STATISTICAL ANALYSIS PLAN

A Randomized, Multicenter, Double-blind, Placebo-controlled Phase 2 Study to Evaluate the Safety and Efficacy of HMI-115 in Women with Moderate to Severe Endometriosis-associated Pain Over a 12-Week Treatment Period

PROTOCOL NO.: HMI-115EM201

NCT05101317

PREPARED FOR:	Hope Medicine (Nanjing) Co.,Ltd Building 16-B2-1 of Tree House Headquarter No.73 Tanmi Road, Jiangbei New Area Nanjing City, 211800 China
PREPARED BY:	Novotech (Australia) Pty Ltd Level 3, 235 Pyrmont Street Pyrmont, NSW, 2009 Australia
DATE OF ISSUE:	2025-02-10
VERSION/STATUS:	Version 1.3_Special Version for Research Paper Publication
VERSION HISTORY:	Version 0.1 (2023-12-21) Version 1.0 (2024-08-13) Version 1.1 (2024-09-03) Version 1.2 (2024-10-22) Version 1.3 (2024-12-17)
AUTHOR:	Simon Xu

Document History

Date	Version number	Protocol/ Study number	Summary of changes
2023-12-21	0.1	HMI-115EM201 v5.0	Initial SAP
2024-08-13	1.0	HMI-115EM201 v6.0	Final SAP for interim analysis
2024-09-03	1.1	HMI-115EM201 v6.0	Some analysis rules were changed requested by the sponsor for interim analysis
2024-10-22	1.2	HMI-115EM201 v6.1	Some analysis rules were changed requested by the sponsor
2024-12-17	1.3	HMI-115EM201 v6.1	Analysis population adjusted based on regulatory authorities' feedback.
2025-02-10	1.3_Special Version for Research Paper Publication	HMI-115EM201 v6.1	Generate a special version for research paper publication only

SAP APPROVAL

By my signature, I confirm that this SAP has been reviewed by Hope Medicine and has been approved for use on the HMI-115EM201 study:

Name	Title / Company	Signature	Date
Chinglin Lai	Biostatistics Consultant / Hope Medicine		
Li Yang	Medical Monitor / Hope Medicine		
Simon Xu	Statistics Manager / Novotech (Australia) Pty Ltd.		

Table of contents

List of Abbreviations

Abbreviation / Acronym	Definition / Expansion
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse event of special interest
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
$AR/R_{ac,}$	Accumulation ratio
ATC	Anatomical therapeutic chemical
$AUC_{0\text{-}tau}$	Area under the curve of plasma concentrations vs time for the dosing interval
BDRM	Blinded data review meeting
BMD	Bone mineral density
BMI	Body mass index
BOCF	Baseline observation carried forward
CI	Confidence interval
CFB	Change from baseline
C_{max}	Maximum concentration
COVID-19	Coronavirus disease 2019
CPSSS	Composite pelvic signs and symptoms score
CS	Clinically significant
CSR	Clinical study report
C_{trough}	Concentration before dosing during multiple dose administration
CV	Coefficient of variation
DXA	Dual energy X-ray absorptiometry
DMC	Data monitoring committee
DYS	Dysmenorrhea
DYSP	Dyspareunia
ECG	Electrocardiogram
eCOA	electronic Clinical Outcome Assessment
eCRF	Electronic case report form
EHP-5	Endometriosis health profile-5
EOS	End of study
EOT	End of treatment
EDV	Early discontinuation visit
EU	European
FAS	Full analysis set

Abbreviation / Acronym	Definition / Expansion
FiH	First-in-human
gCV	geometric coefficient of variation
Geomean	Geometric mean
HEENT	head, eyes, ears, nose, throat
HR	Heart rate
HRQoL	Health-related quality of life
ICE	Intercurrent event
ICF	Informed consent form
IMP	Investigational medicinal product
ISR	Injection site reaction
LLN	Lower limit of normal range
LSM	Least square mean
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed-effect model repeated measures
NADA	Neutralizing anti-drug antibody
NCA	Non-compartmental analysis
NCS	Not clinically significant
NMPP	Non-menstrual pelvic pain
NRS	Numeric rating scales
PAS	Primary Analysis Set
PD	Pharmacodynamics
PGIC	Patient global impression of change
PK	Pharmacokinetics
PKS	PK analysis set
PN	Preferred name
POPPK	Population pharmacokinetics
PT	Preferred term
PPS	Per protocol analysis set
PRL	Prolactin
PRLR	Prolactin receptor
Q2W	Once every 2 weeks
QTc	QT interval corrected for heart rate
QTcB	QT interval corrected for heart rate using Bazett's correction
QTcF	QT interval corrected for heart rate using Fridericia's correction
R	Randomization

Abbreviation / Acronym	Definition / Expansion
RTF	Rich text format
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SAR	Statistical analysis report
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
s.c.	subcutaneous
SD	Standard deviation
SE	Standard error
SoA	Schedule of activities
SOC	System Organ Class
$t_{½}$	Apparent terminal elimination half-life
TEAE	Treatment emergent adverse event
T_{max}	Time of maximum plasma concentration
TVU	Transvaginal ultrasound
ULN	Upper limit of normal range
USA	United States of America
WHO-DD	World Health Organization - Drug Dictionary
WNL	WinNonlin

1. INTRODUCTION

HMI-115 is a highly specific human monoclonal antibody that will be investigated for safety and efficacy in pre-menopausal women with moderate to severe endometriosis associated pain.

A First in-human (FiH) study with HMI-115 has already established the safety and tolerability of HMI-115 when administered subcutaneously in post-menopausal women aged between 45 to 65 years in single doses up to 240 mg and multiple doses up to 90 mg once every 2 weeks (Q2W) for 6 weeks. This study aims to evaluate the safety and efficacy of HMI-115 versus a placebo in pre-menopausal women aged 18 to 49 years diagnosed with moderate to severe endometriosis associated pain.

The analyses described in this Statistical Analysis Plan (SAP) are based upon the following study documents:

- Study Protocol Version 6.1 (19 Sep 2024)
- Electronic Case Report Form (eCRF), Version 6.0 (12 Dec 2023)

Data will be summarized and listed according to the SAP. The details will be reflected in the document of Table, Listing, and Figure Shells.

The planned analyses identified in this SAP may be included in clinical study reports (CSRs), regulatory submissions, or future manuscripts. Also, post hoc exploratory analyses not necessarily identified in this SAP may be performed to further examine study data. Any post hoc, or unplanned, exploratory analyses performed will be clearly identified as such in the final CSR.

In addition, the SAP V1.3_special version for research paper publication is generated for the requirement of research paper publication only. Compared to the SAP V1.3, the change is that the SAF will be same as the PAS for all safety analysis.

PROJECT OVERVIEW

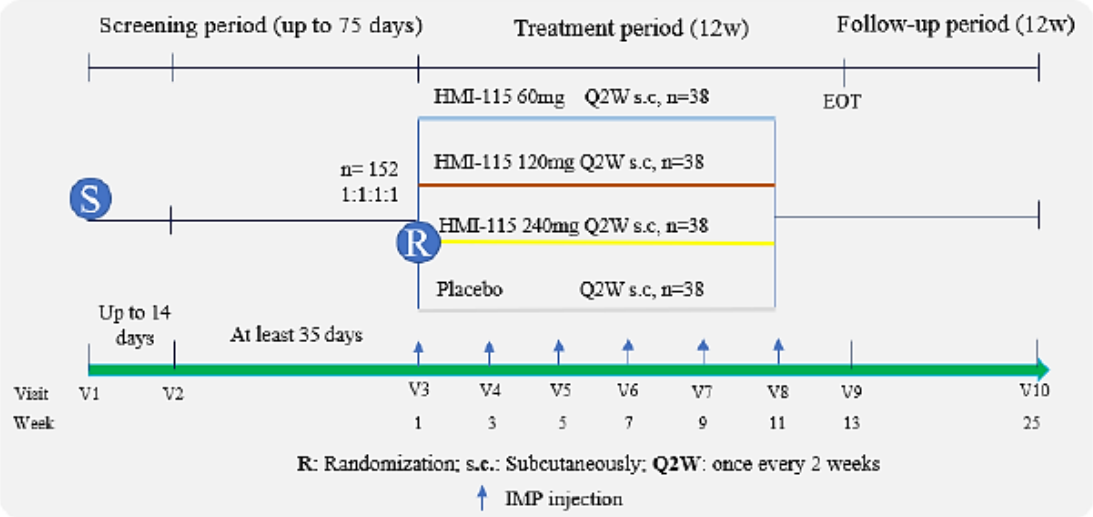
1.1 Study Design

This is a Phase 2, multicenter, double-blind, placebo-controlled, randomized study to assess the safety and efficacy of HMI-115 subcutaneously administered monoclonal antibody versus placebo in pre-menopausal women with moderate to severe endometriosis-associated pain. The planned study regions include China, the United States of America (USA) and European countries.

The study will comprise of three periods:

1. Screening Period of up to 75 days prior to first dose
   - Screening Visit 1
   - Screening Visit 2 (at least 35 days before Visit 3 to ensure at least 5 weeks baseline pain score obtained before randomization)
2. Treatment Period (12 weeks). Subjects will be given subcutaneous (s.c.) injections Q2W (once every 2 weeks).
3. Follow-up Period of 12 weeks.

Figure 1: Study Design of the Study



IMP: investigational medicinal product.

Eligible subjects will be randomized into four arms: placebo, 60 mg, 120 mg, and 240 mg HMI-115 in a 1:1:1:1 ratio with 38 subjects/arm. The randomization will be stratified with two stratification factors 1) region and 2) participation in the intensive Pharmacokinetics(PK) sampling subset.

A subject is considered to have completed the study if she has completed all phases of the study including the last scheduled procedure (follow-up at Week 25 for safety and to cover 5 half-lives) as shown in the Schedule of Activities (SoA) (Section 1.3 of Protocol).

The end of this clinical study is defined as the date of the last visit of the last subject in the study or last scheduled procedure shown in the SoA (Section 1.3 of Protocol) for the last subject in the study globally.

1.2 Objectives

1.2.1 Primary objective(s)

The primary objective is to study the efficacy of HMI-115 for 12 weeks compared with placebo in the management of endometriosis-associated pain based on subjects with surgical diagnosis.

1.2.2 Secondary objective(s)

The secondary objective is to study the safety, tolerability, and persistence of efficacy of HMI-115 compared with placebo.

1.2.3 Exploratory objective(s)

The exploratory objective is to explore the effect of HMI-115 on Quality of life, PK, PK-PD (Pharmacodynamics) relationship, and immunogenicity.

1.3 Endpoints

1.3.1 Efficacy endpoint(s)

All subjects will be administered an e-Diary after adequate training at Visit 2. Subjects are required to record uterine bleeding and endometriosis-associated pain every day at around the same time for the entire duration of the study. The use of rescue medications for endometriosis-associated pain will also be recorded.

Numeric Rating Scale (NRS) for pain will be used to measure Dysmenorrhea (DYS) from Baseline to Week 12 and Baseline to Week 24 with pain scale ranging from 0 (no pain) to 10 (severe pain). Non-menstrual pelvic pain (NMPP) and dyspareunia (DYSP) will be also measured using the NRS pain scale from Baseline to Week 12, and Week 24.

Endometriosis Daily Impact Pain Scale will be used to record the mean change every month from Baseline to Week 12 and 24 with the pain scale ranging from 0 (no pain) to 3 (severe pain).

1.3.1.1 *Primary Efficacy Variable*

- Change in DYS measured by NRS from Baseline to Week 12 measurement

1.3.1.2 *Secondary Efficacy Variable*

- Change in DYS measured by NRS from Baseline to Week 24 measurement
- Change in NMPP measured by NRS from Baseline to Week 12 measurement and Week 24 measurement
- Change in DYSP measured by NRS from Baseline to Week 12 measurement and Week 24 measurement
- Change from Baseline (CFB) in the monthly mean Endometriosis Daily Impact Pain Scale for DYS, NMPP and DYSP at Week 12 measurement and 24 measurement
- CFB by visit in permitted rescue medication use
- Change in menstrual period heaviness (bleeding) from Baseline by visit

1.3.2 Safety and Tolerability endpoints(s)

1.3.2.1 *Adverse Events (AEs)*

All AEs will be collected from the signing of informed consent form (ICF) until the end of study participation at the time points specified in the SoA (Section 1.3 of Protocol).

1.3.2.2 *Vital Signs*

The following vital signs will be measured:

- Blood pressure (systolic and diastolic [mmHg])
- Pulse (beats/min)

- Tympanic body temperature (°C)
- Respiratory rate (breaths/min)

1.3.2.3 *Physical Examinations*

A complete physical examination will include, at a minimum, assessments of General appearance, head, eyes, ears, nose, throat (HEENT), Neck (including thyroid), Lung/pulmonary, Chest (including breast), Cardiovascular, Abdomen, Back, Neurological, Extremities, Skin and Lymph nodes. A brief physical examination will include, at a minimum, assessments of General appearance, HEENT, Neck, Lung/pulmonary, Breast, Cardiovascular, Abdomen, Extremities and Skin.

The time points examinations are shown in the SoA (Section 1.3 of Protocol). Height (cm) and weight (kg) will also be measured and recorded at specified period.

1.3.2.4 *Concomitant Medication*

Concomitant medications will be recorded until last study visit.

1.3.2.5 *Twelve-Lead Electrocardiogram (ECG)*

The following 12-Lead ECG parameters will be collected:

- HR (beats/min)
- PR interval (msec)
- QRS interval (msec)
- QT interval (msec)
- QT interval corrected for heart rate (QTc) (msec) (QTcB and QTcF)

1.3.2.6 *Clinical Laboratory Tests*

Clinical laboratory tests include hematology, coagulation, clinical chemistry, urinalysis, hormones and CA-125. All tests will be obtained as outlined in the SoA (Section 1.3 of Protocol).

1.3.2.7 *Dual Energy X-ray Absorptiometry (DXA) Scan*

DXA scan results contain the lumbar spine, total hip, and femoral neck. The blinded reads of scans and management will be made based on the central review or local readings.

Subjects who prematurely discontinue IMP treatment prior to week 11 will not have a DXA at the end of treatment (EOT) visit or in the Post-treatment Follow-up Period. Subjects who become pregnant during the study will not have protocol required DXA scans.

1.3.2.8 *Transvaginal Ultrasound (TVU)*

- Assessments for TVU include, but are not limited to, the following:
- Endometrial thickness (double layer, mm)
- Other endometrial findings such as heterogeneity or other abnormality
- Presence, size and appearance of ovarian cysts or masses
- Presence, location and size of uterine fibroids
- Endometriomas

The Screening TVU may be repeated per protocol (e.g., ovarian cyst criteria) or as clinically appropriate. An unscheduled (elective) TVU may also be performed as clinically indicated for subject evaluation during the course of the study.

1.3.3 Exploratory endpoints(s)

1.3.3.1 *Other Patient-Reported Outcomes*

- Quality of Life assessment with health-related quality of life questionnaire (HRQoL) (Endometriosis Health Profile-5 [EHP-5])

The EHP-5 is a HRQoL scoring that has a core questionnaire and modular questionnaire with a 4-point rating scale for each item (0 = never, 1=rarely, 2=sometimes, 3=often, 4 = always). Anything that is not applicable to the subject in the questionnaire will be marked not relevant.

- Response to Patient Global Impression of Change (PGIC)

Subjects will score the PGIC as shown in the SoA (Section 1.3 of Protocol). The PGIC questionnaire is a self-reported 7-point scale rating a subject's overall impression of change in their endometriosis-associated pain since start of IMP, where 1=very much improved, 2=much improved, 3=minimally improved, 4=not changed, 5=minimally worse, 6=much worse, 7=very much worse.

1.3.3.2 ***PK***

- $AUC_{0\text{-}10d}$, $AUC_{0\text{-}inf,}$ $\%AUC_{extrap,}$ C_{max}, T_{max}, λz, $t_{1/2}$, MRT, CL/F, and V_z/F following dose 1
- $T_{max\ ss}$, $C_{ss\ trough}$, C_{ss_max}, C_{ss_av}, $t_{1/2}$, $AUC_{0\text{-}tau,ss}$, CL/F, V_z/F et al following last dose
- Accumulation ratio (AR) for both C_{max} and AUC

PK concentration data will be obtained at the time point(s) described in the protocol as follows:

Blood samples will be collected to determine PK profile from Day 1 to Day 169. A full PK sampling strategy will be followed for a subgroup of ten subjects per arm conduct at selected sites (refer Table 1-2 in the Protocol). Only sparse sampling will be obtained from the rest of the subjects (refer Table 1-3 in the Protocol).

1.3.3.3 ***PK-PD***

Relationship between HMI-115 plasma concentration/exposure and pelvic pain score measured by NRS (cyclic = DYS, non-cyclic = NMPP+DYSP, and average)/side effects will be detected.

1.3.3.4 ***Immunogenicity***

- Anti-drug antibody (ADA)
- Neutralizing antibody (NABs).

To assess the overall immunogenicity potential of HMI-115, subject's blood will be tested for anti-drug antibody (ADA) production: ADA (positive/negative and titer) and neutralizing antibodies (NAB) (positive/negative and titer only in subjects with ADA positive result). The time points for ADA determinations in blood are provided in SoA with one additional sampling point at week 17. All samples for ADA should be collected together with the corresponding PK sample.

1.4 Sample Size

The sample size calculation for this 4-arm parallel-group study is based on the assumption of a SD of 2.0 and a paired t-test at the 0.05 alpha level. In this case, with 34 subjects per group, there is an 80% power to detect a 1-point difference in the CFB to Week 12 in DYS measured by NRS within a treatment group. This sample size is also expected to provide adequate precision in the point estimate and 95% CI of the CFB for each treatment group.

With an assumption that up to 10% of subjects may prematurely discontinue the study without providing data for efficacy evaluations, enrollment of approximately 38 subjects per treatment group is planned. Approximately 190 subjects will be screened to achieve 152 subjects randomly assigned to HMI-115 or placebo group and 136 evaluable subjects for an estimated total of 34 evaluable subjects per treatment group.

1.5 Randomization

All subjects will be centrally randomized using an Interactive Response Technology (IRT). The randomization will be stratified with two stratification factors 1) region and 2) participation in the intensive PK sampling subset. Each subject will be assigned a unique number (randomization number) that encodes the subject's assignment to different arms of the study.

Eligible subjects will be randomized into four arms: placebo, 60 mg, 120 mg, and 240 mg HMI-115 in a 1:1:1:1 ratio with 38 subjects/arm.

2. STATISTICAL CONSIDERATIONS

Data will be handled and processed per the sponsor's representative (Novotech) Standard Operating Procedures (SOPs), which are written based on the principles of good clinical practice (GCP).

2.1 General Considerations

- 'Baseline' is defined as the last available pre-treatment assessment (including unscheduled assessments) unless otherwise specified in the SAP.
- 'End of Study (EOS)' is defined as the last available post-treatment assessment.
- 'Study Day' will be calculated relative to the date of First Administration of IMP (i.e., prior to treatment, Study Day = Assessment Date -First Administration of IMP Date; otherwise, Study Day = Assessment Date - First Administration of IMP Date+1). The day of the first IMP administration will be Day 1. In case of partial or missing event date, study day will appear missing.
- Measurements collected at early discontinuation visit (EDV) will be mapped to the scheduled visit according to the analysis visit window in Table 12.2. If EDV measurement allocates to a visit without a scheduled measurement, EDV measurement will be considered for analysis by visit. If EDV measurement allocates to a visit with existing scheduled measurement, EDV measurement will not be considered for analysis by visit. Unscheduled visits will not be mapped to scheduled visits.
- For by-visit summaries, only scheduled assessments will be used in tables, and if there are repeated post-first dose assessments, only the earlier value will be used in summary. In the subject listing, data will be presented under the nominal visits.
- Continuous data will be summarized in terms of the number of observations (n), mean, standard deviation (SD), first quartile (Q1), median, third quartile (Q3), minimum and maximum, unless otherwise stated. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, Q1 and Q3 will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.
- Categorical data will be summarized in terms of frequency counts and percentages. Any planned collapsing of categories will be detailed in the SAP text and the data displays.
- Percentages will be presented to one decimal place. Percentages will not be presented for zero counts. Percentages will be calculated using n as the denominator unless otherwise specified. If sample sizes are small (n<10), the data displays will show the percentages, but any textual report will describe frequencies only.
- For quantitative measurements, change from baseline (CFB) at a particular visit will be calculated as (value at visit – baseline value).

  Changes from baseline in categorical data will be summarized using shift tables where appropriate.
- P-values greater than or equal to 0.001, in general, will be presented to three decimal places after rounding. P-values less than 0.001 will be presented as "<0.001" and p-values greater than 0.999 will be presented as ">0.999".
- Confidence intervals (CIs) will be presented to one more decimal place than the point estimate.

- All values below or above a limit of quantification (e.g. <0.1 or >100) will be displayed as such in data listing. When include in table or figure summary, value may be substituted by specific value as described in corresponding section.
- For some safety parameters, the within-group and between-group differences are analyzed, but nominal p-values are provided for information only. No formal conclusion will be made based on p-values.
- For subjects who prematurely discontinued during the treatment period, their data collected during the Follow-up Period will be included in the analysis.

2.2 Inferential Analyses

The null hypothesis is that there is no difference between each HMI-115 treatment group and placebo in mean change of DYS measured by NRS from Baseline to Week 12 measurement. The alternative hypothesis will be that there is a difference.

All efficacy analyses will be conducted using the PAS. Each comparison will be carried out at the two-sided 5% level of statistical unless otherwise specified.

2.3 Multiple Comparisons and Multiplicity Adjustments.

Not applicable for this study.

2.4 Treatment Groups

Treatment groups will be summarized as follows:

- 60 mg
- 120 mg
- 240 mg
- Placebo
- Overall

Placebo treatment groups will be pooled for all analyses.

3. PARTICIPANT DISPOSITION AND ANALYSIS SETS

3.1 Participant Disposition

A clear accounting of the disposition of all subjects who enter the study will be provided, from screening to study completion.

The number and percentage of subjects entering and completing the clinical study will be presented for all subjects enrolled by treatment group and overall. Subject disposition will be summarized and will include the following information: number and percentage of subjects enrolled, screen failed, reason for screen failures, number and percentage of subjects randomized and dosed, number and percentage of subjects completing the study and the number and percentage of subjects who were withdrawn (including reasons for withdrawal and including discontinuing IMP/discontinuing study).

By-subject listings of informed consent details, eligibility details, randomization details, visit dates and withdrawal/study completion details (including reason for discontinuation and duration of treatment prior to discontinuation) should be provided.

3.2 Protocol Deviations

Major protocol deviations and any action to be taken regarding the exclusion of patients or affected data from specific analyses are defined in the project-specific Protocol Deviation Specification.

All protocol deviations will be reviewed and classified on a Blinded Data Review Meeting (BDRM) shortly before database lock/unblinding. Results and analysis set assignments will be summarized in a BDRM report which will be signed off by all relevant scientific experts.

The number and percentage of subjects with any major protocol deviations and any major protocol deviations related to COVID-19 will be presented for treatment group by deviation category using all subjects randomized.

A by-subject listing of protocol deviations will also be presented. Protocol deviations related to COVID-19 will be identified.

3.3 Analysis Sets

For purposes of analysis, the following analysis sets will be used:

Analysis Set	Description
Enrolled Analysis Set (ES)	All subjects who signed the ICF (including screening failures).
Full Analysis Set (FAS)	All randomized subjects who take at least one dose of IMP. For the FAS, subjects are assigned to a treatment group based on the randomization schedule, regardless of the treatment actually received.
Primary Analysis Set (PAS)	All subjects in the FAS who have surgical diagnosis. PAS will be the primary analysis set for all efficacy analyses.
Per Protocol Analysis Set (PPS)	A subset of the PAS who do not have protocol deviations that could significantly affect the results for the primary endpoints. Subjects' inclusion/exclusion from the PPS will be determined and documented prior to the database lock and unblinding. PPS will be used to demonstrate robustness of results for the primary and secondary efficacy endpoints.
Safety Analysis Set (SAF)	All subjects who received at least one dose of IMP. Subjects will be analyzed according to the IMP they actually received. SAF

	will be used for all safety analyses. In the SAP V1.3 special version for research paper publication, the SAF will be same as the PAS.
PK Analysis Set (PKS)	All subjects in the SAF who received active treatment with at least one drug concentration value, and have no major protocol deviation that could significantly impact PK assessment.
Immunogenicity Analysis Set	The immunogenicity analysis set consists of all subjects in the SAF with at least one immunogenicity value.

All protocol deviations will be reviewed and classified regarding how they affect clinical and analysis outcomes during the BDRM. Subjects may be excluded from the PPS and PKS.

Summary of the number and percentage of subjects in each analysis set with exception of the ES by treatment group and overall, based on all randomized subjects with reasons for exclusion from each analysis set will be provided.

A by-subject listing of analysis set details will be provided on the ES. This listing will be presented by treatment group and will include center, subject identifier, inclusion/exclusion flag for each set and reason for exclusion from each set.

4. DEMOGRAPHIC AND BASELINE INFORMATION

4.1 Demographics

Demographic characteristics (age, ethnicity, race), anthropometric characteristics (height, weight at baseline, body mass index [BMI] at baseline), and diagnosis of endometriosis (method, duration) will be summarized by treatment group and overall using FAS and PAS, respectively. All demographic information will be captured directly from EDC.

The denominator for percentages will be the number of subjects with non-missing data in FAS/PAS for each treatment group or for all subjects as applicable. The duration of endometriosis (months) will be calculated as follows:

Endometriosis Duration (months) = ((Date of ICF – Date of Diagnosis + 1)/365.25)×12

For subjects who only have a year of the diagnosis date, the date of January 1 will be assumed. For subjects who have both the year and month of the diagnosis date, the first date of diagnosis (01mmmyyyy) will be assumed.

By-subject listings of demographic, anthropometric, and disease details will be provided.

4.2 Medical History

Medical history will be coded using the MedDRA 24.0 or a later version. Medical history data will be summarized on the PAS by MedDRA System Organ Class (SOC) and Preferred Term (PT). The summary will be ordered alphabetically for SOC, and then PT within SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (SOC or PT).

All data will be listed by subject including description of the disease/condition, MedDRA SOC and PT, start date, and end date (or ongoing if applicable).

4.3 Menstrual History

Quantitative measurements of historical menstrual cycle and period in days before screening, and menstrual cycle and period in days during study (from informed consent to end of study) will be summarized by treatment group and overall using PAS. Categorical lengths of historical menstrual cycle (<28, ≥28 - ≤ 35, and > 35 days), historical menstrual period (<3, ≥3 - <5, ≥5 - < 7, and ≥ 7 days), menstrual cycle during study (<28, ≥28 - ≤ 35, and > 35 days), and menstrual period during study (<3, ≥3 - <5, ≥5 - < 7, and ≥ 7 days) will also be summarized.

5. TREATMENT EXPOSURE

All study drug administration information, including reason not administered, start date and time of administration etc., will be presented in the by-subject data listings based on SAF.

Duration of exposure, total volume administered and compliance (%) will be summarized descriptively in tables for each treatment group using SAF.

Duration of Exposure (days) = Date of Last IMP Administration – Date of First IMP Administration +14.

Compliance to treatment (%) = (Total volume administered(ml)/Total volume prepared(ml))*100 for whole study per subject.

6. PRIOR AND CONCOMITANT MEDICATIONS

Medications/procedures that start and stop prior to the date of first dose of IMP will be classified as Prior (P) only. If a medication/procedure starts before the date of first dose of IMP and stops on or after the date of first dose of IMP, then the medication/procedure will be classified as both Prior and Concomitant (PC). Medications will be classified as Concomitant (C) only if they have a start date on or after the date of first dose of IMP.

If medication/procedure start and/or stop dates are missing or partial, the dates will be compared as far as possible with the date of first dose of IMP. Medications/procedures will be assumed to be Concomitant only, unless there is clear evidence (through comparison of partial dates) to suggest that the medication/procedure started prior to the first dose of study medication. If there is clear evidence to suggest that the medication/procedure started prior to the first dose of IMP, the medication/procedure will be assumed to be both Prior and Concomitant, unless there is clear evidence to suggest that the medication/procedure stopped prior to the first dose of IMP. If there is clear evidence to suggest that the medication/procedure stopped prior to the first dose of IMP, the medication/procedure will be assumed to be Prior only.

Prior and concomitant medications will be coded according to the World Health Organization Drug Dictionary (WHO-DD) SEP2021 B3 or a later version.

Number and percentage of subjects with prior or concomitant medications will be summarized separately by anatomical therapeutic chemical (ATC) level 3, preferred name (PN) treatment group and overall using PAS. Summaries will be ordered alphabetically for ATC3, and then PN within ATC3.

Prior and concomitant procedures will be coded using the MedDRA 24.0 or a later version. Summaries of prior or concomitant procedures will be ordered alphabetically for SOC, and then PT within SOC.

Summary tables on prior medications/procedures will include “P” only, summary tables on concomitant medications/procedures will include “C” and “PC”.

By-subject listings of medications and procedures will be provided, respectively.

7. EFFICACY

7.1 Analysis and Data Conventions

The null hypothesis is that there is no difference between each HMI-115 treatment group and placebo in mean change of DYS measured by NRS from Baseline to Week 12. The alternative hypothesis will be that there is a difference.

All efficacy analyses will be conducted using the PAS. PPS and FAS will be the supplements for partial efficacy analyses. Each comparison will be carried out at the two-sided 5% level of statistical significance unless otherwise specified.

7.1.1 Multi-center Studies

For the summaries and analyses, the term 'Center' will be used to define each investigator site. This study will be conducted by multiple investigators in China, USA and European countries. Multi-center will be pooled by region.

For the primary efficacy variable, the statistical model to be used for the estimation of treatment effects will adjust for differences between regions, by including region as a main effect term in the model. Adjustment for region will also be performed, wherever possible, in the analysis of the secondary efficacy variables.

The homogeneity of the treatment effect across centers will be investigated. A descriptive summary of primary efficacy variable by treatment group and center will be provided. The center information will be presented in the listings.

7.1.2 Adjustments for Covariates

The primary efficacy analysis will be adjusted for the following covariates:

- Region (stratification variable): China, the USA and Europe
- Baseline measure of DYS measured by NRS

7.1.3 Handling of Dropout or Missing Data

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

Multiple Imputation (MI): Missing mean change in DYS measured by NRS will be imputed using multiple imputation approach under the assumption of missing-at-random (MAR). Details are described in section 7.2.2.

Mixed-Effect Model Repeat Measurement (MMRM): The repeated measures analyses will be conducted using mixed models including measurements of interest at select visits. The mixed model includes baseline efficacy value as continuous covariate, the categorical fixed effects of treatment, the randomization stratification factor of region, visit and treatment-by-visit interaction, and subject as a random effect unless otherwise specified. An unstructured variance covariance matrix will be used, if the model does not converge, variance components will be used. Parameter estimation is based on the assumption of data being missing at random and using the method of restrictive maximum likelihood (REML). The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. Missing data will not be imputed for MMRM analysis.

Last Observation Carried Forward (LOCF): Missing post-baseline data can be imputed using the LOCF. For any missing data in follow-up period, only the available observation in follow-up period can be carried forward.

7.1.4 Multiple Comparisons/Multiplicity

Multiplicity adjustment will not be employed with respect to the multiple dose groups in this proof-of-concept study.

7.1.5 Interim Analyses

In addition to the ongoing safety data monitoring, two unblinded interim analysis (IA) of the efficacy data are planned after 60 subjects or 100 subjects completed Week 12 visits separately, or other IA requested by the sponsor. The objective of the interim analysis is to provide preliminary data, and no adjustment of significance level will be employed. The statistical hypotheses tests for the final analysis will be performed at the 0.05 level as described in Section 9.1. This interim analysis will be performed by an analysis team that is independent from the study team responsible for the study conduct and the final analysis. The TFLs performed for IA will be specified in a separate listing.

7.1.6 Examination of Subgroups

The uniformity of the treatment effect for the primary efficacy variable and NMPP will be examined for the following subgroup:

- Asian vs. Others (combined from the randomization strata of region: China for Asian, the USA and Europe for Others)

Summary tables will be provided by subgroup category. The treatment effects across the subgroups defined will be estimated for the mean change and percentage mean change of DYS and NMPP measured by NRS from Baseline to Week 12 on the PAS and using a similar approach as applied to the analysis for the primary efficacy endpoint. Since the randomization strata factor is used to define one of subgroups, the corresponding strata factor will not be included in the model.

The adjusted estimates of treatment mean differences (each HMI-115 treatment group vs. placebo) with SEs and 95% CIs will be provided as appropriate across the subgroups.

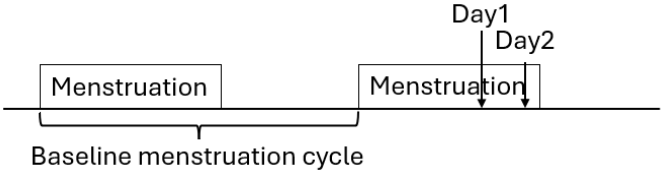
Summaries of the primary efficacy variable by treatment group and subgroups will be produced. No formal statistical analysis will be performed within subgroup.

7.2 Primary Efficacy Endpoint

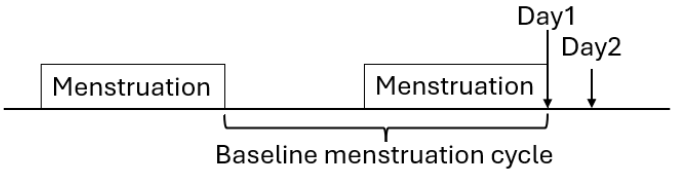
7.2.1 Definition of Endpoint

The primary efficacy endpoint is the mean change of DYS measured by NRS from Baseline to Week 12 measurement. The NRS for overall endometriosis-associated pain ranges from 0 (no pain) to 10 (severe pain). Higher scores mean a worse outcome.

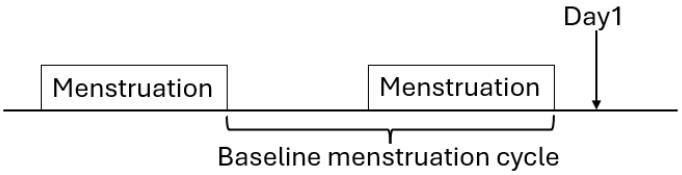
- The Baseline measurement will be derived from the last complete menstruation cycle (menstruation period should be $\geqslant$ 2 days) during the Screening Period prior to Day 1. It will be designated as the baseline menstrual cycle and is defined as following:
    - If the first administration of IMP occurs within the menstruation but not the last day of the menstruation, the baseline menstrual cycle is defined as the interval that first day of the previous menstruation to the day before the menstruation during which the first administration of IMP occurs.



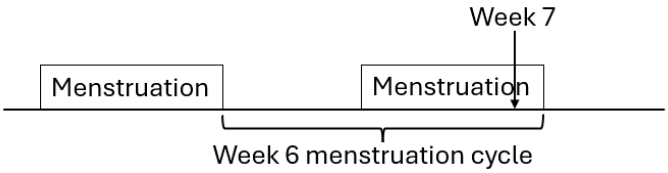
    - If the first administration of IMP occurs on the last day of the menstruation, the baseline menstrual cycle is defined as the interval that begins from the first day after the end of the previous menstruation to the first dosing day.



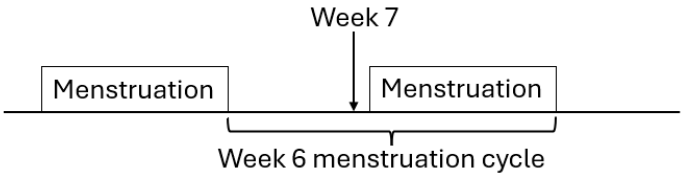
- If the first administration of IMP does not occur during the menstruation, the baseline menstrual cycle is defined as the last menstruation period before the first dose plus the non-menstrual period prior to the last menstruation.



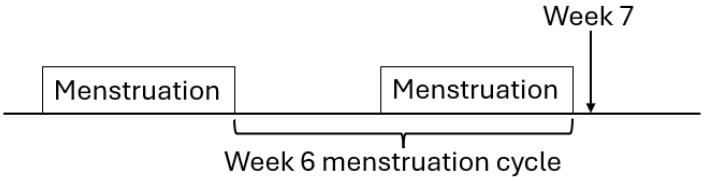
- The Week 6 measurement will be from a complete menstrual cycle (menstruation period should be ≥ 2 days) prior to Week 7.
  - If Week 7 visit occurs during menstruation, the menstrual cycle is defined as the interval that begins from the first day after the end of the previous menstruation to the last day of the menstruation during Week 7 visit occurred. It is expected that the end of this menstruation may be days after the Week 7 visit.



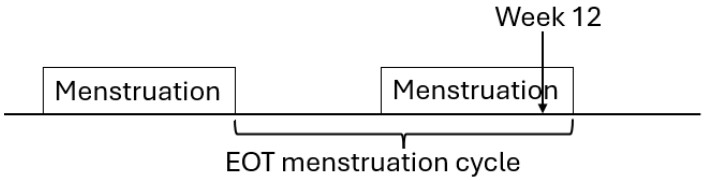
  - If Week 7 visit occurs outside of a menstruation period, the closest menstruation period to Week 7 visit will be identified. This menstrual period may be before or after the Week 7 visit. In either case, the menstrual cycle will begin at the end of the menstruation period immediately before the closest menstruation period to the last day of the closest menstruation period.



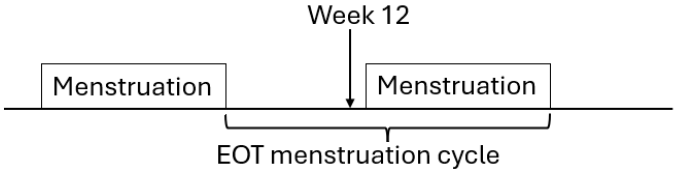
Or



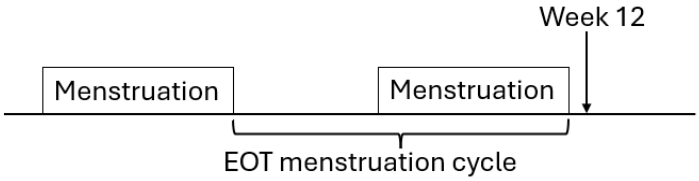
- The Week 12 measurement: Because the Week 12 (EOT) visit may or may not occur during menstruation, the measurement of Week 12 will be defined by the following 2 approaches, depending on the scenarios. The objective of these approaches is to capture a complete menstrual cycle (menstruation period should be ≥ 2 days) and designate it as the EOT menstrual cycle.
  - If EOT visit occurs during menstruation, the EOT menstrual cycle is defined as the interval that begins from the first day after the end of the previous menstruation to the last day of the menstruation during which the EOT visit occurred. It is expected that the end of this menstruation may be days after the EOT visit.



- If EOT visit occurs outside of a menstruation period, the closest menstruation period to the EOT visit will be identified. This menstrual period may be before or after the EOT visit. In either case, the EOT menstrual cycle will begin at the end of the menstruation period immediately before the closest menstruation period to the last day of the closest menstruation period.



Or



The average score is calculated by adding up the daily DYS measured by NRS during analysis window and dividing by the actual number of days with non-missing daily DYS data.

The **primary estimand** is described by the following attributes:

Treatment conditions of interest:	HMI-115 at 60,120 and 240 mg vs. placebo. Placebo treatment is the control group.
Subject population:	Women with moderate to severe endometriosis-associated pain and have surgical diagnosis
Endpoint:	Mean change (change from baseline) in DYS measured by NRS at Week 12 measurement
Population-level summary:	Differences between each HMI-115 treatment group compared with placebo in mean change (change from baseline) in DYS measured by NRS at Week 12, including least square mean (LSM) with its associated standard error (SE) and 95% CIs by using ANCOVA
Intercurrent events (ICEs) and strategies to handle ICEs:	Premature withdrawal from study treatment due to any reasons (Hypothetical strategy)
	At least 15% increase in average pill count of rescue analgesic medications(endometriosis-associated) compared to Baseline. (Composite strategy)

Daily use of rescue analgesic medications will be recorded by the subject in the e-Diary from Visit 2 through the Follow-up period. Protocol specific rescue analgesic medication for endometriosis-associated pain will include four equivalent NSAID (non-steroidal anti-inflammatory drugs) choices and three equivalent opioid choices. For the purposes of the

rescue analgesic analysis, the number of pills for each class of rescue analgesic (NSAID or opioid) will be considered equivalent.

Table 7.1 Analgesic Change During Treatment and Follow-Up Period

Analgesic used at Baseline	**Analgesic dose status at End of Time**	**Responder**
None	None	Responder
	Opioid analgesic and/or NSAID is started	Non-Responder
Tier 1 (NSAID)	Dose stopped, decreases, or is stable (Dose is the same as the screening dose or increases by less than 15% of the screening dose)	Responder
	Dose increases by 15% or more	Non-Responder
	Opioid analgesic is substituted or added	Non-Responder
Tier 2 (Opioid analgesic)	Dose stopped, decreases, or is stable**	Responder
	Dose stopped and NSAID substituted (any dose)	Responder
	Dose decreases and NSAID added (any dose)	Responder
	Dose stable** and NSAID added (any dose)	Non-Responder
	Dose increases by 15% or more	Non-Responder
Tier 1 + Tier 2 (NSAID + opioid analgesic)	NSAID dose stops + opioid analgesic use stops, decreases, or is stable**	Responder
	NSAID use stops + opioid analgesic dose increases by 15% or more	Non-Responder
	NSAID dose decreases + opioid analgesic use stops, decreases, or is stable**	Responder
	NSAID dose decreases + opioid analgesic dose increases by 15% or more	Non-Responder
	NSAID dose stable** + opioid analgesic use stops, decreases, or is stable**	Responder
	NSAID dose stable** + opioid analgesic dose increases by 15% or more	Non-Responder
	NSAID dose increases by 15% or more + opioid analgesic use stops	Responder
	NSAID dose increases by 15% or more + opioid analgesic dose decreases	Responder
	NSAID dose increases by 15% or more + opioid analgesic dose is stable**	Non-Responder
	NSAID dose increases by 15% or more + opioid analgesic dose increases by 15% or more	Non-Responder

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

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

The average pill count for any rescue analgesic use will be calculated by adding up the number of pills (NSAID and/or opioids) at a particular visit time frame and dividing by the actual number of days in the window with non-missing data from the corresponding menstruation cycle.

The pill count will be calculated as follows:

- Tier 1:
  NSAID: 1 pill 200mg Naproxen（220mg Naproxen sodium） = 1 pill 200mg Ibuprofen = 1 pill 25mg Diclofenac = 4 pills 50mg Celecoxib = 1 pill of NSAID use;
- Tier 2:
  Opioids: the pill count for hydrocodone + acetaminophen, codeine phosphate + acetaminophen and tramadol + acetaminophen will be collected from eCOA directly.

First, rescue analgesic use should be categorized as Tier 1 or Tier 2. Then calculating the pill count in each Tier.

If the strength is not the same as the above strength, the pills count can be calculated by (actual strength per pill ÷ standard strength above) *number of pills taken in the last 24 hours for Tier 1.

Tier 1 and Tier 2 are not convertible to each other.

7.2.2 Main Analytical Approach

The following approach (details are provided further down) is to be applied:

1. For subjects who **received at least 15% increase** in average pill count of rescue analgesic medications (**categorized as Non-Responder** per Table 7.1) OR upgraded from Tier 1 to Tier 2 or Tier 1+Tier 2 compared to Baseline measurement cycle, the change from baseline (CFB) in means of DYS measured by NRS at Week 6 measurement and/or Week 12 measurement will be handled per the following two scenarios:
   a. If **CFB ⩾ 0**, the CFB will be kept as it is.
   b. If **CFB < 0**, the CFB will be **set to Zero**.
2. **Any missing** mean changes in DYS measured by NRS at Week 6 and/or Week 12 are **multiple imputed using a MAR approach**.
3. Any (observed or imputed) mean changes in DYS measured by NRS at Week 6 measurement and/or Week 12 measurement **after premature withdrawal** from study treatment (one of ICEs) is **set to missing and then multiple imputed** using a **missing-not-at-random (MNAR) approach**.

Multiple Imputation (MI) Approach

In the MI approach, the mean changes in DYS measured by NRS are multiple imputed: plausible values obtained from a prediction model applying first a MAR approach, followed by a specific MNAR approach applied for hypothetical strategy for handling the specific ICE.

As a monotone missing pattern cannot be expected, the fully conditional specification (FCS) method is used for dealing with arbitrary non-monotone missing data patterns. FCS is based on an iterative algorithm; at each iteration and for each variable of the prediction model, there is a prediction step and an imputation step. When missing values for the last variable in the sequence have been imputed, the algorithm cycles again through each variable, repeating the chain of regression estimation and imputation draw steps.

The full MI approach comprises of 4 steps:

- Prediction Step: current (iteration) values of the observed and imputed values are used to derive the predictive distribution for missing values under MAR, under MNAR mechanism also a function of the selected subset of observations (premature withdrawal from study treatment).
- Imputation Step: updated imputations are generated by draws from the predictive distribution defined by the updated regression model (prediction and imputation are done simultaneously by SAS Proc MI).
- Analysis Step: analysis based on analysis of covariance models (ANCOVA) model (obtaining a point estimate and its standard error for mean change difference in DYS measured by NRS at Week 12) is conducted for each of the multiple imputed datasets by using SAS Proc GLM.
- Combining Step: SAS Proc MIANALYZE is used for combining the point estimates and standard errors obtained in the analysis step from the multiple imputed datasets, in order to provide a single point estimate and confidence interval.

The seed used by in the predication and imputation steps is 264156.

Prediction and Imputation Steps Assuming **MAR**:

The prediction step assuming MAR attempts to predict the mean changes in DYS measured by NRS based on observed data.

The prediction model will be a regression model with the following variables: treatment group, randomization strata of region , baseline DYS measured by NRS, and mean changes in DYS measured by NRS at Week 6 and Week 12. The algorithm cycles are repeated 200 times (“burn-in” iterations) and finally there are 25 draws from the predictive distribution for missing values. SAS Proc MI is used for this process.

The following SAS code for MI based on FCS method will be used:

```
PROC MI data=data_in out=data_out1 nimpute=25 Seed=264156;

CLASS trt01p region;

FCS NBITER=200 REG(chg_dys_w6=trt01p region baseline_dys);

FCS NBITER=200 REG(chg_dys_w12=trt01p region baseline_dys chg_dys_w6);

VAR trt01p region baseline_dys chg_dys_w6 chg_dys_w12;

RUN;
```

Prediction and Imputation Steps Assuming **MNAR** for Premature Treatment Discontinuation:

For the ICE with hypothetical strategy, subsequent to the MAR imputation (described above), mean changes in DYS measured by NRS after such ICE are set to missing and a further multiple imputation is applied based on a specific MNAR mechanism: the now missing mean changes in DYS measured by NRS will be imputed by predictions using only data from subjects who did not experience the ICE with hypothetical strategy. The MNAR prediction model is indicated below.

The following SAS code for MI with MNAR assumption will be used:

```
PROC MI data=data_out1 out=data_out2 nimpute=25 Seed=264156;

BY _imputation_;

CLASS trt01p region;

FCS NBITER=200 REG(chg_dys_w6=trt01p region baseline_dys);

MNAR MODEL(chg_dys_w6/MODELOBS=(ICE2='No'));

FCS NBITER=200 REG(chg_dys_w12=trt01p region baseline_dys chg_dys_w6);
```

MNAR MODEL(chg_dys_w12/MODELOBS=(ICE2='No'));

VAR trt01p region baseline_dys chg_dys_w6 chg_dys_w12;

RUN;

Analysis Step

For each of the 25×25=625 resulting datasets, the **ANCOVA model** with treatment, region as the fixed effects and baseline DYS measured by NRS as covariate will be used to obtain point estimate and its standard error for mean treatment different in DYS measured by NRS at Week 12.

The following SAS code for ANCOVA will be used:

```
PROC GLM data=indata;
  BY _imputation_MAR _imputation_MNAR;
  CLASS trt01p region;
  MODEL cfb = trt01p region baseline;
  LSMEANS trt01p /cl;
run;
```

Combining Step

As a final step, SAS Proc MIANALYZE is used for combining the m=625 point estimates and associated standard errors (from the analysis step above) are combined in order to provide a single point estimate and confidence interval for final statistical inference on the null hypotheses.

A descriptive summary of the average DYS measured by NRS as well as change from baseline by treatment group and visit (measurements of Baseline, Week 6, Week 12, and Week 24) will be presented. The Week 24 measurement is defined in the section 7.3.1.

A by-subject listing of DYS measured by NRS will be provided.

7.2.3 Sensitivity Analyses

In addition to primary analysis, a modified Main Analytical Approach in Section 7.2.2 will be applied for sensitivity analysis.

For Step 1: the two scenarios will be combined as one:

- If **CFB ⩾ 0 or CFB < 0**, the CFB will be **set to Zero**.

Other steps are consistent with the Main Analytical Approach in Section 7.2.2.

7.2.4 Supplementary Analyses

The primary efficacy analysis described in section 7.2.2 will be repeated on the PPS and FAS.

7.2.5 Other Analyses

Percentage Change from Baseline

The percentage change of DYS measured by NRS from Baseline to Week 12 will be performed the same method as analysis of primary outcome in section 7.2.2.

Percentage change from Baseline = (the mean of Visit - the mean of Baseline) / the mean of Baseline)*100%.

The Visit refers to Week 12 measurement or other post-baseline visits.

E-diary Compliance

All subjects will be administered an e-Diary after adequate training at Visit 2. Subjects are required to record uterine bleeding and endometriosis-associated pain every day for the entire duration of the study (until end of study). The use of rescue medications for endometriosis-associated pain will also be recorded. Compliance to e-Diary filling will be summarized for treatment group using PAS including the following variables:

Planned e-Diary Duration (days) = Date of Last Study Visit – Date of Visit 2 +1.

If a subject completed or discontinued the study early, the date of last study visit will be the day of EOS. If a subject is still ongoing, the date of last visit will be replaced by the date of last e-Diary record or last visit date (if applicable), whichever comes last.

- Total days with e-Diary (days) is the total days during study with e-Diary data. All days with non-missing e-Diary regardless of correction or refilling will be included.
- Compliance to e-Diary (%) = (Total Days with e-Diary / Planned e-Diary Duration) × 100.

Based on above calculation formula of e-Diary compliance, another compliance (%) will be calculated to exclude those days with missing and corrected/refilled e-Diary from today days with e-Diary (days).

E-diary data will be listed on the PAS.

Receiver Operating Curve (ROC) Analysis for Determination of Thresholds for Reduction in Pain

The ROC analysis will contain all subjects who received at least one dose of IMP.

- Subjects who a) prematurely discontinue before Week 13, or b) either prematurely discontinue at or after Week 13 but are missing Week 13 PGIC will be assigned a PGIC response of “No change” (PGIC Non-responder).
- Observed values of DYS measured by NRS without any imputation will be used for ROC analysis.

The threshold for reduction in pain will be determined based on a ROC analysis using the PGIC assessment at Week 13 as an anchor and mean change of DYS measured by NRS from Baseline to Week 12 as the independent variable. The PGIC is a 7-point response scale: "Since I started taking the IMP, my endometriosis related pain has: very much improved (1), much improved (2), minimally improved (3), not changed (4), minimally worse (5), much worse (6), very much worse (7)." **The responses of "very much improved (1)" and "much improved (2)" on PGIC will be used to define PGIC responders**, and the threshold for response will be chosen to balance sensitivity and specificity. The threshold with maximum Youden’s index (Youden’s index=sensitivity + specificity – 1) will be selected as optimal threshold. Treatment group will be considered in ROC analysis. No other covariates will be included in the ROC analysis.

Empirical Cumulative Distribution Function (eCDF) Curve

In addition to the anchor-based method, the eCDF method will be used to determine the clinically meaningful improvement. The eCDF curves display a continuous view of the mean score change (both positive and negative) in DYS at Week 12 on the horizontal axis, with the vertical axis representing the cumulative proportion of patients experiencing up to that level of score change. An eCDF curve should be plotted for each distinct anchor category and treatment group.

Dose-Response Relationship

To study the dose-response relationship, the mean CFB to Week 12 measurement will be plotted versus nominal and log-transformed dosage (placebo, HMI-115 60mg, 120mg and 240mg). The dosage (0, 60, 120, 240) will be log-transformed using the function log (dosage+1), where log represents the natural logarithm function. A simple linear regression model will be fitted as a function of dose to characterize the dose-response relationship. In case of lack-of-fit

of the simple linear regression model, multivariate or non-linear regression models may be explored.

The following SAS code will be used:

```
PROC REG data=data;
  MODEL cfb = log_dose / clb;
RUN;
PROC REG data=data;
  MODEL cfb = dose / clb;
RUN;
```

7.3 Secondary Efficacy Endpoint

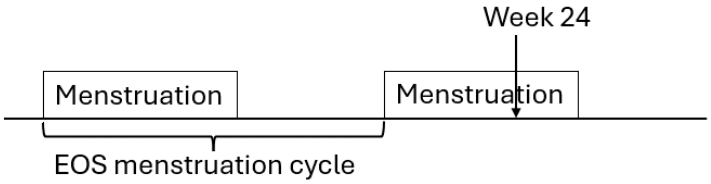
The analysis methods for the secondary endpoints are summarized in the table below.

Secondary Endpoints	Analysis Methods	Missing Data Handling
Change of NMPP measured by NRS from Baseline to Week 12 and 24	Following same steps in Section 7.2.2 for the primary endpoint	
	ANCOVA (as used for the primary endpoint)	MI approach (as used for the primary endpoint)
CFB in the monthly mean Endometriosis Daily Impact Pain Scale for DYS, NMPP and DYSP at Week 12 and 24	Following same steps in Section 7.2.2 for the primary endpoint	
	ANCOVA (as used for the primary endpoint)	MI approach (as used for the primary endpoint)
Change of DYS measured by NRS from Baseline to Week 24	Following same steps in Section 7.2.2 for the primary endpoint	
	ANCOVA (as used for the primary endpoint)	MI approach (as used for the primary endpoint)
Change of DYSP measured NRS from Baseline to Week 12 and 24	Following same steps in Section 7.2.2 for the primary endpoint	
	ANCOVA (as used for the primary endpoint)	MI approach (as used for the primary endpoint)
CFB by visit in permitted rescue medication use	The number of medication and the percentage of days that permitted rescue medication is used will be summarized as two separate endpoints	Missing data will not be imputed
	Analysis of variance (ANOVA)	
Change in menstrual period heaviness (bleeding) from Baseline by visit	The changes of mean scores of heaviness (bleeding) in the menstrual cycle collected from eCOA.	Missing data will not be imputed

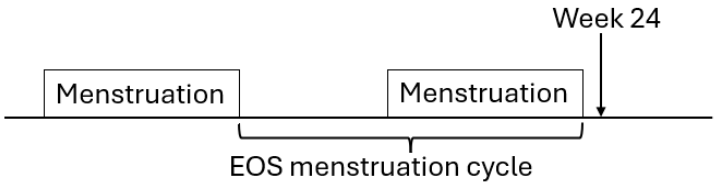
7.3.1 CFB in DYS Measured by NRS at Week 24

The Measurement of Week 24: because the Week 24 (EOS) visit may or may not occur during menstruation, the measurement of Week 24 will be defined by the following 2 approaches, depending on the scenarios. The objective of these approaches is to capture a complete menstrual cycle and designate it as the EOS menstrual cycle.

- If EOS visit occurs during menstruation, the EOS menstrual cycle is defined as the interval that begins from the first day of the previous menstruation to the day before the first day of the menstruation during which the EOS visit occurred.



- If EOS visit occurs outside of a menstruation period, the latest menstruation period prior to the EOS visit will be identified. The EOS menstrual cycle will begin at the end of the menstruation period immediately before the latest menstruation period to the end of the latest menstruation period.



The main analytical approach in Section 7.2.2 will be applied for this analysis.

In addition, sensitivity analysis for primary endpoint defined in section 7.2.3 will be applied for this analysis as well.

7.3.2 CFB in NMPP Measured by NRS at Week 12 and Week 24

The Baseline, Week 12 and Week 24 measurements will follow the definitions in sections of 7.2.1 and 7.3.1 for the time window, respectively.

The mean CFB and %CFB in NMPP at Week 12 and Week 24 will be summarized for each treatment group with point estimates (least square mean [LSM]), the associated standard errors [SEs]) and 95% CIs. Estimated treatment differences (compared for each HMI-115 treatment group against the placebo group) along with corresponding 95% CIs will be presented using the ANCOVA model with treatment group, region as the fixed effects and baseline NMPP as a continuous covariate.

The main analytical approach in Section 7.2.2 will be applied for this analysis.

In addition, sensitivity analysis for primary endpoint defined in section 7.2.3 will be applied for this secondary endpoint as well.

NMPP measured by NRS will be summarized by treatment and visit using observed values.

A by-subject listing will be presented for NMPP.

7.3.3 CFB in DYSP Measured by NRS at Week 12 and Week 24

For DYSP Measured by NRS, the Week 12 and Week 24 measurements will follow the definitions in sections of 7.2.1 and 7.3.1 for the time window, respectively.

The main analytical approach in Section 7.2.2 will be applied for this analysis.

In addition, sensitivity analysis for primary endpoint defined in section 7.2.3 will be applied for this secondary endpoint as well.

The mean CFB and %CFB in DYSP measured by NRS will be summarized by treatment and visit using observed values.

A by-subject listing will be presented for DYSP.

7.3.4 CFB in the Monthly (every menstruation cycle) Mean Endometriosis Daily Impact Pain Scale for DYS, NMPP and DYSP at Week 12 and Week 24

Endometriosis Daily Impact Pain Scale will be used to record the mean change every month from Baseline to Week 12 measurement and Week 24 measurement. For DYS, NMPP and DYSP, responses of "None," "Mild," "Moderate," and "Severe" will be assigned a score of 0, 1, 2, and 3, respectively. For DYSP, the choice of "Not Applicable" applies to subjects not sexually active for reasons other than endometriosis pain. Responses of "Not Applicable" for the assessment of DYSP will be treated as missing values and will not be included in the summaries/analyses.

The average pain scores for DYS, NMPP and DYSP for Week 12 and Week 24 will follow the definitions in sections of 7.2.1 and 7.3.1 for the time window.

The main analytical approach in Section 7.2.2 will be applied for this analysis.

In addition, sensitivity analysis for primary endpoint defined in section 7.2.3 will be applied for this secondary endpoint as well.

Endometriosis Daily Impact Pain Scale for DYS, NMPP and DYSP will be listed.

7.3.5 CFB by Visit in Permitted Rescue Medication Use

Permitted rescue medication will be provided to subjects from Visit 2 for rescue use if needed for endometriosis associated pain throughout the study. Use of analgesic medication will be reported by subjects in the e-Diary on a daily basis from Visit 2 through the Follow-up period.

The average pill count of permitted rescue medications along with CFB will be summarized by visit measurement including Baseline, Week 6, Week 12 and Week 24 (following the definitions in sections of 7.2.1 and 7.3.1 for the time window). The percentage of days with analgesic rescue use which will be calculated based on those days in the visit window with non-missing data will be summarized by visit.

In addition, the pill count of rescue medications at Week 12 separated by Tier 1 and Tier 2 (following the definitions in sections of 7.2.1 for the time window) as the dependent variable, treatment comparison may be compared using an ANCOVA model with treatment and region as the main effects and baseline of pill count as a continuous covariate.

The correlation between the use of rescue medications (in terms of pill count) and DYS and NMPP measured by NRS at Week 12 measurement will be explored by summary statistics including Pearson and Spearman correlation coefficients and graphical presentation (e.g., scatter plot).

Daily use of permitted rescue medications will be listed.

7.3.6 Change in Menstrual Period Heaviness from Baseline by Visit

Descriptive statistics will be presented for the changes of days of menstrual period and menstrual cycle, and the changes of mean scores of heaviness (bleeding) collected from eCOA at Week 6, Week12 and Week 24 measurements.

Scores of heaviness (bleeding) for each subject will be listed.

8. SAFETY

All safety summaries and analyses will be based upon SAF and is as same as PAS to focus on the same population as primary efficacy data analysis. All safety analyses will be based on observed data. Unless otherwise specified, missing data will not be imputed.

8.1 Adverse Events

All AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) 24.0 or a later version.

Treatment Emergent Adverse Events (TEAEs) will be tabulated and are defined as those AEs with a start date on or after the first administration of IMP.

Where dates are missing or partially missing, AEs will be assumed to be treatment-emergent, unless there is clear evidence (through comparison of partial dates) to suggest that the AE started prior to the first administration of IMP. For AEs with completely unknown dates, dates will be shown as NK (unknown) in the listings. For AEs with partially start and/or end dates, the available day, month and year components will be shown in listing.

Frequencies and percentages of subjects with TEAEs will be calculated for each treatment group as follows:

- Overview incidence of all AEs
- A summary of the number and percentage of subjects reporting a TEAE by treatment, SOC, and PT
- A summary of the number and percentage of subjects reporting a TEAE by treatment, causality, SOC, and PT
- A summary of the number and percentage of subjects reporting a TEAE by treatment, severity, SOC, and PT
- A summary of the number and percentage of subjects reporting a AESI by treatment, SOC, and PT
- A summary of the number and percentage of subjects reporting a AESI by treatment, causality, SOC, and PT

Adverse event summaries will be ordered in terms of decreasing frequency for SOC, and PT within SOC, in the overall treatment groups, and then alphabetically for SOC, and PT within SOC. The actual version of the MedDRA coding dictionary used will be noted in the AE tables.

For each subject and each adverse event, the worst severity recorded will be attributed and used in the by-severity summaries. Similarly, the worst causality (most related to treatment) will be attributed and used in the by-causality summaries. If severity or causality is missing, a conservative approach for AE assessment (taking into account the worst case) will be followed.

Below listings will be provided:

- A by-subject listing of AEs
- A by-subject listing of AESIs

8.2 Deaths, Serious Adverse Events, and Other Significant Adverse Events

Below summaries will be provided:

- A summary of the number and percentage of subjects reporting a TEAE resulting in death by treatment, SOC and PT
- A summary of the number and percentage of subjects reporting a TEAE resulting in IMP discontinuation by treatment, SOC and PT

- A summary of the number and percentage of subjects reporting a SAE by treatment, SOC and PT
- A summary of the number and percentage of subjects reporting a SAE resulting in IMP discontinuation by treatment, SOC and PT

Listings to be provided include:

- A by-subject listing of AEs leading to death
- A by-subject listing of TEAEs leading to IMP discontinuation
- A by-subject listing of treatment emergent SAEs

8.3 Clinical Laboratory Evaluation

The following laboratory parameters will be measured for each test:

Laboratory Assessments	Parameters		
Hematology	• Platelet count • Hemoglobin • WBC Count (absolute) • Erythrocytes • Reticulocytes	• Hematocrit • MCV • MCH	**WBC Differential:** • Neutrophils • Lymphocytes • Eosinophils • Monocytes • Basophils
Coagulation	• PT • Fibrinogen	• INR	• APTT
Clinical Chemistry#	• BUN • Creatinine • Glucose, fasting • Sodium • Total protein • Magnesium	• Potassium • Chloride • Calcium • Albumin • GGT • Total bilirubin	• AST (SGOT) • ALT (SGPT) • Alkaline phosphatase
Urinalysis	• Protein • Glucose • Urobilinogen	• Leucocytes • Erythrocytes • Bilirubin	• pH • Ketone • Nitrite • Specific gravity
Sex Hormone	• PRL* • DHEAS*	• LH* • FSH*	• Estradiol* • Progesterone* • Testosterone*
Additional Chemistry#	• FT3, FT4, TSH	• CA-125	• GH
Viral serology	• Anti-HAV • HBs Ag • HBc Ab	• Anti-HCV • HCV RNA (for positive or equivocal anti-HCV)	• HIV antigen • Anti-HIV (1+2)
Bone Biomarkers# (exploratory)	• CTX	• P1NP	• BSAP
Other tests	• eGFR (using CKD EPI formula) • Pregnancy test (serum pregnancy test at Visit 1 only, urine pregnancy test thereafter) • Urine drug screen: amphetamine, barbiturates, benzodiazepines, tetrahydrocannabinol, cocaine, methadone, methamphetamine, opiates, phencyclidine, and tricyclic antidepressants		

Laboratory Assessments	Parameters
	• Testing for COVID-19 infection (Only for subjects enrolled before protocol version 5.0): o SARS-CoV-2 RT-PCR (to be assessed at Screening Visit 1 and prior to randomization. Additional testing may be done depending on Investigator discretion and/or local requirements)
ALT = alanine transaminase; anti-HAV = hepatitis A virus antibody; anti-HCV = hepatitis C virus antibody; anti-HIV = human immunodeficiency virus antibodies; APTT = activated partial thromboplastin time; AST = aspartate transaminase; BSAP = bone specific alkaline phosphatase; BUN = blood urea nitrogen; CKD EPI = Chronic Kidney Disease Epidemiology Collaboration CA-125 = cancer antigen-125; COVID-19 = coronavirus disease 2019;CTX = C-terminal telopeptide of type I collagen; FSH = follicle-stimulating hormone; FT3= Free triiodothyronine; FT4 = Free thyroxine; eGFR = estimated glomerular filtration rate; GGT = gamma-glutamyl transferase; GH = growth hormone; HBsAg = hepatitis B virus surface antigen; HBcAb = Hepatitis B core antibody; DHEAS = Dehydroepiandrosterone-sulfate; INR = international normalized ratio; LH = luteinizing hormone; MCH = mean corpuscular hemoglobin; MCV = mean corpuscular volume; P1NP = procollagen type 1N-terminal propertied; PRL = Prolactin; PT = prothrombin time; RT-PCR: Real-time Reverse Transcriptase Polymerase Chain Reaction; SGOT = serum glutamic-oxaloacetic transaminase; SGPT = Serum glutamic pyruvic transaminase; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; TSH = thyroid stimulating hormone; WBC = white blood cells. * Sex hormone samples will be collected at the menses cycle days 1 to 5. **Subjects should not take any opioids 2 days prior to blood sampling for PRL**. # Samples will be collected under morning fasting conditions.	

All summaries of laboratory parameters will only summarize parameters planned based on the protocol; however, both planned and unplanned laboratory parameters will be provided in by-subject listings.

Clinical laboratory test results will be summarized descriptively for each treatment group at each visit. The within-group CFBs to each visit will be summarized and will be compared with a paired t-test for continuous measurements. The between-group difference will be compared with an ANOVA. Two-sided 95% CI will be constructed for the differences between each of the treatment groups and the placebo group in CFB.

In addition, prolactin results from sex hormone will be summarized separately for each treatment group by category the results as normal, 1 ULN to 2 ULN, 2 ULN to 3 ULN, 3 ULN to 4 ULN, >4 ULN and <1 LLN.

Shift tables for CFBs according to appropriate categories will be provided for each laboratory parameter.

By-subject listings of laboratory assessments including hematology, coagulation, clinical chemistry, urinalysis, sex hormone, additional chemistry, bone biomarker, and other tests (eGFR) will be listed. All values outside the clinical reference ranges will be flagged in the data listings. The abnormal values will be flagged with 'L' for values below the lower limit of the clinical reference range and 'H' for values above the upper limit of the clinical reference range and included in the listings.

Serum pregnancy test at Visit 1 only, and urine pregnancy test at Visits 3, 5, 7, 9 and 10. Subject with pregnancy test results will be listed.

Subjects with urine drug screen test during study will be listed.

Subjects with serology result during study will be listed.

Subjects with COVID-19 testing result will be listed.

8.4 Vital Signs, Electrocardiograms, Physical Examination, Transvaginal Ultrasound, Dual Energy X-ray Absorptiometry, and Others

8.4.1 Vital Signs

Vital signs will be summarized descriptively for each treatment group at each visit. The within group CFBs will be summarized and will be compared with a paired t-test by visit; the between-group difference for each visit will be compared with an ANOVA, and two-sided 95% CI for the differences between each of the treatment groups and the placebo group will be presented.

A by-subject listing of vital signs will be provided.

8.4.2 Physical Examination

Physical examination abnormal findings will be summarized by treatment group and body system.

A by-subject listing of abnormal physical examination findings will be provided.

8.4.3 12-Lead ECG

ECG parameters, i.e., HR, PR, QT, QRS, QTcF, and QTcB will be summarized descriptively for each treatment group at each visit. The within-group CFBs will be summarized and will be compared with a paired t-test by visit; the between-group difference for each visit will be compared with an ANOVA, and two-sided 95% CI for the differences between each of the treatment groups and the placebo group will be presented.

The number and percentage of subjects with abnormal, clinically significant (CS) finding will also be summarized by treatment group at each visit.

ECG data will be listed by subject and visit including changes from baseline.

A summary of the number and percentage of subjects with QTcF intervals exceeding some predefined upper limits (e.g., >450ms, >480ms, >500ms for measured values as well as. >30ms, >60ms for CFBs) will be displayed in a frequency table.

8.4.4 TVU Assessments

TVU data will be summarized descriptively for each treatment group at each visit. CFBs in the number and/or size of ovarian cysts, changes in endometrial thickness will be summarized.

A by-subject listing of TVU findings will be provided.

8.4.5 DXA Scan

Bone mineral density data from DXA scan will be summarized descriptively for each treatment group at each visit. The within-group %CFB will be summarized and will be compared with a paired t-test by visit; the between-group difference for each visit will be compared with an ANOVA, and two-sided 95% CI for the differences between each of the treatment groups and the placebo group will be presented. Data from post treatment Month 6/Month 12 (before Protocol version 5.0) will not be analyzed for clinical study report (CSR).

In addition, DXA scan will be summarized for each treatment group by category the %CFB as $\geqslant$ -3%, -3% to -5%, -5% to -8% and <-8%.

A by-subject listing of DXA scan data will be provided.

8.4.6 Mammogram

Mammogram is required for females aged ≥40 only in the US and Poland sites. A by-subject listing of mammogram results will be provided.

8.4.7 Cervical Cytology Test

Cervical cytology will be assessed at any time during the Screening period. Collected data will be listed.

8.4.8 Sellar MRI with Contrast

Sellar MRI will be assessed at any time during the Screening period. Collected data will be listed.

8.5 Safety Monitoring (Independent Data Monitoring Committee [IDMC])

In order to monitor and protect the safety of study subjects, an IDMC will be established to review all study data pertaining to subjects' safety on regular basis to adjudicate the overall safety of the subjects in this study. The IDMC will convene to provide recommendations on study design and IMP based on review of unblinded safety data. The IDMC will consist of unblinded experts in the fields of gynecology and one independent unblind biostatistician.

In addition, the IDMC will be notified, and a meeting may be called for each case of IMP discontinuation due to an IMP related safety reason.

Study will be paused pending prompt review by IDMC if 2 or more subjects experience similar SAE or severe AE assessed as related to IMP or discontinue IMP for safety reasons assessed as related to IMP.

Study will be paused if any of the following assessed as related to IMP occurs, pending prompt review of the case(s) by the IDMC: death, suicide attempt, serious changes in behavior (such as psychosis or the development of impulse control disorders or addictive behavior), cardiac arrhythmia or AEs pertaining to the pituitary gland.

Refer to the DMC Charter for further details.

9. OTHER ANALYSES

9.1 Pharmacokinetics

9.1.1 Pharmacokinetic Concentrations

PK concentration data for HMI-115 will be listed by treatment and subject. Listings will include nominal PK sampling time, actual sampling times relative to dose administration, deviation from nominal time, and concentration. Concentrations below the lower limit of quantification (LLOQ) will be presented as below the limit of quantification (BLQ) in the listings. Any PK data excluded from PK analysis set will be identified.

Separate concentration for HMI-115 summary tables will be included for intensive and sparse PKS. Concentrations will be summarized by treatment group and nominal timepoint. The following descriptive statistics will be presented for concentrations obtained at each nominal time point: N (number of subjects who received treatment), n (number of subjects with non-missing values), and n (BLQ) (the number of subjects with BLQ samples), arithmetic mean, SD, coefficient of variation (CV%), geometric mean (Geomean), geometric CV% (calculated as: gCV% =SQRT(es²-1)*100; where s is the standard deviation of the log-transformed values), median, minimum and maximum values.

For summary tables, all BLQs will be considered zero, Summary Statistics will not be calculated if non-BLQ concentrations at a scheduled time point is < 3 and will be reported as “not calculable (i.e., NC)”.

The following rules will be followed with regards to the number of decimal places and presentation of data in the tables and listings of concentration data:

Concentration Listings and Tables	Rounding
Individual concentrations	*n* s.f. as supplied by bioanalytical laboratory No rounding
Minimum and Maximum	*n* s.f. capped at 4
Mean/SD/Median/Geomean	*n+1* s.f. capped at 4
CV%/gCV%	*1* d.p.
N/n	Whole number

s.f. = significant figures, d.p.= decimal place

For arithmetic mean graphs, separate figures will be created for extensive and sparse PKS. For linear/linear graphs all BLQ values will be substituted with zero for calculation of arithmetic mean and for log/linear graphs the log transformed arithmetic mean will be displayed (this should not include zero).

For individual linear/linear and log/linear graphs all BLQ values will be substituted as follows:

- BLQs at the beginning of a participant profile (i.e., before the first incidence of a measurable concentration) will be assigned to zero. When using log/linear scale, these timepoints will be considered missing.
- BLQs at the end of a participant profile (i.e., after the last incidence of a measurable concentration) will be set to missing.
- Single BLQs which fall between two measurable concentrations will be set to missing.
- Consecutive BLQs which fall between measurable concentrations will be set to missing. Measurable concentrations after consecutive BLQs will be set to missing.

To visualize the comparison between each treatment, the following descriptive PK graphs will be generated for intensive and sparse PKS.

- Figure 1: Individual Profile for HMI-115 Plasma Concentration Time Data – Linear Scale

- Figure 2: Individual Profiles for HMI-115 Plasma Concentration Time Data – Semi Logarithmic Scale
- Figure 3: Overlaid Individual Profiles for HMI-115 Plasma Concentration Time Data by Treatment – Linear Scale (Spaghetti Plots, only intensive PKS)
- Figure 4: Overlaid Individual Profiles for HMI-115 Plasma Concentration Time Data by Treatment – Semi-Logarithmic Scale (Spaghetti Plots, only intensive PKS)
- Figure 5: Mean (± SD) HMI-115 Plasma Concentration Time Data by Treatment and by PKS Subset (intensive and sparse) – Linear scale (including a separate plot for subgroup – Asian vs. Others)
- Figure 6: Mean HMI-115 Plasma Concentration Time Data by Treatment and by PKS Subset (intensive and sparse) – Semi-Logarithmic Scale (including a separate plot for subgroup – Asian vs. Others)
- Figure 7: Boxplot of C_{trough} Time data by Treatment and by PKS Subset (intensive and sparse) – Linear Scale (including a separate plot for subgroup – Asian vs. Others)

9.1.2 Pharmacokinetic Parameters

PK parameters will be calculated for the intensive PKS by non-compartmental analysis methods from the concentration-time data following these guidelines:

- Actual sampling times relative to dosing rather than nominal times will be used in the calculation of all derived PK parameters.
- There will be no imputation of missing data.
- Handling of BLQ samples for derivation of PK parameters
    - Day 1, all BLQ values pre-dose and in the absorption phase prior to the first quantifiable concentration will be substituted by zeros.
    - For subsequent dosing days, BLQ values pre-dose, in the absorption phase, and between evaluable concentrations will be substituted by missing values before the calculation of the PK variables.
    - BLQs at the end of a participant profile (i.e., after the last incidence of a measurable concentration) will be set to missing.

Pharmacokinetic parameters will be estimated according to the following guidelines:

Parameter	Guideline for Derivation
C_{max}, t_{max}, $C_{trough,}$	Obtained directly from the observed concentration-time data
$AUC_{0\text{-}10d}$, $AUC_{0\text{-}inf,}$ $AUC_{(0\text{-}tau)}$	The $AUC_{0\text{-}10d}$, $AUC_{0\text{-}inf,}$ AUC over the dosing interval after the last administered dose will be calculated by a combination of linear and logarithmic trapezoidal methods. Unless specifically requested and justified, the linear up/log down trapezoidal method will be employed.
%AUCextrap	Percentage of AUC_{inf}（obs, pred）due to extrapolation from t_{last} to infinity: %AUCextrap=$\frac{AUC_{inf} - AUC_{last}}{AUC_{inf}} * 100$
λz , t½	1.The apparent terminal phase rate-constant (λz) will be estimated by linear regression of concentration versus time data presented in a log-linear scale. 2.Data are primarily monotonically decreasing in magnitude and are representative of the actual decline in the log concentration-time curve. 3.Only those data points that are judged to describe the terminal log-linear decline will be used in the regression. 4.A minimum number of three data points in the terminal phase will be used in calculating λz with the line of regression starting at any post-C_{max} data point (C_{max} should not be part of the regression slope). Unless otherwise determined by PK Scientist's best knowledge and judgment, if the adjusted correlation coefficient (R^2 adjusted) is <0.8, it will be excluded from the summary tables and statistical analysis of PK parameters, and λz and all the λz dependent parameters (i.e., t½) will also be flagged and excluded from summary tables and statistical analysis. The reason for exclusion will be listed/footnoted in parameter listings. 5.Unless otherwise determined by PK Scientist's best knowledge and judgment, the interval used to determine λz should be equal or greater than 1.5-fold the estimated t½, and if less than 1.5-fold, λz will be flagged in listings. All the derived parameters (i.e., t½) will also be flagged from listings of PK parameters. The reason for flagging will be listed/footnoted in parameter listings. 6.The t½ will be calculated as follows: t½ = ln2/λz = 0.693/ λz 7.Data points may be dropped from the linear regression if the PK Scientist considers the reported values to be anomalous. Any data points so designated should remain in the listings with a footnote and be identified in the study report with a rationale for exclusion.
MRT	MRT is Mean residence time from the time of dosing to the time of the last measurable concentration. For non-infusion models: $MRT=\frac{AUMClast}{AUClast}$
CL/F	Total body clearance for extravascular administration, calculated as follows： $CL/F=\frac{dose}{AUC_{0-inf}}$（non-steady state）or $CL/F=\frac{dose}{AUC_{0-tau}}$
V_Z/F	V_Z/F is Volume of distribution based on the terminal phase. For non-steady state: $V_Z/F=\frac{dose}{\lambda_z * AUC_{0-inf}}$ For steady state: $V_Z/F=\frac{dose}{\lambda_z * AUC_{0-tau}}$

Parameter	Guideline for Derivation
R_{ac}	Accumulation ratio, calculated from comparison of first dose and last dose: Based on C_{max}: $R_{ac}=\frac{C_{max}(last\ dose)}{C_{max}(first\ dose)}$ Based on AUC: $R_{ac}=\frac{AUC_{(0-tau)}(last\ dose)}{AUC_{(0-tau)}(first\ dose)}$

PK parameters will be listed by subject for the PKS. PK parameters that will be excluded from summary tables and statistical analyses of PK parameters will be flagged and footnoted with the reason for exclusion.

PK parameters will be summarized by dose cohort for the extensive PKS. Summaries for PK parameters will report N (number of subjects who received treatment) and n (number of subjects with non-missing values). Descriptive statistics for calculated PK parameters will include N, n, arithmetic mean, SD, CV%, Geomean, gCV%, median, minimum, and maximum values. For tmax, only N, n, median, minimum and maximum values will be presented. No descriptive statistics will be determined when fewer than three individual PK parameters are available.

The rules followed for presentation of PK parameters data with regards to the number of decimal places/significant digits for the listings of participant level PK parameters and summary tables of PK parameters are as follows:

PK Parameter Listings and Tables	Rounding
Derived Individual parameters	4 s.f.
Directly Derived Individual parameters (C_{max}, C_{trough})	n s.f. as supplied by the WinNonlin but not more than 4 s.f.
Minimum and Maximum	4 s.f.
Mean/SD/Median/Geomean	4 s.f.
CV%/gCV%	1 d.p.
N/n	Whole number
Exceptions for PK Tables	
t_{max} individuals and min/max	2 d.p.
t_{max} median only	2 d.p.

s.f. = significant figures, d.p.= decimal place

9.1.3 PK-PD Relationship Analysis

If appropriate, an exploratory analysis of the relationship between HMI-115 plasma concentration and pelvic pain score measured by NRS (cyclic, non-cyclic, and average) will be performed. The C_{trough} data will be represented graphically against pelvic pain scores (absolute and relative (expressed as %) change from baseline) and values of Prolactin by treatment.

9.2 Immunogenicity

Anti-drug antibodies will be measured using a three-tiered assay approach: screening assay, confirmatory assay and titration assay. Based on results using confirmatory assay, incidence of ADA will be summarized by visit on the immunogenicity analysis set. A shift table from baseline (positive/negative) to each scheduled post-baseline visit will be provided as well.

Samples confirmed as positive within the confirmatory assay, will be further evaluated in a neutralizing assay (NADA), so the by-visit summary of NADA (positive/negative) will also be provided on the Immunogenicity analysis set in all subjects with at least one ADA positive result. Percentage will be calculated for subjects with positive ADA at that visit.

All ADA and NADA data will be listed on the immunogenicity analysis set.

9.3 PGIC

The number and percentage of subjects in each PGIC response category (7-point scale) will be summarized at each visit including Week 5, Week 9, Week 13 and Week 25 for treatment group. The percentage of subjects with response of “Much Improved” or “Very Much Improved” will be combined and labeled as “Much Improved or Better”. The remaining five categories of the PGIC scale will be combined and labeled "Other". The number and percentage of subjects with response of a) "Much Improved or Better" and b) "Other" will be summarized at each visit by treatment group, and comparison will be made between each HMI-115 treatment group and placebo group using a chi-square test, if the counts of equal or more than 20% cells are less than 5, the comparison will use Fisher exact test. In addition, the scores of PGIC will be summarized as continuous data to calculate means of each treatment group.

Score	Response category	Pooled category
1	Very much improved (better outcome)	Much Improved or Better (PGIC Responder)
2	Much improved	
3	Minimally improved	
4	No change	
5	Minimally worse	Other (PGIC Non-responder)
6	Much worse	
7	Very much worse (worse outcome)	

Statistical comparisons between each HMI-115 treatment group and placebo group in score (ranging 1 to 7) at each visit will be performed using ANOVA with treatment group and the randomization stratification factor of region as the fixed effects. If a subject has at least 15% increase of rescue medication use or changed to Opioids/Opioids+NSAIDs from NSAIDs at a particular visit (referring to Appendices Table 12.2) compared to baseline, such subjects will be considered as non-responders (‘No change’ category which the value will be set as 4) in the PGIC responder analysis for that visit.

PGIC data will be listed by subject and visit.

9.4 EHP-5

The EHP-5 is a valid instrument to measure health-related quality of life in endometriosis. The EHP-5 consists of two parts:

- The core instruments have five scale scores covering:
  - Pain (Found it difficult to walk because of the pain?),
  - Control and powerlessness (Felt as though your symptoms are ruling your life?),
  - Emotional well-being (Had mood swings?),
  - Lack of social support (Felt others do not understand what you are going through?),
  - Self-image (Felt your appearance has been affected?)
- A modular questionnaire consisting of six additional questions:
  - Work (Been unable to carry out duties at work because of the pain?),
  - Relationship with child/children (Found it difficult to look after your child/children?),
  - Sexual intercourse (Felt worried about having intercourse because of the pain?),

- Feelings about the medical profession (Felt the doctor(s) think it is all in your mind?),
- Feelings about treatment (Felt frustrated because treatment is not working?),
- Worries about infertility (Felt depressed at the possibility of not having children/more children?).

Each item is rated on a 4-point scale (never = 0, rarely = 1, sometimes = 2, often = 3, always = 4). Scores on the EHP-5 core and modular questionnaire then are standardized on a scale of 0 (indicating best health status) to 100 (indicating worst health status). If the 'Not Relevant' box was ticked for items on modular questionnaire the score could not be computed for that dimension. After standardization, each item will be scored as (never = 0, rarely = 25, sometimes = 50, often = 75, always = 100).

EHP-5 data (absolute standardized score and CFB) will be summarized descriptively for treatment group at each visit including Baseline (collected on Day 1 before 1st dose of IMP), Week 5, Week 9, Week 13 and Week 25. A negative CFB score indicates improvement in quality of life. CFB will be compared using the MMRM and ANCOVA methods as used for the primary endpoint for each item individually. Missing data will be imputed using LOCF for ANCOVA model.

EHP-5 data will be listed by subject and visit.

9.5 Composite Pelvic Signs and Symptoms Score (CPSSS)

The CPSSS consists of 5 components that address dysmenorrhea (pain during menstruation), dyspareunia (painful intercourse), non-menstrual pelvic pain, pelvic tenderness, and pelvic induration (hardening). Each component was scored on a scale of 0 to 3 (0 = none, 1 = mild, 2 = moderate, and 3 = severe). For each component, subjects will be classified as responders if they reported a 1 point or greater reduction (improvement) from baseline. For DYSP, the choice of "Not Applicable" applies to subjects not sexually active for reasons other than endometriosis pain. Responses of "Not Applicable" for the assessment of DYSP will be treated as missing values and will not be included in the summaries/analyses.

CPSSS score raw data will be summarized descriptively by treatment group and visit (Baseline and Week 13/EOT). And for each component, the percentages of responders at each visit will be presented for each treatment group.

Total pelvic pain score is the sum score of dysmenorrhea, dyspareunia and non-menstrual pelvic pain, and the maximum possible value will be 9. The total physical sign score is the sum of pelvic tenderness and induration, and the maximum possible value will be 6. Total symptom and sign severity score is the sum score of all the 5 components, and the maximum possible value will be 15 (a lower score indicates less signs and symptoms of endometriosis or better functioning). For the 3 total scores, each will be summarized with descriptive statistics.

CPSSS data will be listed by subject and visit.

10. CHANGES TO THE PLANNED ANALYSIS

Surgically diagnosed subjects are this study's target subjects, and it complies with the request of the FDA and the global protocol to include subjects with surgical diagnosis only in the primary analysis, so the primary analysis population is changed from the FAS (subjects with randomized and taking at least one dose of IMP) to the subjects who have surgical diagnosis in the FAS. Subjects population with non-surgical diagnoses will be used as supplementary analysis. Compared to the Protocol version 6.1, the below corresponding updates have been done in the SAP version 1.3:

1) The primary objective is updated to study the efficacy of HMI-115 for 12 weeks compared with placebo in the management of endometriosis-associated pain based on subjects with surgical diagnosis.
2) A new analysis set - PAS is added in Section 3.3, which will be the primary analysis set for all efficacy analyses instead of FAS. While FAS will NOT be the primary analysis set. In addition, PPS will be a subset of the PAS instead of FAS. Both FAS and PPS will be the supplement for partial efficacy analyses.
3) For the subgroup analysis as defined in Section 7.1.6, the subgroup – “Surgical diagnosis vs. Non-surgical diagnosis” is removed.
4) The TFL Shells are updated accordingly following the SAP V1.3.

11. SOFTWARE

[1] SAS® Version 9.4 of the SAS System for Personal Computers. Copyright © 2002-2003. SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

[2] Phoenix®WinNonlin® Software Version 8.2.

12. APPENDICES

12.1 Visit Name

Visits in the TFLs will be presented as follows:

Table 12.1: Visit Name for Scheduled Visit

Study Period	Visit Long Name	Visit Short Name
Screening	Visit 1 Screening	Screening V1
	Visit 2 Screening	Screening V2
Treatment Period	Visit 3 (Week 1)	Week 1
	Visit 4 (Week 3)	Week 3
	Visit 5 (Week 5)	Week 5
	Visit 6 (Week 7)	Week 7
	Visit 7 (Week 9)	Week 9
	Visit 8 (Week 11)	Week 11
	End of Treatment (Week 13)	Week 13/EOT
Follow-up Period	Follow-up Visit (Week 25)	Week 25/FU
For data summary, unscheduled visits will not be mapped to next scheduled visit while EDV will be mapped to scheduled visit as defined in section 2.1.		

12.2 Analysis Visit Window for Daily Assessment

Table 12.2: Analysis Visit Window for Scheduled Visit

		Time Interval (Study Day)	
Study Analysis Visit	Target Study Day	Start	End
Screening V1			
Screening V2			-1
Baseline		-35	-1
Week 1	Day 1	1	8
Week 3	Day 15	9	22
Week 5	Day 29	23	36
Week 7	Day 43	37	50
Week 9	Day 57	51	64
Week 11	Day 71	65	78
Week 13/EOT	Day 85	79	127
Week 25/FU	Day 169	≥128	

EOT=end of treatment; FU=follow-up.