
```

"meanbase" macro variable is the mean value of baseline over the same selection of subjects used in the model

4.6.4.4. Change from baseline to each scheduled assessment in the mean worst pelvic pain score defined as the mean of the 5 highest daily pain scores reported during the previous 4-week period assessed on the NRS

The actual value and change from baseline of the mean worst pelvic pain score, defined as the mean of the 5 highest daily pain scores reported during each 28-day period will be summarized for each 28-day period. The baseline mean worst pelvic pain score will be the average of the mean pelvic pain score of each of the two baseline menstrual cycles. An analysis of covariance will be performed for each period.

If less than 5 days of pain score data is available, the mean worst pelvic pain will be defined as the mean of the available daily pain scores.

4.6.4.5. Change from baseline to each scheduled assessment in mean daily dyspareunia scores

reported during the previous 4-week period on the dyspareunia VRS

The actual value and change from baseline in mean daily dyspareunia scores will be summarized as a continuous variable for each 28-day period. An analysis of covariance will be performed for each period.

Subjects not sexually active for reasons other than endometriosis will have missing values for the corresponding 4-week periods. Subjects not sexually active for reasons other than endometriosis at baseline will be excluded from the analysis.

4.6.4.6. Change from baseline to each scheduled assessment in mean daily dyschezia scores reported during the previous 4-week period assessed on the dyschezia NRS

The actual value and change from baseline in mean daily dyschezia scores will be summarized as a continuous variable for each 28-day period. An analysis of covariance will be performed for each period.

4.6.4.7. Change from baseline to each scheduled assessment in non-opioid, opioid and combined analgesic use for EAP during the previous 4-week period based on pill count in the eDiary

Non-opioid, opioid and combined analgesic use for EAP will be measured using the mean of daily pill count collected in the eDiary over each 28-day period. For non-opioid and combined analgesic use, descriptive summary (actual values and changes from baseline) and analysis of covariance for each period will be performed. Due to a low number of subjects using any opioids, opioids use will not be analyzed as a continuous variable, only a listing with by subject opioid use mean daily pill count will be provided.

The number and percentage of subjects reporting no analgesic use for EAP, and the proportion of subjects reporting no opiate use for EAP at baseline and during the preceding 4-week period at Month 3 (last 28 days prior to and including last treatment administration date) will be presented. The proportion of subjects with no analgesic use for EAP, and with no opiate use for EAP will be conducted using a logistic regression model, with treatment group as the main effect. For these proportions, in addition to eDiary analgesic use, data from the concomitant medication page will be used, as described in Section 4.6.3.6.

4.6.4.8. Change from baseline to each scheduled assessment in opioid analgesic use for EAP as reported in the eDiary during the previous 4-week period based on morphine milligram

equivalent (MME)

Opioid analgesic use for EAP will also be assessed based on morphine milligram equivalent (MME). The mean daily amount for each 28-day period reported in eDiary as pill count will be converted in MME.

No opioids are allowed in Ukraine. Subjects from Ukraine will therefore be excluded from the analysis in MME. If some opioids are nevertheless recorded for subjects from Ukraine, they will not be considered, as the conversion in MME would be unknown. Those opioids records will however be considered for other analgesic use analyses not requiring conversion in MME.

Due to a low number of subjects using any opioids, opioids use will not be analyzed, only a listing with by subject mean daily opioid use in MME will be provided.

4.6.4.9. Change from baseline to each scheduled assessment in the number of days of analgesic use (including any class) for EAP during the previous 4-week period as assessed in the eDiary

The actual value and change from baseline in the number of days with analgesic use for EAP collected in the e-diary will be summarized for each 28-day period. The number of days will be analyzed using a negative binomial model, including baseline as a covariate The estimated least square means treatment differences along with 97.5% CI will be provided. The baseline value will be computed by averaging the number of days with analgesic use of the two baseline menstrual cycles.

4.6.4.10. Change from baseline to each scheduled assessment in the number of days of opioid analgesic use for EAP during the previous 4-week period as assessed in the eDiary

The actual value and change from baseline in the number of days with opioid analgesic use for EAP collected in the e-diary will be calculated for each 28-day period. The baseline value will be computed by averaging the number of days with opioid analgesic use of the two baseline menstrual cycles.

Due to a low number of subjects using any opioids, number of days with opioid analgesic use will not be analyzed, only a by subject listing will be provided.

4.6.4.11. Change from baseline to each scheduled assessment in the number of pelvic pain-free days

(assessed on the VRS) during the previous 4-week period

The actual value and change from baseline of number of pelvic pain-free days collected in the e-diary will be summarized for each 28-day period. The number of days will be analyzed with a negative binomial model, including baseline as a covariate. The estimated least square means treatment differences along with 97.5% CI will be provided. The baseline value will be computed by averaging the number of pelvic pain-free days of the two baseline menstrual cycles.

4.6.4.12. Change from baseline to each scheduled assessment in ability to perform daily activities during the previous 4-week period, as assessed in the eDiary (daily function NRS)

Daily activities score for each subject will be calculated by averaging over the corresponding 28 calendar days immediately prior to and including the day each visit (or the 28 calendar days immediately prior to and including the last dose date for subjects who discontinue treatment prior to Month 6). The baseline mean of daily activities score will be calculated by averaging over the two baseline menstrual cycles. The actual value and change from baseline of daily activities score will be summarized as a continuous variable and change from baseline for each 28-day period will be evaluated using analysis of covariance.

4.6.4.13. Change from baseline to each scheduled assessment in the number of days with no difficulty in doing daily activities due to EAP during the previous 4-week period as assessed in the eDiary (daily function NRS)

The actual value and change from baseline of number of days with no difficulty in doing daily activities due to EAP collected in the e-diary will be summarized for each 28-day period. The number of days will be analyzed with a negative binomial model, including baseline as a covariate. The estimated least square means treatment differences along with 97.5% CI will be provided. The baseline value will be computed by averaging the number of days of the two baseline menstrual cycles.

4.6.4.14. Change from baseline to each scheduled assessment in the number of days when dyspareunia was a problem during the previous 4-week period (including days when sexual intercourse was avoided because of anticipation of pain) as assessed on the dyspareunia VRS

The actual value and change from baseline of number of days with answer to Sexual Intercourse in the last 24H questionnaire is “Intercourse was interrupted due to pain” or “No, I avoided sexual intercourse because of anticipation of pain” collected in the e-diary will be summarized for each 28-day period. The number of days will be analyzed with a negative binomial model, including baseline as a covariate. The

estimated least square means treatment differences along with 97.5% CI will be provided. The baseline value will be computed by averaging the number of days of the two baseline menstrual cycles.

4.6.4.15. Change from baseline to each scheduled assessment in the number of days when sexual intercourse was avoided because of anticipation of pain during the previous 4-week period as assessed on the dyspareunia VRS

The actual value and change from baseline in the number of days with answer to Sexual Intercourse in the last 24H questionnaire is "No, I avoided sexual intercourse because of anticipation of pain" collected in the e-diary will be summarized for each 28-day period. The number of days will be analyzed with a negative binomial model, including baseline as a covariate. The estimated least square means treatment differences along with 97.5% CI will be provided. The baseline value will be computed by averaging the number of days of the two baseline menstrual cycles.

In addition, the proportion of days when sexual intercourse was avoided because of anticipation of pain will be also summarized for actual values and changes from baseline.

The proportion is defined as:

> Number of days when sexual intercourse was avoided because of anticipation of pain/ (Number of days in period – Number of days with missing data – Number of days with no sexual intercourse for other reasons)

The number of days in each period will be equal to 28, except at baseline, for which it will be the number of days in each baseline menstrual cycle.

4.6.4.16. Change from baseline to each scheduled assessment in the number of days with uterine bleeding (including spotting) during the previous 4-week period measured by eDiary

The actual value and change from baseline in the number of days with uterine bleeding ("Bleeding", "Spotting", "Heavy Bleeding") collected in the e-diary will be summarized for each 28-day period. The number of days will be analyzed with a negative binomial model, including baseline as a covariate. The estimated least square means treatment differences along with 97.5% CI will be provided. The baseline value will be computed by averaging the number of days with uterine bleeding during the two baseline menstrual cycles.

4.6.4.17. Change from baseline to each scheduled assessment in the number of days when school or work was missed due to EAP in the previous 4-week period as reported in the eDiary

The actual value and change from baseline in the number of days when school or work was missed due to EAP as collected in the e-diary will be summarized for each 28-day period. The number of days will be analyzed with a negative binomial model, including baseline as a covariate. The estimated least square means treatment differences along with 97.5% CI will be provided. The baseline value will be computed by averaging the number of days during the two baseline menstrual cycles when school or work was missed due to EAP.

4.6.4.18. Change from baseline to each scheduled assessment in the number of days when the subject had to go to bed or lie down due to EAP in the previous 4-week period as reported in the eDiary

The actual value and change from baseline in the number of days when the subject had to go to bed or lie down due to EAP as collected in the e-diary will be summarized for each 28-day period. The number of days will be analyzed using a negative binomial model, including baseline value as a covariate. The estimated least square means treatment differences along with 97.5% CI will be provided. The baseline value will be computed by averaging the number of days during the two baseline menstrual cycles when the subject had to go to bed or lie down due to EAP.

4.6.4.19. Change from baseline to each scheduled assessment in the Pain, Control and powerlessness, Emotional well-being, Social support, Self-image dimensions and the Modular sexual relationship questionnaire of EHP-30 scores

The EHP-30 is a disease-specific self-administered questionnaire used to measure health-related quality of life in women with endometriosis. The EHP-30 is composed of two parts: a core questionnaire containing five scales that are applicable to all women with endometriosis and a modular part containing six scales which do not necessarily apply to all women with endometriosis. This study will employ the Core EHP-30 and modular section C (Part 2) as outlined in Appendix D of protocol.

The questionnaire is composed of five scales and of the sexual relationship scale in the modular part:

- Pain (Questions 1-11): See Section 4.6.3.4 for items included in Pain dimension
- Control and powerlessness (Questions 12-17):
    - Generally felt unwell?
    - Felt frustrated because your symptoms are not getting better?
    - Felt frustrated because you are not able to control your symptoms?

  - Felt unable to forget your symptoms?
  - Felt as though your symptoms are ruling your life?
  - Felt your symptoms are taking away your life?
- Social support (Questions 24-27):
  - Felt unable to tell others how you feel?
  - Felt others do not understand what you are going through?
  - Felt as though others think you are whining?
  - Felt alone?
- Emotional well-being (Questions 18-23):
  - Felt depressed?
  - Felt weepy/tearful?
  - Felt miserable?
  - Had mood swings?
  - Felt bad-tempered or short-tempered?
  - Felt violent or aggressive?
- Self-image (Questions 28-30):
  - Felt frustrated that you cannot always wear the clothes you would choose?
  - Felt your appearance has been affected?
  - Lacked confidence?
- Modular sexual relationship questionnaire

Refer to Appendix B in Section 7.2 for scoring. Each scale is standardised on a scale of 0 – 100, where 0 indicates the best health status and 100 the worst health status. Scale scores for each scale are calculated from the total of the raw scores of each item in the scale divided by the maximum possible raw score of all the items in the scale, multiplied by 100.

Each item of each scale will be summarized by category at each visit (Day 1, Month 1, Month 3, Month 6). The actual value and change from baseline of each scale score will be summarized as a continuous variable.

Each scale score will be analyzed at each visit using an analysis of covariance.

4.6.4.20. Change from baseline to each scheduled assessment in the Health-Related Productivity

Questionnaire (HRPQ) scores

The Health-Related Productivity Questionnaire (HRPQ) is a validated, 9-item, self-reported questionnaire that has been used to evaluate health-related productivity.

The HRPQ includes items on employment status; effect of health issues or treatment on working scheduled hours (absenteeism); effect of health issues or their treatment on work output (presenteeism); effect of health issues or their treatment on planned hours of household chores; effect of health issues or their treatment impact work output for household chores actually performed; how long since health issues developed; and effect of health issues on life.

The questionnaire follows a skip pattern so that subjects can answer only the items applicable to them according to whether they work outside the home (eg, full- or part-time employment). Two subject groups are defined: employed and household (employed and not employed, combined).

See Appendix C in Section 7.3 for HRPQ Scoring.

At Day 1, the following will be summarized:

- Employment status,
- Time since symptoms developed,
- Number of subjects and time for which Endometriosis symptoms or its treatment(s) forced to work part-time when wanted to work full-time,
- Number of subjects and time for which Endometriosis symptoms or its treatment(s) kept from having a job when wanted to work full-time,
- Number of subjects and time for which Endometriosis symptoms or its treatment(s) kept from having a job when wanted to work part-time.

The following scores (actual values and changes from baseline) will be summarized at each visit (Day 1, Month 1, Month 2, Month 3, Month 4, Month 5, Month 6):

- For “Employed” group:
    - Number of hours of work lost due to absenteeism (“How many hours of work did you miss because of illnesses/treatments the last week?”)
    - Number of hours of work lost due to presenteeism (“For the hours that you did work during the past week, how did illnesses/treatments impact your work output?”)
    - Total number of hours of work lost
    - Percent of scheduled work lost due to absenteeism (Number of hours lost/Number of scheduled hours*100)

- Percent of scheduled work lost due to presenteeism (Number of hours lost/Number of scheduled hours*100)
- Percent of scheduled work lost in total (Number of hours lost/Number of scheduled hours*100)

- For "Household" group:
  - Number of hours household work lost due to absenteeism ("How many hours of work did you miss because of illnesses/treatments the last week?")
  - Number of hours household work lost due to presenteeism ("For the hours of household chores that you did during the past week, how did illnesses/treatments impact your work output?")
  - Total number of hours of household work lost
  - Percent of scheduled household work lost due to absenteeism (Number of hours lost/Number of scheduled hours*100)
  - Percent of scheduled household work lost due to presenteeism (Number of hours lost/Number of scheduled hours*100)
  - Percent of scheduled household work lost in total (Number of hours lost/Number of scheduled hours*100)

The change from baseline in the percentage of scheduled work lost due to absenteeism/presenteeism/total for employed and household groups will be analyzed with an analysis of covariance.

4.6.4.21. Number of non-study endometriosis related health visits, number of days in hospital and type of procedures performed based on Health Resource Utilization Questionnaire (HRUQ) at each scheduled assessment

The number of non-study endometriosis related health visits, number of subjects by type of clinician/clinic visit, number of subjects by type of therapeutic procedures, and number of nights spent in hospital will be summarized at each visit (Day 1, Month 1, Month 2, Month 3, Month 4, Month 5, Month 6). In case of many clinician/clinic types and therapeutic procedures, types may be regrouped for the descriptive summaries.

The change from baseline of number of non-study endometriosis related health visits and number of nights in hospital will be analyzed with a negative binomial model, including baseline as a covariate. The estimated least square means treatment differences along with 97.5% CI will be provided.

4.6.4.22. Change from baseline to each scheduled assessment in the Physician/Subject Surgery Intention Question (PSIQ/SSIQ)

The PSIQ/SSIQ answers will be summarized as continuous at Day 1 and Month 6. Comparison between treatment arms at Month 6 will be done using analysis of covariance.

4.6.4.23. Change from baseline to each scheduled assessment in the PROMIS Fatigue – Short Form 6a

The PROMIS Fatigue Short Form 6a is composed of 6 items on 5-point scales (1-5). The total score is measured as the sum of answers of each item and will range from 6 to 30. The higher the total score, the more severe the symptom. Missing data will not be imputed.

Each item will be summarized by category at each visit (Day 1, Month 1, Month 3, Month 6). The total score will be summarized as a continuous variable.

An analysis of covariance will be performed on the total score.

4.6.4.24. Change from baseline to each scheduled assessment in the EuroQoL 5-Dimension 5-Level (EQ-5D-5L) questionnaire

The EuroQoL 5-dimension questionnaire (EQ-5D) is a self-administered, generic utility instrument developed by the EuroQoL Group in 1990. The EQ-5D consists of 2 parts, a descriptive system of 5 dimensions and a 20-cm vertical visual analogue scale (VAS) with endpoints 0 and 100. It is collected by the subject in the eDiary at Day 1, Month 1, Month 3, Month 6.

The descriptive system includes five single-item dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Subjects must choose between five levels of difficulty (from 1 to 5) in accomplishing tasks in each dimension. The health states will be described in terms of five-digit numbers based on the answers to each of the five questions. The states will then be converted to a single index score using the TTO (Time Trade-Off) method. The TTO method has been chosen since this is the only one available for the United States (US) and also because for economic studies where cost utility analysis is anticipated, most health economists are recommending a value set that has been derived using a choice-based method. An index score based on the individual dimensions will be calculated using a scoring algorithm (see Reference document EQ-5D-5L_UserGuide_2015.pdf) and the US Index from Crosswalk value set downloaded from the EuroQol website.

In the present study, the available US-specific scoring will be used (see Appendix D in Section 7.4. how to compute EQ-5D-5L crosswalk index values with SAS using the United States value set as an example).

Answers to each dimension will be summarized by category at each visit. The index value and VAS Score will be summarized as continuous variables (actual values and changes from baseline). Index value and VAS Score will be analyzed via an analysis of covariance

4.6.4.25. Response at each scheduled assessment according to Patient Global Impression of Change (PGIC) (and Post-treatment Patient Global Impression of Change, PPGIC)

Responses to the Patient Global Impression of Change questionnaire range from 0- Very much worse to 6-Very much improved. Responses will be summarized by category at each visit (Month 1, Month 2, Month 3, Month 4, Month 5, Month 6). A generalized linear model analysis for ordinal data will be performed. Odds-ratio and 97.5% confidence intervals estimates of having higher ordered responses categories, of each treatment group versus Placebo will be presented.

SAS Code (may be updated at analysis stage to fit the data): Generalized Linear Model for Ordinal Data

```
Proc genmod data=dataset order=internal descending ;
Freq count;
Class treatment ;
Model pgic=treatment / dist=multinomial link=clogit;
Estimate 'logor 75 vs pbo' -1 1 0 / exp alpha=0.025 ;
Estimate 'logor 200 vs pbo' -1 0 1 / exp alpha=0.025 ;
Ods output estimates=estimates ;
Run;
```

4.6.4.26. Change from baseline to each scheduled assessment in the monthly PGIS (mPGIS) score

Responses to the Patient Global Impression of Severity range from 0- No symptoms to 4-Severe. Responses will be summarized by category at each visit (Day 1, Month 1, Month 2, Month 3, Month 4, Month 5, Month 6). A generalized linear model analysis for ordinal data will be performed. Odds-ratio and 97.5% confidence intervals estimates of having higher ordered responses categories, of each treatment group versus Placebo will be presented.

4.6.5. Additional Efficacy Analyses

4.6.5.1. Clinically meaningful reduction at Month 3 for the two co-primary endpoints with

discontinued subjects considered as non-responders

A supportive analysis of the co-primary endpoint analysis will be conducted that considers all subjects who prematurely discontinue the study treatment before the Month 3 visit as non-responders. These are composite endpoints that estimate the proportion of subjects with clinically meaningful reduction in pain, a stable or decreased use of rescue medication, and who complete 3 months of treatment. The same analyses as described in Section 4.6.2.1.2 will be performed, i.e., using a composite strategy for discontinuation as opposed to a while on treatment strategy used for the primary estimand.

The analysis will be performed on the Full Analysis Set and the Per Protocol Set.

4.6.5.2. Clinically meaningful reduction at Month 3 for the two co-primary endpoints with multiple imputation approach for subjects who discontinued the study early

As supportive analysis of the co-primary endpoint analysis, in order to estimate the effect of treatment policy, a reference based multiple imputation approach will be used for the co-primary endpoints for subjects who discontinue the study early, under the assumption that the efficacy of the linzagolix treated subjects gradually transitions to that observed in the placebo subjects, i.e., using a treatment policy strategy for discontinuation as opposed to a while on treatment strategy used for the primary estimand.

Details on multiple imputation methodology are described in section 4.2.8.

DYS and NMPP mean scores changes, and ibuprofen and analgesic use (pill counts) will be simultaneously imputed using multivariate imputation by a fully conditional specification method, with 20 burn-in iterations before each imputation. The predictive mean matching method for continuous variables will be used where the imputed values will be randomly taken from the 5 closest observed values whose predicted values are closest to the predicted value for the missing value from the simulated imputation model. The imputation model will include baseline value as a covariate. Following imputation, the corresponding binary endpoints of response for DYS, NMPP including use of analgesics, will be computed.

For each imputed dataset, the logistic regression analyses for DYS and NMPP as described in section 4.6.2.1.2 will be performed. The combined estimate of odds ratios will be obtained from each of the m models estimates.

SAS Code for multiple imputation:

```
Step 1: Multiple imputation
     proc mi data=dataset seed=1910 nimpute=500 out=outmi;
          class treatment ;
         fcs regpmm ( DYS_chg = DYS_base / k=5 ) nbiter=20;
         fcs regpmm ( ibu = ibu_base / k=5 ) nbiter=20;
         fcs regpmm ( narco = narco_base / k=5 ) nbiter=20;
          mnar model( DYS_chg / modelobs=(treatment='placebo') );
         mnar model( ibu / modelobs=(treatment='placebo') );
         mnar model( narco / modelobs=(treatment='placebo') );
```

```
          var DYS_base ibu_base nar_base DYS_chg ibu narco;
        run;

      data outmi2 ;
          set outmi ;
          /* Derive binary endpoints as described in primary endpoint section*/
      Run;

  Step 2: Statistical analysis for treatment effect: Logistic Regression
      PROC LOGISTIC data=outmi2 order=internal;
             By _imputation_ ;
             Class treatment(ref=placebo);
             Model DYS_resp(event= "Yes")= treatment DYS_base  /alpha=0.05;
             Oddsratio treatment ;
             Lsmeans treatment / ilink exp cl alpha=0.05 at means diffs=control("1")
             adjust=bon;
            Ods output OddsRatiosWald=or
            lsmeans=lsm diffs=diffs ParameterEstimates=es;

      run;

  Step 3: Combine estimates from each imputation dataset
  Data or2 ;
      Set diffs ;
      Logodds=estimate;

      Logoddsse=(AdjUpper-AdjLower)/(2*2.24);
  Run;
  proc mianalyze data=or2;
    ods output parameterestimates=mian_logodds;
    modeleffects logodds;
    stderr logoddsse;
  run;

* Back-transform to get odds ratio ;
  data mian_odds; set mian_logodds;
    estimate_back = exp(estimate); *pooled odds ratio;
    lcl_back=estimate_back*exp(-2.24*stderr); *pooled lower limit;
    ucl_back=estimate_back*exp(+2.24*stderr); *pooled upper limit;
  run;
* pooled p-value ;
      %macro trtby(n=);
      data es_;
        set es;
        where variable in ('treatment') and CLASSVAL0="&n";
      run;
      proc sort data=es_ out=lgs_anova1_es;
        by variable _imputation_ ;
      run;
      ** calculate Rx;
      data lgs_anova2_es;
```

```
        set lgs_anova1_es;
        by variable _imputation_ ;
        StChi=SQRT(WaldChiSq);
      run;
      proc means data=lgs_anova2_es noprint;
        by variable ;
        var StChi;
        output out=lgs_anova3_es(drop=_:) mean=mnStChi;
      run;
      data lgs_anova4_es;
        merge lgs_anova2_es lgs_anova3_es;
        by variable ;
        StDif2=(StChi-mnStChi)**2;
      run;
      proc means data=lgs_anova4_es noprint;
        by variable ;
        var df WaldChiSq StDif2;
        output out=lgs_anova5_es(drop=_t: rename=(_freq_=M)) mean=DF mnChi2 mnStDif2
sum=smDf smChi2 smStDif2;
      run;
      data lgs_anova6_es_&n;
        set lgs_anova5_es;
        by variable ;
        Rx=(1+1/M)/(M-1)*smStDif2;
        Dx=(mnChi2/DF-(M+1)/(M-1)*Rx)/(1+Rx);
        DF_d=DF**(-3/M)*(M-1)*(1+1/Rx)**2;
        Pval=1-CDF("F",Dx,DF,DF_d);
        trtn=&n;

      run;
%mend;
%trtby(n=2);
%trtby(n=3);
/* Bonferroni adjustment */
      data bonf_2 ;
            set LGS_ANOVA6_ES_2 ;
            pvalbon=pval * 2 ;
            if pvalbon > 1 then pvalbon = 1;

      run;
```

SAS Code may be updated to fit the data.

The analysis will be performed on the Full Analysis Set.

4.6.5.3. Clinically meaningful reduction at Month 3 for the two co-primary endpoints with stable or decreased use of analgesics for EAP assessed across all days

An additional analysis of DYS and of NMPP will be conducted where the assessment of stable or decreased use of analgesics for EAP will be assessed over all days rather than separately for bleeding days

and non-bleeding days. This same assessment will then be incorporated into the DYS and NMPP endpoints. The same logistic regression analysis as described at the start of Section 4.6.2.1.2 will be performed.

The analysis will be performed on the Full Analysis Set.

4.6.5.4. COVID-19 Considerations

ObsEva in consultation with its Independent Data safety Monitoring Committee (IDMC) evaluated the overall benefit/risk of continuing to conduct this study during the COVID-19 pandemic and concluded that the integrity of the trial, the safety and wellbeing of trial participants and the safety of clinical trial staff were adequately assured by the measures put in place, as described below.

To ensure continuity in patient's safety monitoring, treatment, and study data collection, remote study visits, use of local laboratory and shipment of materials and study drugs from site to subject were put in place, whenever feasible, for subjects who cannot attend their visits due to quarantine, travel restrictions, site closure or other unforeseen reasons. Sites were to schedule telephone calls to subjects as replacement for on-site visits to identify adverse events and ensure continuous medical care and oversight. Phone scripts and detailed instructions for the study conduct during COVID-19 pandemic were distributed to all sites.

Any COVID-19 related study protocol deviations were being documented and regularly reviewed and analyzed by ObsEva in order to take appropriate corrective and/or preventive actions as needed.

In addition, remote monitoring visits were being approved by ObsEva when the monitors were not able to conduct in-person visits to the study sites to perform their oversight activities. The clinical monitoring plan was updated accordingly to define the content of remote monitoring visits.

ObsEva believes that COVID-19 related protocol changes do not impact the clinical data interpretability considering that:

- co-primary efficacy endpoints and most of the study secondary efficacy endpoints are patient-driven data collected by the patient on her e-diary and not during an on-site visit,
- very few protocol deviations related to COVID-19 have been reported,
- although COVID-19 could potentially affect pain values, very few participating subjects were affected with the condition, thus the impact on pain values is expected to be negligible

4.6.5.5. Daily Patient Global Impression of Severity (dPGIS), Daily Event Cancelled, Daily Difficulty

Sleeping

The dPGIS, Event Cancelled and Difficulty Sleeping daily diaries were collected for subjects enrolled prior to amended Protocol V4.0 only.

The dPGIS mean scores and changes from baseline, the number of days with difficulty sleeping, the number of days for which events were cancelled and the number of days with no event planned will be computed for each 4-week period and reported in listings only.

The dPGIS scores will be calculated by averaging over the corresponding 28 calendar days immediately prior to and including the day of each scheduled assessment (or prior to and including the last dose date for subjects who discontinue treatment).

4.7. Pharmacokinetic Evaluations

Pharmacokinetic analyses will be conducted using the Pharmacokinetic Set.

PK samples are collected at each monthly visit.

For descriptive statistics of plasma concentrations mean (arithmetic and geometric), standard deviation, median, 1st and 3rd quartiles, minimum, maximum, coefficient of variation (CV%) and number of observations will be provided. Concentrations below the limit of quantification (LOQ) will be assigned a value of zero. Missing values will not be imputed, and if sufficient data are missing for a given subject, that subject may be considered non-evaluable for pharmacokinetic analysis and would not be included in the PK Set.

All plasma concentration data will be displayed in listings. PK assessment day and time, IMP intake status and IMP intake time on the day of PK assessment and on the previous days will be included in the listings.

Explorative analyses of correlations between plasma concentrations and intrinsic PK factors such as body weight/BMI, race, age will be performed, as appropriate, and will be reported separately.

4.8. Safety Analyses

Safety analyses will be conducted using the Safety Population.

4.8.1. Extent of treatment exposure and compliance

4.8.1.1. Extent of treatment exposure

Duration of treatment (in weeks and in days) will be summarized for each treatment group. Duration is defined as: [(date of last treatment administration as described in section 4.2.11) - (date of first administration as collected in eCRF) +1] / 7 (weeks), [(date of last treatment administration as described in section 4.2.11) - (date of first administration as collected in eCRF) +1] (days).

Time on study (in weeks and in days) will be summarized for each treatment group, defined as: [(end of study date as described in section 4.2.11) – (first dose date) +1.]/7 (weeks), [(end of study date as described in section 4.2.11) – (first dose date) +1.] (days).

Exposure data will be reported in listings.

4.8.1.2. Compliance

Lingazolix/Placebo Grey blister (200 mg or matching placebo), Lingazolix/Placebo Pink blister (75 mg or matching placebo) and Add-back therapy compliance from the study treatment accountability page will be averaged among the Day 1-Month 3 and Month 3-End of study periods and summarized by treatment group.

If Lingazolix/Placebo and Add-back therapy compliance from accountability data are missing, compliance will be computed from daily e-diary data for the corresponding 3-month period. Compliance will be computed as the number of days with pink tablet/grey tablet/red capsule taken divided by the number of days in the period (Month 3 visit date/early discontinuation visit date-Day 1 or Month 6 visit date/early discontinuation visit date-Month 3 visit date). Data from “Today” will be used primarily. If data from today is missing, data from “Yesterday” will be considered. If the e-diary is not completed for a day (neither “Today” nor “Yesterday), it will be assumed that no intake of drug was taken on that day.

Study treatment overdose and misuse will be reported in a listing.

Compliance will be reported in a listing.

4.8.1.3. Question Regarding Treatment Received

Subjects were asked at treatment termination which treatment they believed they received during the blinded treatment period. The number of subjects who believed they received active treatment, placebo or unknown will be summarized by treatment group on the Safety Population.

4.8.2. Adverse Events

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) and displayed in tables and listings using System/Organ/Class (SOC) and Preferred Term.

Analyses of adverse events will be performed for those events that are considered treatment emergent (TEAEs), where treatment emergent is defined as AEs with a start date on or after the first dose of study drug through 30 days after discontinuation of study drug or the Month 6 date, or any event that was present at baseline but worsened in intensity or was subsequently considered drug-related by the investigator through 30 days after discontinuation of study drug or the Month 6 date.

Adverse Event tables will be presented for the following periods, as defined in Section 4.2.11:

- Day 1 to Month 6 date -1 or treatment discontinuation date (treatment period)
- Month 6 date or (treatment discontinuation date+1) to Month 6 FU (follow-up period):
    - TEAE: AEs starting within 30 days after discontinuation of study drug or Month 6 visit date
    - Post-treatment AE: Adverse events starting more than 30 days after end of treatment
- Day 1 to Month 6 FU

If, for a subject not entering the follow-up period or the extension, an AE occurs within the 30 days following Month 6 date-1 or treatment discontinuation, then the AE will be counted in the treatment period.

For each of the treatment and follow-up periods, an overall summary table will be prepared presenting, by treatment group and overall, the number and percentage of subjects with

- any TEAE,
- Severe TEAE
- any TEAE assessed by the Investigator as related to Linzagolix,
- any TEAE assessed by the Investigator as related to add-back therapy,
- any TEAE leading to permanent discontinuation of IMP,
- any serious adverse event,
- any serious Treatment Emergent Adverse Event,
- any serious Treatment Emergent adverse event related to Linzagolix,
- any serious Treatment Emergent adverse event related to add-back therapy,
- Any TEAE leading to permanent discontinuation of IMP
- any fatal TEAE (where outcome is “Fatal”)

In these tabulations, each subject will contribute only once (i.e., via the most related occurrence or the most intense occurrence) to each of the incidence rates in the descriptive analysis, regardless of the number of episodes experienced.

Adverse events are summarized by subject incidence rates, therefore, in any tabulation, a subject contributes only once to the count for a given adverse event (SOC or preferred term).

The number and percentage of subjects with any treatment-emergent adverse event, with any treatment-emergent adverse events assessed by the Investigator as related to treatment (definite, probable, or possible relationship), and with any treatment-emergent serious adverse event and with any treatment-emergent non-serious adverse event will be summarized by treatment group and overall. Tabulation by severity will also be made. In these tabulations, each subject will contribute only once (i.e., the most related occurrence or the most intense occurrence) to each of the incidence rates in the descriptive analysis, regardless of the number of episodes.

The ongoing status of each Adverse Event will be derived based on Adverse Event Outcome as the following:

AE Outcome	AE Ongoing Status
Fatal	No
Not recovered/not resolved	Yes
Recovering/resolving	Yes
Recovered/resolved	No
Recovered/resolved with sequelae	No
Unknown	Unknown
Missing	Missing

The ongoing status collected in eCRF will be ignored.

No formal hypothesis-testing analysis of adverse events incidence rates will be performed.

All adverse events occurring on study will be listed in subject data listings.

By-subject listings also will be provided for the following: subject deaths; serious adverse events; and adverse events leading to withdrawal.

4.8.3. Bone Mineral Density (BMD) by DXA

BMD of the femoral neck, total hip and lumbar spine will be assessed by DXA at Baseline, Month 6 and Month 6 follow-up. Unacceptable DXA assessments (accepted="No") will not be summarized (hip, and spine and femur considered separately).

In case of repeated values due to BMD loss ≥5%, the initial scan will be used in summaries. Initial and repeated scan values will be reported in the listings.

Calibrated values should be provided once all cross-calibration data are made available to the vendor in charge of the scan assessment. These calibrated values will be used in the summary tables if available, otherwise initial values will be considered.

BMD values and corresponding Z-scores will be summarized for each treatment group. The within-group percent changes from baseline for BMD values will also be summarized, including 2-sided 95% confidence intervals.

Percent change from baseline to Month 6 and Month 6 follow-up will be described and will also be summarized in categories of percent change with the following classes:

- No change or increase,
- Decrease ≤ 3%,
- Decrease > 3% and ≤ 5%,
- Decrease > 5% and ≤ 7%,
- Decrease > 7% and ≤ 8%,
- Decrease > 8%.

The percent changes from baseline to Month 6 will be compared between each of the linzagolix groups and the placebo group via analysis of covariance with treatment group as the main effect (three values) and including the baseline as a covariate.

Sample SAS code (SAS code will be fully validated at the analysis stage) for analysis of covariance:

```
Proc GLM data=dataset order=internal;
        Class Treat;
        Model BMD_pchg = BaselineBMD Treat;
        LSMEANS Treat / at means pdiff= control('1') alpha=0.05  cl adjust=bon;
        ODS      OUTPUT      lsmeans=lsm      diff=pvals      lsmeancl=lsmeancl
lsmeandiffcl=lsmeandiffcl;
Run;
```

All data will be listed.

A listing of BMD for subjects who discontinued treatment will be provided.

Subgroup analyses by race, age groups 1 and 2, weight and BMI groups 1 and 2 will be performed for BMD, including descriptive summaries, percent changes categories summaries and analysis of covariance.

Due to the war situation in Ukraine, data from subjects in Ukraine sites may be incomplete, missing or uncleaned. A sensitivity analysis of the BMD will be performed, excluding data collected after 24FEB2022 for subjects from Ukraine sites, for the Follow-up period only (Follow-up Safety set), as the main treatment period was completed before the start of the war. Listings will contain country information to identify those subjects.

4.8.4. **Laboratory Data**

Clinical laboratory values will be expressed using conventional SI units.

Hematology, coagulation parameters, chemistry and lipids are assessed at screening and at each visit during the treatment period, as well as at M1 FU and M3 FU visits. E2, progesterone (P4) and LH will be assessed at each visit during the treatment period (not at the screening visit), as well as at M1 FU and M3 FU visits. Serum levels of anti-müllerian hormone (AMH) and fasting glucose will be assessed on Day 1 only. Serum levels of sex hormone-binding globulin (SHBG) will be assessed on Day 1 and Month 3 and Month 6 visits. Follicle-Stimulating Hormone (FSH) will be assessed at Month 3 FU at local laboratories for subjects who did not resume menses.

Blood samples will be analyzed by the central laboratory. Due to the COVID-19 situation, some samples may be collected by local laboratories.

Laboratory parameters values will be summarized for each visit for each treatment group and overall including within-group changes from baseline. In the event of repeated values, the last non-missing value per study day/time will be used.

Semi-quantitative laboratory results will be imputed as follows: “<X.X” will be imputed as “X.X/2” and “>X.X” will be imputed as “X.X”.

Evaluation of shifts for changes from baseline to all visits, according to the normal ranges with categories “Low”, “Normal”, “High” where available (for ALT, AST, ALP, Bilirubin, Albumin, Calcium, Creatine Kinase) and according to normal/abnormal/abnormal clinically significant information collected otherwise, will be provided.

For Liver Function Test (LFT) parameters (ALT, AST, ALP, total bilirubin, albumin), evaluation of shifts for changes from baseline to the worst (highest) on-treatment value (including unscheduled assessments) according to the normal ranges with categories “Low”, “Normal”, “High” will be provided.

Assessments from local laboratories will not be included in summaries and shift tables, as normal ranges may not be comparable.

All laboratory data will be provided in data listings.

A subset listing will be presented for all abnormal clinically significant laboratory values.

For the lipid panel (Triglycerides, HDL, LDL, Total Cholesterol, LDL/HDL ratio), the percent change from baseline to each on-study visit will also be summarized.

For serum levels of E2, P4, SHBG and LH, only the actual values at each time point will be summarized. Note: in the event of repeated values, the last non-missing value per study day/time will be used.

In addition, for E2, the proportion of subjects with the following categories:

- E2 <20 pg/mL,
- E2 between ≥20 and <60 pg/mL,
- E2 ≥60 pg/mL

at each visit and for each treatment group will be provided in tables and graphically (bar charts).

The relationship between E2 serum levels and bone mineral density loss will be explored graphically in different ways:

- First, the E2 serum level measured at Month 6 will be plotted on the x-axis against the percent BMD change measured at Month 6 on the left side y-axis and the percentage of subjects with clinically meaningful reduction of DYS (resp. NMPP) with stable or decreased use of analgesics for EAP on the right side y-axis.
- Second, bar charts will be produced at each visit up to Month 6 and presented with the percent BMD change from baseline to Month 6 on the y-axis and the proportion of subjects with the E2 categories described above on the x-axis. Similar graphs will be produced for each of the BMD sites (femoral neck, hip, and spine).
- Lastly, bar charts of the proportion of subjects with the E2 categories described above on the x-axis will be plotted against the percentage of responders according to the primary endpoint.

For P4, since P4 above 10 nMol/L may suggest luteal activity and hence ovulation, the proportion of subjects with P4 >10.0 nMol/L at least once from Study Day 1 to Month 3, from Month 3 to Month 6, and from Study Day 1 to Month 6 will also be provided in tables and graphically (bar charts).

The number and percentage of subjects at the following categories

- LDL ≥ 160 mg/dL
- LDL ≥ 190 mg/dL
- HDL < 40 mg/dL

will be reported at each visit.

Shift tables from baseline to each visit will be provided for the following LDL and Triglycerides categories (after rounding to integer values if necessary).

LDL:

- 0: ≤ 130 mg/dL
- 1: 131 to 159 mg/dL
- 2: 160 to 189 mg/dL
- 3: ≥ 190 mg/dL

Triglycerides:

- 0: ≤ 150 mg/dL
- 1: 151 to 300 mg/dL
- 2: 301 to 500 mg/dL
- 3: 501 to 1000 mg/dL
- 4: >1000 mg/dL

4.8.5. Vital Signs and Physical Examinations

Vital signs include: Height (cm), Weight (kg), Body Mass Index, Systolic Blood Pressure (mmHg), Diastolic Blood Pressure (mmHg) and Heart Rate (beats/min).

Height is assessed at Screening; weight is assessed at Screening, Month 6 and Month 6 of the follow-up period. Other Vital Signs are assessed at Screening, Day 1, every month during the treatment period and at Months 1 and 3 from the follow-up period.

Physical Examination is assessed at Screening, Months 3 and 6 of treatment and follow-up periods.

The actual value and change from baseline to each on study evaluation will be summarized by treatment group and overall for vital signs.

By-subject listings of vital sign measurements will be presented in data listings.

Physical examination results at each time point will be summarized by treatment group and overall; shifts from baseline in physical examination findings to each on study visit will also be presented. All physical examination findings will be presented in a data listing.

4.8.6. **Electrocardiogram**

Local 12-lead ECG readings of QTcF are performed at Screening, Day 1 (pre and post-dose) and every month from the treatment period, as well as Months 1,3,6 from the follow-up period.

Baseline will be defined as the highest value prior to first dose.

Actual values and changes from baseline of ECG results will be summarized descriptively, as well as the number and percent of subjects with normal, abnormal and clinically significant abnormal results at baseline and each study visit by treatment group and overall.

All ECG data for each subject will be provided in data listings.

4.8.7. **Concomitant Medications**

Prior and Concomitant medications will be coded using the WHO Drug dictionary. Frequencies and percentages will be presented for each treatment group and overall by Anatomic Therapeutic Class (ATC) and preferred name. Vitamin D and Calcium supplementation as collected in the eCRF at each visit will be summarized as well.

The use of concomitant medications will be included in a by-subject data listing.

Prior medications are those the subject used prior to the first day of treatment.

All medications administered between the date of the first dose of study drug and the date of the last dose of study drug, inclusive, (i.e., all medications starting or ongoing during the time interval) are concomitant. A medication taken prior to Study Day 1 and continuing post Study Day 1 will be considered both prior and concomitant.

Prior interventional surgeries/procedures for endometriosis/endometriosis symptoms will also be summarized.

4.8.8. **Transvaginal Ultrasound (TVUS)**

TVUS will be performed at Screening, Month 3, Month 6, Month 3 and Month 6 of the Follow-up period.

Actual values and changes from baseline will be summarized by treatment group and overall for uterus length, width, and depth in mm and corresponding uterine volume in cm3, as well as endometrium thickness in mm.

The number and percentage of subjects with any ovarian endometrioma with a diameter of 7 cm or greater, and presence of any uterus, left ovary or right ovary abnormality over time will be presented by treatment group and overall.

4.8.9. **Other Examinations**

At least two endometrial biopsies for histological assessment were obtained for each subject:

- One at screening, unless an endometrium biopsy was performed within the past 6 months of the screening visit, which shows no atypical hyperplasia or adenocarcinoma and for which slides are available for current study assessment through retrospective central laboratory reading.
- One at Month 6 if the endometrium thickness via TVUS is > 5 mm
- One at Month 1 FU, 2 FU, 3 FU, 4 FU, 5 FU or 6 FU if not obtained at Month 6 or if the preceding biopsy diagnosis is different than "benign endometrium"

A summary at each time point will be performed (number and percentage of subjects with normal, abnormal, not assessable).

Biopsy diagnosis will be summarized by classification and primary diagnosis at each visit. Subjects will be counted only once per classification and diagnosis. A subject can be counted in several diagnoses.

Possible classifications and diagnoses are:

- Benign Endometrium without Hyperplasia or Atypia: Proliferative, Atrophic, Metaplastic Changes, Secretory, Indeterminate, Inactive, Polyp, Endometritis, Menstrual phase, Other
- Hyperplasia: Simple with Atypia, Simple without Atypia, Complex with Atypia, Complex without Atypia
- Malignant Endometrial Neoplasm: Carcinoma, Sarcoma, Other Malignant, Not applicable.

If any diagnosis does not correspond to any on the list, the diagnosis will be coded as: Other Malignant if classification is Malignant Endometrial Neoplasm, and Other for any other classification.

For biopsies with reported primary diagnosis "Hyperplasia" or "Malignant Endometrial Neoplasm" at other visits than screening, a second read of the biopsy was to be performed. If the diagnosis of the two reads was different, then the sample was to be re-read a 3rd time. In case of multiple readings, the following rules will be applied for the statistical analysis:

- If 2 readings out of 3 lead to the same diagnosis, one of these 2 records will be selected.
- Otherwise, the record corresponding to the most severe diagnosis should be selected considering the following order, from less severe to more severe: simple without Atypia hyperplasia < complex without Atypia hyperplasia < simple with atypia hyperplasia < complex with Atypia hyperplasia < endometrial malignant neoplasm < other malignant neoplasm.

In addition, in order to perform a quality check of the readings, 15% of biopsies were sent blinded for a second read when the primary diagnosis was "Benign Endometrium". In case of multiple readings when the primary diagnosis from the first read is "Benign Endometrium", this first read will be used for the analysis.

Endometrial biopsy results will be reported in listings.

Due to the war situation in Ukraine, some biopsy samples may have been collected and analyzed by local laboratories. Those assessments will be flagged in the listings, for the Follow-up period only, as the main treatment period was completed before the start of the war. Information about local biopsy assessment is reported in Data Management Close queries file. This file will be provided by data management department and reviewed by Sponsor. Biopsy assessments performed locally can be identified in the queries list from Endometrial Biopsy page, by subject and visit and by word search, with answer to query containing the word "local".

Gynecological Examination is assessed at Screening, Month 3, Month 6 and Month 3 of Follow-Up period.

Breast examination (by palpation) will be performed at the Screening, Month 6 and Month 3 FU visits.

A summary at each time point will be performed (number and percentage of subjects with normal, abnormal, abnormal significant) for Gynecological and Breast Examinations. Results will also be reported in listings.

PAP Smear is collected at Screening. A listing of abnormality information and a listing of diagnosis results will be provided, including possible unscheduled assessments.

4.8.10. **Columbia-Suicide Severity Rating Scale (C-SSRS)**

The C-SSRS questionnaire prospectively assess the occurrence of treatment-emergent suicidal ideation and behavior.

Subjects will complete:

- either the "Baseline" C-SSRS version, capturing lifetime history of suicidal ideation and behavior, for subjects under screening at the time of/after protocol amendment 3 implementation,

- or the "already enrolled subjects" C-SSRS version, for subjects already randomized at the time of protocol amendment 3 implementation, and who are providing answers to the C-SSRS for the first time during the study

- the "since last study visit" C-SSRS version, dedicated to subjects for whom the C-SSRS was completed at the previous study visit, for use at all remaining study visits.

The "Baseline" C-SSRS version is to be completed during Screening, or at Day 1 if not done during Screening.

The following C-SSRS data will be summarized by time point (baseline "Lifetime" or Prior to Study Entry, "Since Study Start" and each post-baseline time point):

- Incidence of the following suicidal ideation:
    - Wish to be dead.

- Non-specific active suicidal thoughts.
- Active suicidal ideation with any methods (not plan) without intent to act.
- Active suicidal ideation with some intent to act, without specific plan.
- Active suicidal ideation with specific plan and intent.

- Incidence of the following suicidal behavior:
  - Actual attempt.
  - Interrupted attempt.
  - Aborted attempt.
  - Preparatory acts or behavior.
  - Completed suicide (Actual attempt leading to death).

The number of subjects with any suicidal ideation, any suicidal behavior, and any suicidal ideation or behavior will be summarized at each time point.

The number of subjects experiencing the following, at any time post-baseline, will also be summarized

- Suicidal ideation or behavior
- Emergence of suicidal ideation.
- Worsening of suicidal ideation.
- Emergence of suicidal behavior.

Emergence of suicidal ideation/behavior is defined as having no suicidal ideation/behavior at baseline and having reported any type of suicidal ideation/behavior at any time post-baseline (including "Since Study Start" answers). Worsening of suicidal ideation is defined to occur when the most severe suicidal ideation rating since study start at any time post-baseline is more severe than its rating at baseline. The Suicidal ideation rating is defined as the following, as recommended in the C-SSRS scoring and analysis guide in https://cssrs.columbia.edu/wp-content/uploads/ScoringandDataAnalysisGuide-for-Clinical-Trials-1.pdf:

Maximal suicidal ideation category during the period (baseline, post-baseline up to Month 6, FU period) where categories range from 0 to 5:

0. No ideation
1. Wish to be dead
2. Non-specific active suicidal thoughts.

3. Active suicidal ideation with any methods (not plan) without intent to act.
4. Active suicidal ideation with some intent to act, without specific plan.
5. Active suicidal ideation with specific plan and intent.

If data of suicidal ideation or behavior is missing at baseline then the subject will not be included in summaries of emergence or worsening of suicidal ideation or behavior. The "already enrolled subjects" C-SSRS version ("Prior to Entry Study") will be used as baseline for subjects already randomized at the time of implementation of Amendment 3. If a subject completed this "prior to Entry Study" questionnaire at several visits, the worst severity answers will be considered as baseline in the analysis.

The same analysis will be repeated at any time post-baseline from Month 1 FU to Month 6 FU, on the follow-up Safety analysis set.

4.8.11. **Time to the first post-treatment menses**

The first post-treatment menses will be defined from the Vaginal Bleeding daily ediary, as the first day for which the subject recorded bleeding after the last treatment administration date of the main study, i.e., the first day with "BLEEDING" or "HEAVY BLEEDING" after the last treatment administration date of the main study.

Time to first post-treatment menses is defined as:

Time to first post-treatment menses (days) = (day of first post-treatment menses – last treatment administration date) + 1

Only data from the follow-up treatment period will be included, and the analysis will be provided on the Follow-up Safety Analysis Set, by treatment group.

Subjects who have not recorded any bleeding up to the end of the follow-up period will be considered as right-censored data at the last ediary vaginal bleeding record date before the end of the main follow-up period (as defined in section 4.2.11) with time of censoring = last non-missing ediary vaginal bleeding record date – last treatment administration date +1.

The probability of subjects having bleeding will be displayed by treatment group in the form of cumulative probability curves estimated using the non-parametric Kaplan-Meier method. Point estimates and corresponding 2-sided 95% CIs using the Greenwood's variance estimate will be provided at 28 days, 56 days, 84 days, 112 days, 140 days, 168 days, 196 days after the last treatment administration date. Percentiles (25th, 50th, and 75th) and corresponding 2-sided 95% CIs will also be provided; "N/A" will be presented for any of the percentiles not reached.

Subjects with or without first post-treatment menses, time to first post-treatment menses, end of study status and end of study date will be presented in a by-subject listing.

4.8.12. **Unblinding of subjects**

Subjects that were unblinded during the course of the study will be summarized and listed. The cumulative unblinding history report from IWRS will be provided by the Sponsor in an excel format. This file will only contain unblinding status but no treatment information.

4.9. Exploratory Analyses

Bone turnover markers will be analyzed on the Safety Population.

Blood samples for exploratory bone biomarkers include, but may not be limited to, collagen type 1 β-carboxy-telopeptide (CTx), procollagen 1 Intact N-Terminal (P1NP), bone-specific alkaline phosphatase (B-ALP) and osteocalcin as part of clinical chemistry, and are collected at Day 1, Month 3, Month 6 and Month 3 Follow-up. These exploratory data will be kept blinded up to the unblinding of the database.

Actual values and changes from baseline will be summarized overtime by treatment group. Bone marker data will be reported in listings as well.

5. CHANGES TO PLANNED ANALYSES

The following change to the protocol-defined statistical analyses is presented in this statistical analysis plan:

For secondary endpoints of change from baseline to Month 6 that are collected daily using the eDiary, the protocol states (Sections 13.5.3.1, 13.5.3.2, 13.5.3.3, 13.5.3.5, 13.5.3.6, 13.5.4) that the endpoints will be based on the 28 calendar days immediately prior to and including the day of the Month 6 visit (or the 28 calendar days immediately prior to and including the last dose date for subjects who discontinue treatment prior to Month 6).

In this SAP this definition has been modified to not include the Month 6 day, as for subjects entering the Extension Study, the Extension Treatment is received at Month 6 (i.e., on the Month 6 day), and for subjects entering the Main Study Follow-up, no treatment is received at Month 6 (i.e., on the Month 6 day).

The end of the 28-day period prior to Month 6 is defined in Section 4.2.11 as the last treatment administration date in the Main Study.

The following endpoints are impacted:

- Ranked Secondary Endpoints:
    - Change from baseline to Month 6 in DYS (VRS) and in NMPP (VRS) (Section 4.6.3.1)
    - Change from baseline to Month 6 in dyschezia (NRS) (Section 4.6.3.2)
    - Change from baseline to Month 6 in overall pelvic pain (NRS) (Section 4.6.3.3)
    - Change from baseline to Month 6 in dyspareunia (VRS) (Section 4.6.3.5)
    - No analgesics use/ no opiate use for EAP during the preceding 4-week period at Month 6 (Section 4.6.3.6)
- Additional Secondary Endpoints (Section 4.6.4): endpoints based on daily e-Diary entries

Updates from SAP Amendments since SAP V1.0 are summarized in the table below:

SAP Version	Update description	SAP Section	Reason
V2.0	Addition of MCT analysis using Month 6 data for ranked secondary endpoints - Addition of corresponding M6 threshold analysis population - Addition of random sample for Month 6 MCT (N=200) - Clarification of blinding/unblinding process for M6 - For threshold analysis population at Month 3, available anchor data needed among PGIC or PGIS or DYSME or NMPP only (since only the primary endpoint MCT are assessed at Month 3)	Section 3.1 Section 4.6.1 Section 4.3	FDA Advice/Information request Dated 16Aug2021 (Comment 14a)
V2.0	Correction of definition for threshold analysis population and random sample to clarify that PGIC is needed at Month 3 (or Month 6) only (not at baseline)	Section 3.1 Section 4.6.1	FDA Advice/Information request Dated 16Aug2021 (Comment 14b)
V2.0	For MCT analysis of EHP-30 Pain Domain, increase of the size of the random sample to n=200 instead of n=100	Section 4.6.1	FDA Advice/Information request Dated 16Aug2021 (Comment 15)
V2.0	Clarification of rationale for the while on treatment strategy	Section 4.6.2.1.1	FDA Advice/Information request

			Dated 16Aug2021 (Comment 19a)
V2.0	Added subgroups of Age and BMI	Section 4.2.6	FDA Advice/Information request Dated 16Aug2021 (Comment 20)
V2.0	Clarification of how the theoretical Month X date will be used if the subject missed the Month X visit: The theoretical Month X date will be the reference end date of the Month X 28-day period in which the subject may have available daily diary data.	Section 4.2.11	FDA Advice/Information request Dated 16Aug2021 (Comment 22)
V2.0	Clarification of "the last 28 days on treatment will be used at all subsequent scheduled visits for those discontinued subjects" applicability and of the twos ways to perform descriptive summaries on efficacy endpoints	Section 4.6.4	FDA Advice/Information request Dated 16Aug2021 (Comment 24)
V2.0	Data cleaning rule for considering daily ediary records entered after 00:00 on the correct day.	Section 4.2.8.3	Blinded data review finding
V2.0	Deletion of potential major deviations non-applicable on eDiary completion compliance	Section 3.2	Efficacy analyses on daily eDiary endpoints are already considering subjects with a minimum of daily diary data completed
V2.0	Update of visit window rule for baseline questionnaires: the	Section 4.2.9	Blinded data review finding: update to consider baseline

	most recent data entered before first dose regardless of questionnaire name		questionnaires entered in the wrong visit page (Month 1) but correct date (before first dose)
V2.0	Addition of definition of study completion status (different from eCRF completion guideline definition): Subjects who completed the treatment and follow-up periods, or completed the treatment period and entered the extension study.	Section 4.4	Per eCRF completion guideline, subjects who discontinued treatment before 3 months are considered as completed study. Addition of new relevant definition for analysis.
V2.0	Addition of C-SSRS worsening score derivation clarification and source document	Section 4.8.10	Details were missing in the SAP
V2.0	Minor updates (Typographical errors, wording update, SAS code details)		N/A
V3.0	Addition of results from Month 3 Meaningful Change Threshold analysis for Primary Endpoints	Section 4.6.1.3	Planned update with obtained relevant meaningful change thresholds for the primary endpoints
V3.0	Clarification of some major deviations definitions	Section 3.2	Outcome of blinded data review
V3.0	Addition of Screened set and Randomized set definitions	Section 3.1	Definitions were missing from the SAP
V3.0	Addition of visit windows rules for PK analysis	Section 4.2.9	Outcome of blinded data review
V3.0	Efficacy analyses repeated on Per Protocol Set only for Primary	Section 4.6	Repetition of efficacy analyses of all secondary

	and Ranked Secondary endpoints.		endpoints deemed unnecessary
V3.0	Analysis of analgesic use in Morphine Milligram Equivalents (MME) - In countries where several opioids were allowed, take the average of possible MME - Subjects from Ukraine excluded from MME analysis	Section 4.6.3.9 Section 4.6.4.8	- In order to not lose a lot of subjects from countries with several opioids are allowed and MME unknown, the average of possible MME will be taken - Some subjects from Ukraine recorded some opioids even if not allowed (confirmed eDiary data entry issue). MME is not assessable
V3.0	Change from baseline to each scheduled assessment in non-opioid, opioid and combined analgesic use for EAP during the previous 4-week period based on pill count in the eDiary; Change from baseline to each scheduled assessment in opioid analgesic use for EAP as reported in the eDiary during the previous 4-week period based on morphine milligram equivalent (MME) The mean daily pill count will be presented instead of total pill count	Section 4.6.4.7 Section 4.6.4.8	Mean daily pill count and mean daily MME amount are more appropriate than total pill count/MME in the 28-day period, as the mean adjusts with missing days.

V3.0	Update of opioid analgesic use (pill count and MME) and endpoints analyses: Summary and analysis of covariance removed. A listing will be presented. Addition of analysis of proportion of subjects with no analgesic use/no opiate use for EAP at Baseline and Month 3. Update of number of days with any opioid analgesic use endpoint analysis: Summary and negative binomial analysis removed. Only a listing will be presented.	Section 4.6.4.7 Section 4.6.4.8 Section 4.6.4.10	Due to a low number of subjects using any opioid, the planned analyses were not appropriate.
V3.0	Clarification of analgesic status derivation - The "**" criteria applies in all cases - Increase of exactly= 15% of the baseline dose is considered as stable	Section 7.1: Appendix A	Derivation rules of analgesic status needed minor clarification
V3.0	Addition of summary of time on treatment and time on study in days.	Section 4.8.1.1	Time on treatment/study in days is as relevant as time in weeks.
V3.0	Addition of Weight Group, Age Group 2, BMI Group 2 and deletion of Ethnicity group subgroup analyses for BMD.	Section 4.8.3	Subgroups analyses added per FDA Advice/Information request Dated 16Aug2021 (Comment 20) Ethnicity subgroup analysis is not relevant

V3.0	C-SSRS Already enrolled "Prior to Entry Study" questionnaire: If a subject completed this "prior to Entry Study" questionnaire at several visits, the worst severity answers will be considered as baseline in the analysis.	Section 4.8.10	Blinded Data Review finding: Some subjects completed at several visits the "Prior to Entry Study" questionnaire, while it should have been completed only once.
V3.0	EQ-5D-5l questionnaire : Addition of detail on source documentation	Section 4.6.4.24	Detail on source documentation was missing in the SAP
V3.0	Minor updates (Typographical errors, wording update, SAS code details)		N/A
V4.0	Clarification that the assessment of stable or decreased use of analgesics for EAP is assessed separately for DYS and NMPP, using days with uterine bleeding and days with no uterine bleeding respectively.	Section 4.6.2.1.1 Section 4.6.2.1.2	Derivation rules of analgesic status for each co-primary endpoint needed clarification.
V4.0	Added an additional analysis of DYS and NMPP where the assessment of stable or decreased use of analgesics for EAP will be assessed over all days rather than separately for bleeding days and non-bleeding days	Section 4.6.5.3	Add additional analysis.
V4.0	Addition of results from Month 6 Meaningful Change Threshold analysis for Ranked Secondary Endpoints	Section 4.6.1.4 Section 4.6.1.1.2	Planned update with obtained relevant meaningful change thresholds for the Ranked Secondary Endpoints
V5.0	Addition of sensitivity analysis excluding data from Ukraine site	Section 4.6.4.1 Section 4.8.3	Due to the war situation in Ukraine, data from subjects

	after the start of war on 24FEB2022, for proportion of responders for DYS and NMPP and BMD analysis, for Follow-up period only.		in Ukraine sites may be incomplete, missing or uncleaned. Sensitivity analysis is added to assess the impact.
V5.0	Assessments from Local Laboratories (due to Covid-19 situation) will be excluded from summary tables of laboratory data	Section 4.8.4	Local laboratories may not have comparable normal ranges, thus it does not make sense to pool them with central laboratory data in numeric summaries.
V5.0	Biopsy assessments from Local Laboratories (due to war situation in Ukraine) should be flagged in listings for the follow-up period. Information is available in Data Management Closed Queries File.	Section 4.8.9	Due to war situation in Ukraine, there were some Biopsies assessed locally. This information should be incorporated in listings.
V5.0	Addition of listings for PAP Smear for both abnormality and diagnosis information.	Section 4.8.9	PAP Smear listings were missing from original SAP Version.
V5.0	Addition of table and listing of subjects who were unblinding during the course of the study	Section 4.8.12	At least one patient was subject to unblinding during the course of the study, table and listing is added for complete reporting.

6. REFERENCES

Cappelleri, & Bushmakin. (2020). *Biopharmaceutical Report.*

Cappelleri, & Bushmakin. (2021). *Biopharmaceutical Report.*

Griffiths, & al, e. (2018). Psychometric Properties in the Face of Missing Data- A Simulation Study Assessing the Effect of Missing Data on Test-Retest Reliability in Diary Studies. *Quality of Life Research 25th Annual Conference.*

Ratitch, B., Lipkovich, I., & O'Kelly, M. (2013). Combining Analysis Results from Multiply Imputed Categorical Data. *PharmaSUG*, Paper SP03.

7. APPENDICES

7.1. Appendix A: Analgesic change during treatment period

Use of No Analgesics at Baseline		
Analgesic used during Screening	**Analgesic dose status at end of study**	**Assessment of Change**
None	None**	Stable/Decrease
	Narcotic analgesic and/or ibuprofen is started	Increase
Use of Only Ibuprofen at Baseline		
Analgesic uses at Baseline	**Analgesic dose status at end of study**	**Assessment of Change**
ibuprofen	Dose stopped, decreases, or is stable**	Stable/Decrease
	Dose increases by 15% or more	Increase
	Narcotic analgesic is substituted or added	Increase
Use of Only Narcotic Analgesic at Baseline		
Analgesic uses at Baseline	**Analgesic dose status at end of study**	**Assessment of Change**
Narcotic analgesic	Dose stopped, decreases, or is stable**	Stable/Decrease
	Dose stopped and ibuprofen substituted (any dose)	Stable/Decrease

	Dose decreases and ibuprofen added (any dose)	Stable/Decrease
	Dose stable and ibuprofen added (any dose)**	Increase
	Dose increases by 15% or more	Increase
Use of Ibuprofen and Narcotic Analgesic at Baseline		
Analgesic uses at Baseline	**Analgesic dose status at end of study**	**Assessment of Change**
ibuprofen + narcotic analgesic	Ibuprofen dose stops + narcotic analgesic use stops, decreases, or is stable**	Stable/Decrease
	Ibuprofen use stops + narcotic analgesic dose increases by more than 15%	Increase
	Ibuprofen dose decreases + narcotic analgesic use stops, decreases, or is stable**	Stable/Decrease
	Ibuprofen dose decreases + narcotic analgesic dose increases by more than 15%	Increase
	Ibuprofen dose stable + narcotic analgesic use stops, decreases, or is stable**	Stable/Decrease
	Ibuprofen dose stable + narcotic analgesic dose increases by more than 15%	Increase

	Ibuprofen dose increases by more than 15% + narcotic analgesic use stops	Stable/Decrease
	Ibuprofen dose increases by more than 15% + narcotic analgesic dose decreases	Stable/Decrease
	Ibuprofen dose increases by more than 15% + narcotic analgesic dose is stable**	Increase
	Ibuprofen dose increases by more than 15% + narcotic analgesic dose increases by 15% or more	Increase

**Stable =Dose is the same as the screening dose or increases by less than or equal to 15% of the screening dose. In addition, a subject can increase the monthly total dose of analgesics by one pill of analgesic (either ibuprofen or narcotic) and still be considered stable.

7.2. Appendix B: EHP-30 Scoring

The mapping of the response categories for the EHP-30 core questionnaire and the modular questionnaire is as follows:
0 = Never, 1 = Rarely, 2 = Sometimes, 3 = Often, 4 = Always
For the modular questionnaire, records indicating "Not Applicable" will be excluded from summary/analysis.

Missing data will not be imputed.

The EHP-30 core questionnaire is structured into the following five dimensions.
1. Pain (Questions 1 – 11)
2. Control and Powerlessness (Questions 12 – 17)
3. Emotional Well-Being (Questions 18 – 23)
4. Social Support (Questions 24 – 27)
5. Self-Image (Questions 28 – 30)

Each dimension for the core and modular questionnaire is calculated on a scale from 0 to

100.
0 = best possible health status as measured by the questionnaire; 100 = worst possible health status as measured by the questionnaire. Within each dimension, assessment will be based on the following equation:

Sum of scores for each dimension / 4 (maximum score per item) * number of items in the dimension

Sum of the scores for each item in the dimension
Core questionnaire: Pain
((Ql + Q2 + Q3 + Q4 + Q5 + Q6 + Q7 + Q8 + Q9 + Q10 + Q11)/44) × 100
2. Core questionnaire: Control and Powerlessness
((Q12 + Q13 + Q14 + Q15 + Q16 + Q17)/24) × 100
3. Core questionnaire: Emotional well-being
((Q18 + Q19 + Q20 + Q21 + Q22 + Q23)/24) × 100
4. Core questionnaire: Social Support
((Q24 + Q25 + Q26 + Q27)/16) × 100
5. Core questionnaire: Self-Image

((Q28 + Q29 + Q30)/12) × 100

Modular questionnaire: Sexual intercourse
((C1 + C2 + C3 + C4 + C5)/(total or 20)) × 100
For Modular questionnaire: Sexual intercourse, if a "not applicable" box is checked for one or more of the five items (Questions C1 – C5), the score is calculated by omitting those "not applicable" items from the numerator and the denominator.
If any of the components of a dimension on the core or modular questionnaire is missing, no score will be calculated for that dimension. For example, if answer for Q1 is missing, no score for Core questionnaire: Pain is calculated. If answer for C1 is missing, no score for Modular questionnaire: Sexual intercourse will be calculated.

7.3. Appendix C: HRPQ Scoring

Data Cleaning
1. No imputation of data should be conducted. Missing data fields should be treated as missing.

2. For respondents indicating they are not currently employed (checking box 1c), Responses 2 – 4 should be coded as missing.
3. Code Response 3b to zero if Response 3c is checked. Code Response 6b to zero if Response 6c is checked.
4. The hours of work missed at work or in the home cannot be greater than the hours of work for which they were scheduled.
a. If Response 3b is > Response 2, Response 3b = Response 2
b. If Response 6b is > Response 5, Response 6b = Response 5
Note: Respondents with no reported scheduled hours (zero hours or missing values) of paid work (Response 2) or planned hours of work in the household (Response 5), should be excluded from the analyses regarding productivity in those respective venues.
-

Calculation of hours of lost work due to absenteeism by class of work
5. Calculate the number of hours absent by class of work:
a. Workplace: Response 3b
b. Household: Response 6b

Calculation of hours of lost work due to presenteeism by class of work
6. Calculate the number of hours worked by class of work:
a. Workplace: Response 2 – Response 3b
b. Household: Response 5 – Response 6b
7. Calculate the hours of work lost due to presenteeism by class of work:
a. Workplace: Calc 6a × (Response 4/100)
b. Household: Calc 6b × (Response 7/100)

Calculation of total [absenteeism + presenteeism] hours of lost work by class of work
8. Calculate the sum by class of work:
a. Workplace: Calc 5a + Calc 7a
b. Household: Calc 5b + Calc 7b

Calculation of the % of scheduled work lost due to absenteeism, presenteeism, and total productivity loss by class of work
9. Calculate the % of work lost due to absenteeism
a. Workplace: (Calc 5a/Response 2) × 100
b. Household: (Calc 5b/Response 5) × 100
10. Calculate the % of work lost due to presenteeism
a. Workplace: (Calc 7a/Response 2) × 100

b. Household: (Calc 7b/Response 5) × 100
11. Calculate the total % of work lost
a. Workplace: Calc 9a + Calc 10a
b. Household: Calc 9b + Calc 10b

Calculation of the workplace productivity impacts by current employment status
12. Subset subjects by their current employment status to report productivity impacts among those employed full-time and those employed part-time

Calculation of hours of lifetime productivity lost until the time of measurement due to illness impacts on workforce participation
Note: Calculations 13 and 14 should be made for only those respondents who have non-zero non-missing responses to Question 2
13. Mean weekly hours of scheduled work in previous week for subjects full-time employed
a. Mean of Response 2 among subjects selecting Response 1a at baseline
14. Mean weekly hours of scheduled work in previous week for subjects part-time employed at baseline
a. Mean of Response 2 among subjects selecting Response 1b at baseline
Note: Calculations 15 through 17 should be made for only those respondents indicating they have experienced that workforce participation impact
15. Hours of lifetime productivity lost due to illness induced part-time employment when subject would have continued full-time employment
a. ((Response 9a × 4.33 wks/mo) + (Response 9b × 52 wks/yr)) × (Calc 13 – Calc 14)
16. Hours of lifetime productivity lost due to illness induced premature retirement when subject would have continued full-time employment
a. ((Response 9c × 4.33 wks/mo) + (Response 9d × 52 wks/yr)) × Calc 13
17. Hours of lifetime productivity lost due to illness induced premature retirement when subject would have continued part-time employment
a. ((Response 9e × 4.33 wks/mo) + (Response 9f × 52 wks/yr)) × Calc 14
18. Total hours of lifetime productivity lost due to illness impact on workforce participation

a. Calc 15 + Calc 16 + Calc 17

7.4. Appendix D: EQ-5D TTO Scoring

Computing EQ-5D-5L crosswalk index values with SAS using the United States (US) value set:

The variables for the 5 dimensions of the EQ-5D-5L descriptive system should be named 'mobility', 'selfcare', 'activity', 'pain', and 'anxiety'. If they are given different names the syntax code below will not work properly. The 5 variables should contain the values for the different dimensions in the EQ-5D health profile (i.e. 1, 2, 3, 4, or 5). The variable 'EQindex' contains the values of the EQ-5D-5L crosswalk index values on the basis of the US set of weights.

You can copy and paste the syntax below directly into a SAS syntax window.

```
*****************************************************************

*SAS syntax code for the computation of index*

*values with the US TTO value set*

*****************************************************************;


data Euroqol.US_tto;
        set Euroqol.EQ5D_states;
        EQindex = .;


if (mobility=1 and selfcare=1 and activity=1 and pain=1 and anxiety=1) then EQindex = 1.000;
if (mobility=1 and selfcare=1 and activity=1 and pain=1 and anxiety=2) then EQindex = 0.876;
if (mobility=1 and selfcare=1 and activity=1 and pain=1 and anxiety=3) then EQindex = 0.844;
if (mobility=1 and selfcare=1 and activity=1 and pain=1 and anxiety=4) then EQindex = 0.700;
if (mobility=1 and selfcare=1 and activity=1 and pain=1 and anxiety=5) then EQindex = 0.550;
if (mobility=1 and selfcare=1 and activity=1 and pain=2 and anxiety=1) then EQindex = 0.861;
if (mobility=1 and selfcare=1 and activity=1 and pain=2 and anxiety=2) then EQindex = 0.820;
if (mobility=1 and selfcare=1 and activity=1 and pain=2 and anxiety=3) then EQindex = 0.809;
if (mobility=1 and selfcare=1 and activity=1 and pain=2 and anxiety=4) then EQindex = 0.669;
if (mobility=1 and selfcare=1 and activity=1 and pain=2 and anxiety=5) then EQindex = 0.524;
if (mobility=1 and selfcare=1 and activity=1 and pain=3 and anxiety=1) then EQindex = 0.827;
if (mobility=1 and selfcare=1 and activity=1 and pain=3 and anxiety=2) then EQindex = 0.806;
if (mobility=1 and selfcare=1 and activity=1 and pain=3 and anxiety=3) then EQindex = 0.800;
if (mobility=1 and selfcare=1 and activity=1 and pain=3 and anxiety=4) then EQindex = 0.661;
if (mobility=1 and selfcare=1 and activity=1 and pain=3 and anxiety=5) then EQindex = 0.517;
if (mobility=1 and selfcare=1 and activity=1 and pain=4 and anxiety=1) then EQindex = 0.682;
if (mobility=1 and selfcare=1 and activity=1 and pain=4 and anxiety=2) then EQindex = 0.663;
```

```
if (mobility=1 and selfcare=1 and activity=1 and pain=4 and anxiety=3) then EQindex = 0.659;
if (mobility=1 and selfcare=1 and activity=1 and pain=4 and anxiety=4) then EQindex = 0.544;
if (mobility=1 and selfcare=1 and activity=1 and pain=4 and anxiety=5) then EQindex = 0.426;
if (mobility=1 and selfcare=1 and activity=1 and pain=5 and anxiety=1) then EQindex = 0.463;
if (mobility=1 and selfcare=1 and activity=1 and pain=5 and anxiety=2) then EQindex = 0.450;
if (mobility=1 and selfcare=1 and activity=1 and pain=5 and anxiety=3) then EQindex = 0.446;
if (mobility=1 and selfcare=1 and activity=1 and pain=5 and anxiety=4) then EQindex = 0.369;
if (mobility=1 and selfcare=1 and activity=1 and pain=5 and anxiety=5) then EQindex = 0.289;
---------------------------------------------------------------------------------------------------- = -------:
---------------------------------------------------------------------------------------------------- = -------:
---------------------------------------------------------------------------------------------------- = -------:
---------------------------------------------------------------------------------------------------- = -------:
---------------------------------------------------------------------------------------------------- = -------:
---------------------------------------------------------------------------------------------------- = -------:
if (mobility=5 and selfcare=5 and activity=5 and pain=1 and anxiety=1) then EQindex = 0.178;
if (mobility=5 and selfcare=5 and activity=5 and pain=1 and anxiety=2) then EQindex = 0.165;
if (mobility=5 and selfcare=5 and activity=5 and pain=1 and anxiety=3) then EQindex = 0.162;
if (mobility=5 and selfcare=5 and activity=5 and pain=1 and anxiety=4) then EQindex = 0.113;
if (mobility=5 and selfcare=5 and activity=5 and pain=1 and anxiety=5) then EQindex = 0.063;
if (mobility=5 and selfcare=5 and activity=5 and pain=2 and anxiety=1) then EQindex = 0.152;
if (mobility=5 and selfcare=5 and activity=5 and pain=2 and anxiety=2) then EQindex = 0.132;
if (mobility=5 and selfcare=5 and activity=5 and pain=2 and anxiety=3) then EQindex = 0.127;
if (mobility=5 and selfcare=5 and activity=5 and pain=2 and anxiety=4) then EQindex = 0.083;
if (mobility=5 and selfcare=5 and activity=5 and pain=2 and anxiety=5) then EQindex = 0.037;
if (mobility=5 and selfcare=5 and activity=5 and pain=3 and anxiety=1) then EQindex = 0.145;
if (mobility=5 and selfcare=5 and activity=5 and pain=3 and anxiety=2) then EQindex = 0.124;
if (mobility=5 and selfcare=5 and activity=5 and pain=3 and anxiety=3) then EQindex = 0.118;
if (mobility=5 and selfcare=5 and activity=5 and pain=3 and anxiety=4) then EQindex = 0.075;
if (mobility=5 and selfcare=5 and activity=5 and pain=3 and anxiety=5) then EQindex = 0.030;
if (mobility=5 and selfcare=5 and activity=5 and pain=4 and anxiety=1) then EQindex = 0.078;
if (mobility=5 and selfcare=5 and activity=5 and pain=4 and anxiety=2) then EQindex = 0.060;
if (mobility=5 and selfcare=5 and activity=5 and pain=4 and anxiety=3) then EQindex = 0.055;
if (mobility=5 and selfcare=5 and activity=5 and pain=4 and anxiety=4) then EQindex = 0.015;
if (mobility=5 and selfcare=5 and activity=5 and pain=4 and anxiety=5) then EQindex = -0.026;
if (mobility=5 and selfcare=5 and activity=5 and pain=5 and anxiety=1) then EQindex = -0.024;
if (mobility=5 and selfcare=5 and activity=5 and pain=5 and anxiety=2) then EQindex = -0.037;
if (mobility=5 and selfcare=5 and activity=5 and pain=5 and anxiety=3) then EQindex = -0.040;
if (mobility=5 and selfcare=5 and activity=5 and pain=5 and anxiety=4) then EQindex = -0.074;
if (mobility=5 and selfcare=5 and activity=5 and pain=5 and anxiety=5) then EQindex = -0.109;
```

```
if (mobility = .) or (selfcare = .) or (activity = .) or (pain = .) or (anxiety = .) then EQindex = . ;

output;
run;
```

7.5. Appendix E: Mapping rules for ediary data

Table 8 Mapping Rules for [redacted] Device

Case #	impacted visit	Patient status	Schedule of [redacted] daily diary	impact on data	Cytel action on IMP intake data	Cytel action on Menstrual data	Cytel mapping	cut off date (from eCRF)
1	Month 6	patient was included in EXT	remained in MAIN TREATMENT (due to Covid remote visit or to site error in diary completion)	data present in the Main datasets should be considered for the analysis of the extension study	mapping	n/a	from MAIN TREAT to EXT TREAT	First IMP intake in EXT

2	Month 6	patient was included in EXT	rescheduled to MAIN FU (due to site error in diary completion)	data present in the Main datasets should be considered for the analysis of the extension study	n/a - missing in final DB	n/a	from MAIN FU to EXT TREAT	First IMP intake in EXT
3	Month 6	patient entered MAIN FU	rescheduled to EXT TREATMENT (site error in diary completion)	data present in the Extension datasets should be considered for the analysis of the Main study	"Is the subject eligible to the 6 months treatment free follow-up?" is YES --> delete answers to YIMPI and IMPI	n/a	from EXT TREAT to MAIN FU	MONTH 6 visit date

4	Month 6	patient entered MAIN FU	remained in MAIN TREATMENT (due to Covid remote visit or to site error in diary completion)	Data regarding IMP intake will be present for this period	"Is the subject eligible to the 6 months treatment free follow-up?" is YES --> delete answers to YIMPI and IMPI	n/a	n/a - remain in MAIN DB	MONTH 6 visit date
5	Month 6	whatever status: DUPLICATE ENTRIES	patient entered diary data in both Main and Extension studies each day	duplicates data will be present across the datasets	CASE by CASE review - [redacted] cleaning needed			

6	Day 1	patient entered MAIN TREATMENT	remained in SCREENING (due to site error in diary completion)	Data regarding IMP intake will not be present for this period and data regarding menstrual periods will be present	n/a - missing in final DB	delete the answers to “is it your Menstrual period?” and start date questions, if any	n/a - remain in MAIN DB	First IMP intake in MAIN

Table 9 Mapping Rules for Signant Health Device

Case #	impacted visit	Patient status	Schedule of Signant Health daily diary	impact on data	Cytel action on IMP intake data	Cytel action on Menstrual data	Cytel mapping	cut off date (from eCRF)
1	Month 6	patient was included in EXT	remained in MAIN TREATMENT (due to Covid remote visit or to site error in diary completion)	data present in the Main datasets should be considered for the analysis of the extension study	mapping	n/a	from MAIN TREAT to EXT TREAT	First IMP intake in EXT
2	Month 6	patient was included in EXT	rescheduled to MAIN FU (due to site error in diary completion)	data present in the Main datasets should be considered for the analysis of the extension study	mapping	n/a	from MAIN FU to EXT TREAT	First IMP intake in EXT

3	Month 6	patient entered MAIN FU	rescheduled to EXT TREATMENT (site error in diary completion)	data present in the Extension datasets should be considered for the analysis of the Main study	delete answers to Yesterday IMP intake and to Today IMP intake if any	n/a	from EXT TREAT to MAIN FU	MONTH 6 visit date
4	Month 6	patient entered MAIN FU	remained in MAIN TREATMENT (due to Covid remote visit or to site error in diary completion)	Data regarding IMP intake will be present for this period	delete answers to Yesterday IMP intake and to Today IMP intake if any	n/a	n/a - remain in MAIN DB	MONTH 6 visit date
5	Day 1	patient entered MAIN TREATMENT	remained in SCREENING (due to site error in diary completion)	Data regarding IMP intake will not be present for this period and data regarding menstrual periods will be present	mapping	delete the answers to “is it your Menstrual period?” and start date questions, if any	n/a - remain in MAIN DB	First IMP intake in MAIN

6	Any Monthly visit of the Main Treatment period	patient in MAIN TREATMENT	Set-up in screening period (due to site error in diary completion)	Data regarding IMP intake will not be present for this period and data regarding menstrual periods will be present	mapping	delete the answers to “is it your Menstrual period?” and start date questions, if any	n/a - remain in MAIN DB	First IMP intake in MAIN

Obseva-18-OBE2109-003_SAP_5.0_13JUN2022

Final Audit Report 2022-06-14

Created:	2022-06-13
By:	[redacted]
Status:	Signed
Transaction ID:	CBJCHBCAABAAh0b_RMbUzkIW4ODMVG9CthKNFw6JonnJ

"Obseva-18-OBE2109-003_SAP_5.0_13JUN2022" History

[redacted]