STATISTICAL ANALYSIS PLAN

Protocol Title:

A 12-Week, Randomized, Double-blind, Placebo-controlled, Parallel-group Phase 3 Study to Evaluate the Efficacy and Safety of HP-5000 Topical System (Patch) in Subjects with Osteoarthritis Pain of the Knee

Date of SAP:

15-Feb-2023

NCT Number:

NCT04683627

Appendix 16.1.9 Documentation of Statistical Methods

HP-5000-US-07 - Statistical Analysis Plan - Version 1.0 – 28 September 2022

DocuSign Envelope ID: D2BB5456-E458-4E7C-8650-F63E715880F5

STATISTICAL ANALYSIS PLAN

A 12-Week, Randomized, Double-blind, Placebo-controlled, Parallel-group Phase 3 Study to Evaluate the Efficacy and Safety of HP-5000 Topical System (Patch) in Subjects with Osteoarthritis Pain of the Knee

AUTHOR: [redacted]

VERSION NUMBER AND DATE: V1.0, 28 SEPTEMBER 2022

Copyright © 2009, 2010, 2012, 2016, 2018, 2019 [redacted]. All rights reserved. The contents of this document are confidential and proprietary to [redacted] Holdings Inc. and its subsidiaries. Unauthorized use, disclosure or reproduction is strictly prohibited.

DocuSign Envelope ID: D2BB5456-E458-4E7C-8650-F63E715880F5

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

Statistical Analysis Plan V1.0 (Dated 28 September 2022) for Protocol [Final]HP-5000-US-07 protocol_amendment#1_Jan7_Cleaned.

	Name	Signature	Date
Author:			30-Sep-2022
Position:	Biostatistics Team Lead		
Company:			

Upon review of this document, the undersigned approves this version of the Statistical Analysis Plan, authorizing that the content is acceptable for the reporting of this study.

	Name	Signature	Date
Approved By:			29-Sep-2022
Position:	Statistical Scientist		
Company:			
Approved By:			30-Sep-2022
Position:	Senior Director, Biostatistics – Product Development		
Company:	Noven Pharmaceuticals, Inc.		

DocuSign Envelope ID: D2BB5456-E458-4E7C-8650-F63E715880F5

Approved By:	[redacted]	29-Sep-2022
Position:	Senior Director Clinical Development	
Company:	Noven Pharmaceuticals, Inc.	

MODIFICATION HISTORY

Unique Identifier for this Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
V1.0	28Sep2022	[redacted]	Not Applicable – First Authorized Version

TABLE OF CONTENTS

DocuSign Envelope ID: D2BB5456-E458-4E7C-8650-F63E715880F5

DocuSign Envelope ID: D2BB5456-E458-4E7C-8650-F63E715880F5

DocuSign Envelope ID: D2BB5456-E458-4E7C-8650-F63E715880F5

LIST OF ABBREVIATIONS

AE	Adverse event
APAP	Acetaminophen
ATC	A–Anatomical, T-Therapeutical/ Pharmacological and C-Chemical WHODrug dictionary class
BMI	Body mass index
CI	Confidence interval
ECG	Electrocardiogram
eCRF	Electronic case report form
eDiary	Electronic diary
HR	Heart rate
ICF	Informed consent form
IEC	Independent Ethics Committee
IRB	Institutional Review Board
ITT	Intent-to-Treat
LOCF	Last observation carried forward
LOE	Lack of efficacy
LS	Least squares
MAR	Missing-at-random
MI	Multiple imputation
MMRM	Mixed Model Repeated Measure
MNAR	Missing-not-at-random
Mini-OAKHQOL	Mini-OA Knee and Hip QOL Questionnaire
NRS	Numeric rating scale
OA	Osteoarthritis

PR	PR interval of the electrocardiogram; time duration between the P and R waves
PT	Preferred term
PGIC	Patient Global Impression of Change
PGA	Patient Global Assessment
QOL	Quality of life
QRS	QRS interval of the electrocardiogram; duration of the QRS complex
QT	QT interval of ECG, duration between the Q and T waves
OTC	Over the counter
QTcB	QT interval of ECG corrected for heart rate using Bazett's formula
QTcF	QT interval of ECG corrected for heart rate using Fridericia's formula
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SAS®	Statistical Analysis Software
SD	Standard deviation
SMQs	Standard MedDRA Queries
SOC	System organ class
TEAE	Treatment-emergent adverse event
TLF	Table, listing, and figure
WHO	World Health Organization
WOMAC	West Ontario and McMasters Universities Osteoarthritis Index
WPAI-OA	Work Productivity and Activity Impairment Questionnaire for Osteoarthritis

1. INTRODUCTION

This statistical analysis plan (SAP) is based on protocol version 2.0, dated 07 January 2021.

2. STUDY OBJECTIVES

2.1 PRIMARY EFFICACY OBJECTIVE

The primary objective of this study is to evaluate the effect of HP-5000 on pain among subjects with Osteoarthritis (OA) of the knee at Week 12.

Primary Efficacy Endpoint for this objective is:

- West Ontario and McMasters Universities Osteoarthritis Index (WOMAC) LK3.1 OA (pain) change from baseline at Week 12

2.2 KEY SECONDARY EFFICACY OBJECTIVES

The key secondary objectives of this study are:

- To evaluate the effect of HP-5000 on physical function among subjects with OA of the knee at Week 12
- To evaluate the effect of HP-5000 on stiffness among subjects with OA of the knee at Week 12

Key Secondary Efficacy endpoints for these objectives are:

- WOMAC LK3.1 OA (physical function) change from baseline at Week 12
- WOMAC LK3.1 OA (stiffness) change from baseline at Week 12

2.2.1 OTHER SECONDARY EFFICACY OBJECTIVES

- To evaluate the effect of HP-5000 on pain among subjects with OA of the knee at Weeks 1, 2, 4, and 8
- To evaluate the effect of HP-5000 on physical function among subjects with OA of the knee at Weeks 1, 2, 4, and 8

- To evaluate the effect of HP-5000 on stiffness among subjects with OA of the knee at Weeks 1, 2, 4, and 8
- To evaluate the effect of HP-5000 on pain intensity as measured by numeric rating scale (NRS) scores on a daily diary
- To evaluate the onset of effect of HP-5000 on pain among subjects with OA of the knee
- To evaluate the pattern of use of rescue medication in the study
- To evaluate skin irritation, discomfort, and adhesion following administration of HP-5000
- To evaluate the safety and tolerability of HP-5000
- To evaluate the effect of HP-5000 treatment on Quality of Life (QOL)

Other Secondary Efficacy Endpoints for these objectives are:

- WOMAC LK3.1 OA (pain) change from baseline at Weeks 1, 2, 4, and 8
- WOMAC LK3.1 OA (physical function) change from baseline at Weeks 1, 2, 4, and 8
- WOMAC LK3.1 OA (stiffness) change from baseline at Weeks 1, 2, 4, and 8
- WOMAC LK3.1 OA (composite score) change from baseline at Weeks 1, 2, 4, 8, and 12
- Change from baseline in pain intensity assessed on an 11-point NRS at Weeks 1, 2, 4, 8, and 12
- Change from baseline in pain intensity assessed on an 11-point NRS of a weekly average of all available daily pain scores at each week from Week 1 through 12
- Change from baseline in Patient Global Assessment (PGA) at Weeks 1, 2, 4, 8, and 12
- Patient Global Impression of Change (PGIC) at Weeks 1, 2, 4, 8, and 12
- Treatment Response: Reduction in the WOMAC LK3.1 OA Index (pain) of ≥30%, ≥50%, ≥70%, and ≥90% at Weeks 1, 2, 4, 8, and 12
- Treatment Response: Reduction in the WOMAC LK3.1 OA Index (physical function) of ≥30%, ≥50%, ≥70%, and ≥90% at Weeks 1, 2, 4, 8, and 12
- Treatment Response: Reduction in the WOMAC LK3.1 OA Index (stiffness) of ≥30%, ≥50%, ≥70%, and ≥90% at Weeks 1, 2, 4, 8, and 12
- Rescue Medication Use during Treatment Phase: Proportion of subjects who used, number of days used, and total number of tablets

- Onset of the effect: Change from baseline pain intensity assessed on an 11-point NRS
- Change from baseline in Mini-OA Knee and Hip QOL (Mini-OAKHQOL) scores at Weeks 4 and 12
- Change from baseline in Work Productivity and Activity Impairment Questionnaire for Osteoarthritis (WPAI: OA) impairment scores at Weeks 4 and 12

2.3 Safety Objective

The safety objective is to evaluate the safety and tolerability of HP-5000.

Safety Endpoints associated with this objective are:

- The incidence of treatment-emergent adverse events (TEAEs), severity of TEAEs, adverse events (AEs) leading to discontinuation of the study drug, serious adverse events (SAEs), and other significant AEs
- Change from baseline in clinical laboratory tests (SI units), 12-lead electrocardiogram (ECG) findings, body weight, physical examination findings, and vital signs
- Dermal assessments: irritation, discomfort, and adhesion

2.4 Estimands

The estimand for the primary efficacy analysis is the difference between the ITT population treatment groups (HP-5000 *vs*. placebo) in change from baseline WOMAC pain scores at Week 12 taking into account intercurrent events using the composite strategy specified in the section 16.1.5.

Discontinuation of study treatment due to reasons related to study treatment (e.g. AEs, lack of efficacy (LOE), etc.), rescue medication use within 24 hours of WOMAC pain assessment, and removal/detachment of the patch for more than 24 hours before WOMAC pain assessment are defined as intercurrent events.

3. Study Design

3.1 General Description

This is a multicenter, randomized, double-blind, placebo-controlled Phase 3 study in the US

evaluating the efficacy and safety of HP-5000 in subjects with OA pain of the knee.

The study will consist of up to a 28-day Screening Phase that will include a Washout Period of current prescription and over the counter (OTC) analgesics; a 12-week Double-blind Treatment Phase, and a 1-week safety Follow-up Phase.

Screening Phase: Subjects will be seen in the clinic where the study will be described to them and will be asked to sign the informed consent form (ICF). No screening procedures may begin prior to obtaining informed consent. After obtaining a written informed consent, subjects will be screened for participation in the study. The target knee for the purpose of the study will be defined as the one which causes the subject more pain than the other knee and will be identified at the Screening visit using the NRS score. During the Screening Phase, subjects will complete a daily electronic diary (eDiary) to record pain severity of their target knee. Subjects who satisfy the entry criteria will be instructed to begin the Washout Period for 7 to 14 days, or at least 5 half-lives, whichever is longer, prior to the Baseline Visit (Day 0).

During the Washout Period, current prescription and OTC analgesics will be discontinued. However, rescue medication (acetaminophen [APAP]) for the treatment of any other aches the subject might experience during the trial, such as headache, reduction of fever, and the target knee pain of OA (not for the non-target knee) will be permitted (a maximum of 2 g/day) except within 3 calendar days prior to the Baseline Visit (i.e., rescue medication is prohibited for 2 days until clinic visit on Day 0 and on the day of Baseline Visit [V3]). During the Washout Period, subjects will complete a daily eDiary to record rescue medication usage in addition to pain severity of their target knee. Following completion of the Washout Period, subjects will return to the clinic for their Baseline Visit (Day 0). Eligible subjects will be randomized to either HP-5000 or placebo in a 1:1 ratio.

Treatment Arm	[redacted] **Treatment**
HP-5000	HP-5000 [redacted] patch Each patch contains [redacted] diclofenac sodium.
Placebo	HP-5000 [redacted] Placebo patch Each patch contains 0% [redacted] diclofenac sodium.

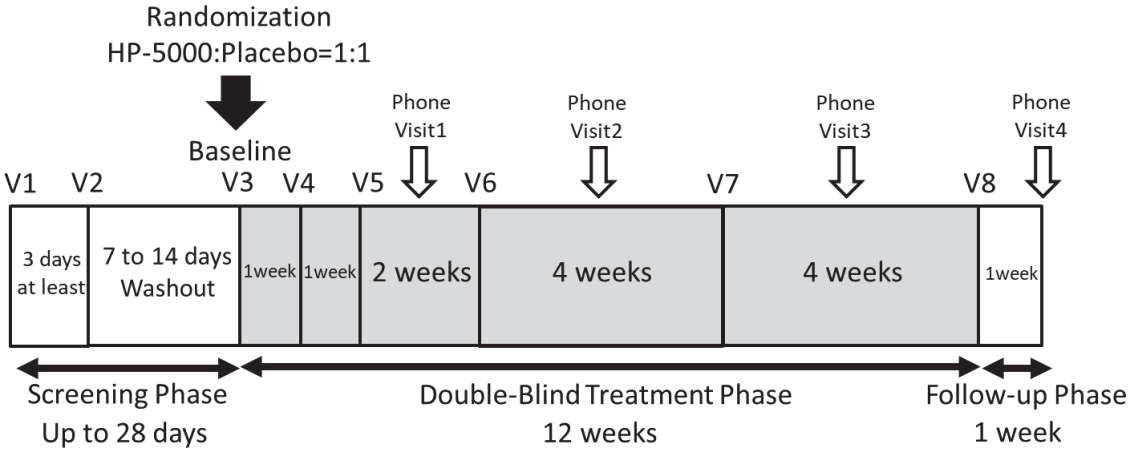
Double-blind Treatment Phase: Subjects will apply a single patch to the target knee [redacted] After the removal of the patch [redacted], a new patch will be applied to a different site on the target knee. Subjects will be instructed to rotate the patch application site [redacted] to alternate sides (inner and outer) of the target knee. During the Double-blind Treatment Phase, subjects will also complete a daily eDiary to record patch application and removal time, adhesion, irritation, discomfort, and pain assessments; and

amount of rescue medication (APAP) taken daily, if applicable. A maximum of 2 g/day of APAP will be allowed as rescue medication except from Baseline Visit until Visit 4, within 2 calendar days prior to all subsequent visits (V5, V6, V7, and V8), and on the day of the clinic visits (V4, V5, V6, V7, and V8).

NOTE: Subjects are NOT allowed to apply HP-5000 on the non-target knee at any time during the study.

Follow-up Phase: Subjects will have a safety visit approximately 7 days after the Week 12 End-of-Treatment Visit. The investigator or qualified study staff will contact all subjects by phone for continued safety monitoring of AEs. At the discretion of the investigator, subjects may have to return to the clinic for their Follow-up Visit.

Figure 1: A Schematic of the Study Design



Sample Size: Based on the HP-5000 Phase 2 study results, the assumed effect size of the test product in the pivotal study will be about [redacted]. The sample size of [redacted] subjects per treatment arm (or total of about [redacted] subjects) will be included in order to have 80% power to demonstrate statistically significant difference versus placebo using a 2-sided overall Type I Error rate of 0.05.

3.2 Schedule Of Events

The schedule of events can be found in Section 10.2 of the protocol.

3.3 CHANGES TO ANALYSIS FROM PROTOCOL

Not Applicable

4. PLANNED ANALYSES

4.1 DATA SAFETY MONITORING BOARD

There will be no Data Safety Monitoring Board for this study.

4.2 INTERIM ANALYSIS

No interim analyses are planned.

4.3 FINAL ANALYSIS

The analyses detailed in this SAP will be performed by [redacted] Biostatistics following sponsor authorization of this SAP, Database Lock, Sponsor authorization of Analysis Sets and Unblinding of Treatment.

5. ANALYSIS SETS

Analysis of efficacy and safety endpoints will be performed based on the analysis sets defined in this section. The inclusion/exclusion of subjects from each analysis set will be determined prior to the final analyses (and approved by Noven) based on blinded data review.

5.1 INTENT-TO-TREAT SET[ITT]

Intent-to-treat (ITT) analysis set includes all consented and randomized subjects. Regardless of any protocol deviations and/or treatment compliance, analyses performed on the ITT set will be based on the randomized treatment assignment and all available data. The ITT set will be used as the primary set for analysis of primary and secondary efficacy endpoints. If a subject is randomized incorrectly or is administered the incorrect study drug, analyses of the ITT set will be based on the assigned treatment whereas all analyses of the safety analysis set will be based on the actual treatment.

5.2 Safety Analysis Set [SAF]

Safety analysis set (SAF) includes all subjects who have had at least 1 patch of double-blind study drug applied and who have at least 1 post-dose safety measurement during the Double-blind Treatment Phase. In the unlikely event that errors may have occurred in treatment arm assignments, then analyses using the SAF set will be based on treatment actually received. The SAF set will be used for the analysis of dermal evaluations and safety endpoints.

6. General Considerations

6.1 Reference Start Date and Study Day

Relative study day will be calculated from the date of Day 0, which is the day of Baseline Visit (V3) and will be used to show the start and/or stop days of treatment, study procedures, and assessments.

If the date of the treatment, procedure, or event is on or after the Day 0 date then:

Relative Study Day = (date of variable of interest – Day 0 date) + 1.

If the date of the treatment, procedure, or event is prior to the Day 0 date then:

Relative Study Day = (date of variable of interest – Day 0 date).

If the date of the treatment, assessment, or event is partial or missing, relative study day, and any corresponding event durations will appear missing in the listings.

6.2 Baseline and Post Study assessment

Baseline is defined as the last non-missing measurement taken on or prior to the first patch application date. In the case where the last non-missing measurement and the reference start date coincide, that measurement will be considered pre-baseline if the assessment is planned per protocol to take place prior to first study medication administration.

Post-study assessment is defined as a non-missing measurement taken on or after the date of the last patch removal. In the case where the dates coincide, time of the measurement should be after the last patch removal time to be considered post study assessment. This will be applied to identify post study medications and adverse events.

6.3 VISIT WINDOWING CONVENTIONS

Analyses presented by visit or study day will be based on the protocol defined scheduled visits including scheduled, unscheduled, and early discontinuation visits. Visit windows for unscheduled visits and early discontinuation visits are defined in Table A, which provides the mapping of relative day ranges to the scheduled target days and the study periods. If more than one assessment is available in the same 'Range of Relative Study Days' (window), the assessment closest to the Scheduled Target Day will be selected and assigned to that Scheduled Target Day. If two or more assessments are available in the same window and are equidistant from the Scheduled Target Day, the latest assessment will be selected.

All study listings will include scheduled, unscheduled, re-test, and early discontinuation data.

Table A: Mapping of Relative Day Ranges to Schedule Target Day

Study Phase	Study Period	Range of Relative Study Days	Scheduled Target Day
Screening	Screening (Visit 1)	28-14 days before Day 0 datetime	Day -28 to -14
	Washout (Visit 2)	14-7 days before Day 0 datetime	Day -14 to -7
Double Blind Treatment	Baseline (Visit 3)	After first patch application Day 0 datetime	Day 0
	Week 1 (Visit 4)	1 to 10 days relative to Day 0 datetime	Day 7
	Week 2 (Visit 5)	11 to 17 days relative to Day 0 datetime	Day 14
	Week 4 (Visit 6)	18 to 38 days relative to Day 0 datetime	Day 28
	Week 8 (Visit 7)	39 to 73 days relative to Day 0 datetime	Day 56
	Week 12 (Visit 8)	After the last dose of treatment (patch removal) and before the start of the Safety Follow-up Period (one day after actual Day 84 visit) which will be ≥74 days after Day 0 datetime.	Day 84

Follow-up	Safety Follow-up	>84 days after Day 0 datetime for treatment completers (83 days on-treatment and after last patch removal) and after Day 84 post-treatment assessments or after early discontinuation visit (for subjects who terminated early).	Day 91

6.4 Statistical Tests

Unless otherwise indicated, all testing of statistical significance will be 2-sided, and a difference resulting in a p-value of <0.05 will be considered statistically significant.

6.5 Common Calculations

For quantitative measurements, change from baseline will be calculated as:

Value at Visit X – Baseline Value

Percent change from baseline will be calculated as:

$$\frac{\text{Value at Visit X} - \text{Baseline Value}}{\text{Baseline Value}} \times 100.$$

6.6 Software Version

All analyses will be conducted using SAS version 9.4 or higher.

7. Statistical Considerations

7.1 Adjustments for Covariates and Factors To Be Included in The Analysis

For details on the inclusion of the covariates in the models, see the specific analysis sections 16.1, 16.2, and 16.3.

7.2 Multicentre Studies

This study will be conducted by multiple investigators at multiple centers in the United States.

Data from all sites will be pooled and statistical analyses will not be adjusted for investigational site.

7.3 Missing data

The total and subscale scores of any assessment with more than one item, will be derived from individual items. Any individual missing item in any scale will not be imputed. If one or more items are missing at a visit, then the associated total score or subscale score will be set to missing.

Missing WOMAC scores will be imputed and used in sensitivity analyses of the primary and key secondary efficacy as per section 16.1.4 and 16.1.5.

7.4 Multiplicity Issues

For handling multiplicity in this study, a fixed-sequence procedure will be applied for primary and key secondary endpoints, where the order in which the hypotheses are tested is pre-specified as following: (1) WOMAC pain score, (2) WOMAC physical function score, and (3) WOMAC stiffness score. Testing begins with the first hypothesis about WOMAC pain score, and each test is carried out as long as a significant result with a level of significance of α=0.05 is observed in all preceding tests. The fixed-sequence procedure controls the family-wise error rate because, for each hypothesis, testing is conditional upon rejecting all hypotheses earlier in the sequence.

All statistical tests of hypotheses in this analysis will be 2-sided with a Type I Error rate of 0.05.

7.5 Active control Studies Intended To show Non inferiority or Equivalence

Not applicable for the study.

7.6 Examination Of Subgroups

Subgroup analyses will be performed for the primary endpoint on the ITT set. It should be noted that the study is not designed to detect treatment differences within subgroups.

The following subgroups may be assessed and described within the exploratory analysis section; however, categories for sub-grouping will be confirmed further based on empirical data:

- Age: (<65, 65 - <=75, >75), or data driven groups

- Sex: Male/Female
- Race: As per data collected in the CRF
- Body mass index (BMI): (<18.5, 18.5 to <25.0, 25.0 to <30, >= 30.0)
- Severity: (Kellgren-Lawrence grade) of target knee based on X-ray before baseline
- Average and Worst NRS of Target Knee: (5 – 7 = moderate, 8 – 10 = severe) at baseline
- Contralateral Knee Involvement: (Yes/No)

8. OUTPUT PRESENTATIONS

The table, listing, and figure (TLF) shells provided with this SAP describe the presentations for this study and therefore the format and content of the summary TLFs to be provided by [redacted] Biostatistics.

Continuous variables will be summarized using descriptive statistics (number of subjects [n], mean, standard deviation [SD], median, 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 and median 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.

Categorical variables will be summarized using frequency counts and percentages. Unless otherwise stated, the calculation of percentage will be based on the number of subjects in the analysis set of interest.

All p-values will be presented to four decimal places. P-values less than 0.0001 will be presented as “<0.0001”. Unless otherwise indicated, all testing of statistical significance will be 2-sided, and a difference resulting in a p-value of <0.05 will be considered statistically significant. Confidence intervals (CIs) will be presented to two more decimal places than the raw data.

Source data for summary tables and statistical analyses will be presented as subject data listings.

9. DISPOSITION AND WITHDRAWALS

The numbers of subjects who are randomized, completed study treatment, discontinued from study treatment, and withdrew from the study, along with reasons for withdrawal will be tabulated overall and by treatment group for each analysis set. The withdrawal reasons will be presented under three categories: “study treatment related”, “not study treatment related”, and “uncertain”. The specific reasons will be categorized under those three.

The number of subjects in each analysis population will be reported.

9.1 PROTOCOL DEVIATIONS

Protocol deviations are any changes to, deviations, or departures from the study design or procedures that are under the investigators’ control and that have not been reviewed and approved by the Institutional Review Board (IRB) / Independent Ethics Committee (IEC).

Critical deviations are deviations from protocol-related procedures that threaten integrity of data, adversely affect subjects and/or could invalidate acceptability of a study (or part of it). Major deviations are unapproved changes or deviations from protocol-related procedures that could affect integrity, completeness, accuracy, and reliability of the data or adversely affect subject’s rights, safety, or well-being. Minor deviations are deviations from accepted procedures that will not adversely affect subjects or data integrity.

The clinical lead will provide an excel file which includes all the critical, major, and minor deviations for each type along with a flag if the deviation was COVID-19 related to the biometrics team. The statistical programmer will use this file to create an analysis dataset.

Randomized subjects with major and minor protocol deviations will be listed. For the ITT Set, protocol deviations will be summarized by deviation criterion, treatment group, and overall. The deviation criterion has three levels critical, major, and minor. The number of subjects enrolled in error will also be summarized by treatment group and overall. Information indicating if the deviation is COVID-19 related will be listed.

10. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Demographic data and other baseline characteristics will be presented for the SAF Set. No formal statistical testing will be carried out for comparing demographic or other baseline characteristics between treatment groups.

The following demographic characteristics will be presented in summary tables and listings:

Age and age categories (< 65, 65 - ≤75, > 75)
Sex
Race
Ethnicity
Height (cm)
Weight (kg)
BMI (kg/m^2)
BMI by category (kg/m^2) (<18.5, 18.5 to <25.0, 25.0 to <30.0, >= 30.0)
Age at the onset of OA of the knees (years)

The following baseline characteristics will be presented:

Target knee (left or right)
NRS pain average (target knee) by category (mild: < 5, moderate: >=5 to <7, severe ≥ 7)
Is the diagnosis bilateral (Y/N)?
Kellgren-Lawrence grade for the target knee (0-4)
WOMAC pain score
WOMAC stiffness score
WOMAC physical function score
WOMAC composite score
NRS average pain score (target and non-target knee)
NRS worst pain score (target and non-target knee)
NRS average pain score frequency counts
NRS worst pain score frequency counts
WOMAC pain score by category (0 – 4, 5 – 8, 9 – 12, 13 – 16, 17 – 20)

11. MEDICAL HISTORY

Medical History information will be presented for the ITT Set. Medical History will be coded using Medical Dictionary for Regulatory Activities (MedDRA) currently in effect at the time of final database lock.

Any medical condition or clinically significant laboratory abnormality with an onset date after signing the ICF and before the time of dosing is considered to be pre-existing and should be documented as Medical History (and not as an AE). Pre-existing diseases or conditions will NOT be considered AEs unless there is an increase in the frequency or severity, or a change in the

nature of the disease or condition during the Double-blind Treatment Phase.

The number and percentage of subjects with at least one pre-existing medical condition, as collected in the *Medical History* eCRF pages will be summarized by system organ class (SOC), preferred term (PT), treatment group, and overall.

Additionally, similar data will be summarized for prior surgeries, procedures, or therapies as collected in the *Surgeries, Procedure or Therapies* eCRF page, where the reason ticked is "Medical History".

12. CONCOMITANT SURGERIES, PROCEDURES OR THERAPIES

Concomitant surgeries, procedures, or therapies will be presented for the SAF set. Each of these surgeries, procedures, or therapies will be categorized as either medical history that worsened during treatment or an AE per the definitions given in section 11 and section 18.1, respectively.

'Concomitant' surgeries, procedures, or therapies are those that:

- started on or after the date of first dose of study medication, which is the first patch application time, and started prior to the date of last dose of study medication, which is patch removal of past patch applied.
- AND stopped on or after the date of first dose of study medication or were ongoing at the completion of Week 12 post-dose assessments visit (or last treatment day which is removal of patch for subjects who discontinue treatment), or after the last dose of study medication (if planned assessments are not performed).

See Appendix 1 for handling of partial dates, if it is not possible to define a surgery, procedure, or therapy as prior, prior and concomitant, concomitant, or post-treatment, it will be classified by the worst case (i.e., concomitant).

13. PRIOR AND CONCOMITANT MEDICATIONS

Prior medications will be presented for the ITT Set and concomitant medications will be presented for the SAF Set. Medications will be captured on the *Concomitant Medications* eCRF page and coded using the World Health Organization (WHO) Drug dictionary (WHODDE01SEPT2020 or later version).

'Prior' medications are medications which started and stopped prior to the date of first dose of

study medication which is time of first patch application.

'Concomitant' medications are medications that:

- started on or after the date of first dose of study medication, which is the first patch application time, and started prior to the date of last dose of study medication, which is patch removal of past patch applied.
- AND stopped on or after the date of first dose of study medication or were ongoing at the completion of Week 12 post-dose assessments visit (or last treatment day which is removal of patch for subjects who discontinue treatment), or after the last dose of study medication (if planned assessments are not performed).

See Appendix 1 for handling of partial dates for medications, if it is not possible to define a medication as prior, prior and concomitant, concomitant, or post-treatment, the medication will be classified by the worst case (i.e., concomitant).

Prior and concomitant medications will be summarized separately by treatment group, Anatomical Therapeutic Chemical (ATC) Level 2 and Standardized name (preferred term). All prior and concomitant medications will be listed flagging prior medication as "yes/no".

14. STUDY MEDICATION EXPOSURE

Exposure to study medication will be presented for the SAF Set. Investigational product administration for each subject will be summarized in terms of the total number of patches applied by visit, number of patches removed early (due to safety reasons: irritation, discomfort, etc.) by visit, and summary of the reason for early removal.

Each subject's duration of exposure (days) will also be calculated and summarized by treatment arm.

Derivation:

Duration of exposure (days) = latest of (last patch application date and last patch removal date) - date of 1st patch application + 1

Refer to Appendix 4 for more details.

A by-subject listing of all drug exposure information collected will be presented.

The daily amount of rescue medication used and the daily amount of study drug used will be calculated and presented in a by-subject listing. Deviations from the planned doses (overdose, missed dose, or timing) will be recorded on the eCRF and presented in by-subject listing.

15. STUDY MEDICATION TREATMENT COMPLIANCE

Treatment compliance with study drug regimens will be presented via the counting of returned unused patches, patch application, removal, and by questioning the subject, at every post-randomization visit. Treatment compliance will be calculated as number of patches given less the number of patches returned divided by the number of days for required applications, expressed as a percentage. The total number of days between study visit dates will be used for the calculation to determine the number of days for required applications. Unscheduled patch applications will be considered in the calculation.

Number and percentage of subjects who are compliant (defined as using ≥ 80% and ≤ 120% of study medication), noncompliant (defined as using less than 80% or more than 120% of study medication) at each treatment arm will be presented by visit and overall. A summary table will also present overall subject compliance with study drug. A by-subject listing of all drug exposure information collected will be presented.

16. EFFICACY OUTCOMES

16.1 PRIMARY EFFICACY

16.1.1 PRIMARY EFFICACY VARIABLE AND DERIVATION

The primary efficacy variable is the change from baseline to Week 12 in the WOMAC pain score.

This score, is derived by summing the sub-scores for the 5 following items collected in the *Western Ontario and McMaster Universities Osteoarthritis Index* eCRF page taking into account missing observations as described in section 7.3:

Question: How much pain have you had:

[redacted]

[redacted]

Each item is rated from 0 to 4, with 0 being none and 4 being extreme.

Therefore, the WOMAC pain score will either have an integer value from 0 to 20 or a missing value for each visit.

16.1.2 Missing Data Methods for Primary Efficacy Variable

The primary analysis method for evaluating the primary efficacy endpoint is the likelihood-based analysis of repeated measures, a mixed model with repeated measures (MMRM) under an assumption that the missing observations of change from baseline WOMAC pain score are missing at random. To assess the robustness of the MAR assumption, sensitivity analyses which utilize multiple imputations and different assumption about unobserved outcomes will be performed, as detailed in Section 16.1.4.

16.1.3 Primary Analysis of Primary Efficacy Variable

The estimand in the primary analysis for efficacy is the difference between treatment groups (HP-5000 *and* placebo) in the change from baseline to Week 12 in WOMAC pain score. The analysis will be performed using the ITT analysis set taking into account intercurrent events using the composite strategy.[4,5] The study treatment discontinuation due to reasons related to study treatment (e.g. AE, LOE, etc.), rescue medication use within 24 hours of WOMAC pain assessment, removal/detachment of the patch for more than 24 hours before WOMAC pain assessment are defined as intercurrent events. Only the reasons that are in the database will be presented in the table.

The treatment effect on the primary efficacy variable which is change from baseline to Week 12 in WOMAC pain score will be evaluated using an MMRM model for the ITT Set and as well as taking into account following intercurrent events, and composite strategy (see section 16.1.4).

The MMRM model will include change from baseline in WOMAC pain score as the repeated dependent variable, with effects for treatment (HP-5000 and placebo), visit, treatment-by-visit interaction, and the baseline WOMAC pain score as a covariate. Initially an unstructured covariance matrix will be assumed. If convergence is not achieved with this initial structure, additional variance-covariance structures will be tested and the one with the smallest AIC will be chosen. The Kenward-Roger method will be used to estimate the denominator degrees of freedom and correct for bias in the estimated variance-covariance of the fixed effects. This MMRM with the covariance structure that converges will be used to compare the HP-5000

treatment to placebo. The sample SAS code for any MMRM can be found in Appendix 2.

For each visit, the least-squares mean (LSM) estimate, standard error, and 95% CIs will be presented for each treatment group. The least square mean difference of the treatment from placebo estimate, standard error, 95% CI, effect size, and p-value for each visit will also be presented in this table. The effect size is calculated as the difference in the LS mean estimate divided by the pooled estimate of within subject standard deviation obtained from the variance-covariance matrix used in the model. Descriptive statistics for the WOMAC pain score and change from baseline in WOMAC pain score will be presented in a table by visit and treatment group.

Additionally, if the normality assumption is not supported by the data, an analysis of covariance (ANCOVA) on rank-transformed data will be used as utility analysis and additional supportive analyses may be performed if necessary.

16.1.4 Composite strategy analysis for primary efficacy

Composite strategy will be applied to the stratum of subjects experiencing intercurrent events.

For the intercurrent events of study treatment discontinuation due to reasons related to study treatment (AE, LOE, etc.), the obtained WOMAC pain score will be addressed in the analysis by expecting the long-term benefit from the treatment for that subject to be zero from the time of intercurrent event. For the other intercurrent events, rescue medication use within 24 hours of WOMAC assessment, removal/detachment of study patch for more than 24 hours before WOMAC assessment, the obtained WOMAC pain score will not be included in the analysis.

16.1.5 Sensitivity Analysis of Primary Efficacy Variable

Further details of sensitivity analyses are provided in Appendix 3.

Sensitivity analyses will be performed to assess the impact of assumptions about missing data patterns on the primary and key secondary efficacy inferences. Additional analysis due to intercurrent events occurring in the trial will be performed.

The intercurrent events are defined as:

1. Discontinuation due to reasons related to study treatment (e.g. AE, LOE, lack of compliance, deviation from protocol, etc.),
2. Rescue medication drug use within 24 hours of WOMAC pain assessment,

3. Removal/detachment of the patch for more than 24 hours before WOMAC pain assessment.

Table B provides details on how these intercurrent events will be used for the sensitivity analyses.

Table B: Sensitivity Analysis Methods for Intercurrent Events

Analysis	Endpoint	Methods for Intercurrent Events		
		Discontinuation related to Study Medication	Rescue Medication	Patch Removal/ Detachment
Primary Efficacy	Mean change in WOMAC pain score from baseline to Week 12	1. Pattern Mixture Model	1. Pattern Mixture Model	1. Pattern Mixture Model
		2. Copy Reference Model	2. Copy Reference Model	2. Copy Reference Model

Pattern-Mixture Model and Copy Reference Sensitivity Analyses of MMRM Approach:

Robustness of the primary results under MNAR (Missing not at random) mechanisms will be assessed using pattern-mixture model and copy reference imputation approaches. All sensitivity analyses will be based on the pattern mixture model using multiple imputation for the dropout reasons and their assumed pattern of missing data as follows:

- LOE as MNAR with other missing data as MAR
- LOE and AE as MNAR with other missing data as MAR
- Any reason (all dropout) as MNAR

a. Multiple Imputation: Pattern Mixture Model (PMM)

The missing data due to early discontinuation will be imputed assuming MNAR following the approach described in Permutt (2016)[1]. That approach will be implemented by imputing missing data first under the MAR assumption in each treatment group using multiple imputation and then adding a pre-defined Δ1 to each imputed value in the placebo group and adding a different pre-defined Δ2 to each imputed value in the active treatment group, then varying Δ1 and Δ2 over an enclosure of a plausible range of values. As noted in the article by Permutt (2016)[1], this is considered the most appropriate kind of sensitivity analysis for the missing data problem. Also, this imputation approach was documented on Page 89 of the NRC report (2010)[2].

b. Multiple Imputation: Copy Reference (Placebo) Approach

This is placebo-based imputation for missing data which will be implemented assuming missing data from the placebo group as MAR and also as MNAR. This will be performed by first implementing placebo-based imputation for missing data in the active treatment group and the placebo group (as MAR). Then, an adjustment for MNAR in both treatment groups will be implemented by adding a common shift (Δ) to the imputed data, worsening the imputed data further, in both the placebo and the active treatment groups as described in Koch and Wienner (2016)[3]. This step is necessary because it is possible to have a lower treatment discontinuation rate in the active treatment group than in the placebo group and it is important to consider a MNAR approach in the placebo group in such a case. Using a similar argument as in the previous paragraph, the shift Δ will assume similar values to those of Δ1 (while the case of Δ=0 represents the common copy placebo method, where missing data from placebo are considered as MAR).

16.1.6 Supplementary Analysis of Primary Efficacy Variable

A supplementary analysis will be done on the primary efficacy variable; the MMRM will be completed as described in Section 16.1.3 for the change from baseline to Week 12 in WOMAC pain score, but no adjustment for intercurrent events will be done.

16.2 Secondary Efficacy

The secondary efficacy analysis will be performed for the ITT Set.

16.2.1. Key Secondary Efficacy Variables and Derivation

16.2.1.1 First Key Secondary Efficacy Variables and Derivation

The endpoint to address the first Key Secondary Objective is the WOMAC LK3.1 OA (physical function) change from baseline at Week 12 (Visit 8)

The WOMAC physical function score, is derived by summing the sub-scores for the 17 following items, collected in the *Western Ontario and McMaster Universities Osteoarthritis Index* eCRF page taking into account missing observations as described in section 7.3:

Question: How much difficulty have you had:

[redacted]

[redacted]

Each item is rated on a scale of 0 to 4, with 0 being none and 4 being extreme. Therefore, the WOMAC physical function score, for a specific visit, will be one of the integer values from 0 to 68 or missing.

16.2.1.2 Second Key Secondary Efficacy Variables and Derivation

The endpoint to address the second Key Secondary Objective is the WOMAC LK3.1 OA (stiffness) change from baseline at Week 12 (Visit 8).

The WOMAC stiffness score is derived by summing the sub-scores for the 2 following items, as collected in the *Western Ontario and McMaster Universities Osteoarthritis Index* eCRF page taking into account missing observations as described in section 7.3:

[redacted]

Each item is rated on a scale of 0 to 4, with 0 being none and 4 being extreme.

Therefore, the WOMAC stiffness score, for a specific visit, will be one of the integer values from 0 to 8 or missing.

16.2.2. Analysis of Key Secondary Efficacy Variables

The MMRM models for the key secondary efficacy variables will include change from baseline in WOMAC physical function (or stiffness) score as the repeated dependent variable, with effects for treatment (HP-5000 and placebo), visit, treatment-by-visit interaction, and the baseline WOMAC physical function (or stiffness) score as a covariate. Initially an unstructured covariance matrix will be used to model the correlation among repeated measurements within subject.

If the unstructured covariance matrix fails to converge, the same procedure for selection of covariance structures as was specified for the primary efficacy will be done as listed in section 16.1.3. Descriptive and inferential statistics, comparable to those described in sections 16.1.3 through 16.1.6 for the primary efficacy variable will be presented for each of the key secondary efficacy variables.

16.2.3. Multiplicity Method for Primary and Key secondary efficacy

The fixed-sequence procedure will be applied for primary and key secondary endpoints as described in Section 7.4. All statistical tests of hypotheses in this analysis will be 2-sided with a Type I Error rate of 0.05.

16.3 Other Secondary Efficacy

Other secondary efficacy analyses will be performed for the ITT Set. For each below secondary efficacy variable, tables will be produced that display descriptive and inferential statistics by visit and treatment group. For continuous change from baseline variables descriptive and inferential statistics, comparable to those described in section 16.1.3 for the primary efficacy variable will be presented. With no adjustment for intercurrent events.

For the secondary variables expressed as proportion of subjects with a response that is categorized as binary a logistic regression model will be used to obtain the point estimate of the odds ratio between treatment groups (HP-5000 and placebo) and the associated 95% confidence interval.

16.3.1 WOMAC LK3.1 OA (pain) change from Baseline at Weeks 1, 2, 4, and 8

The change from baseline will be analyzed using a MMRM analysis. As described in section

16.1.3 for the WOMAC pain score, the same MMRM model will be applied.

A plot of the LS Mean change from baseline by visit will be presented.

16.3.2 WOMAC LK3.1 OA (PHYSICAL FUNCTION) CHANGE FROM BASELINE AT WEEKS 1, 2, 4, AND 8

The change from baseline will be analyzed using a MMRM analysis. As described in section 16.1.3 for the WOMAC physical function score, the same MMRM model will be applied.

A plot of the LS Mean change from baseline by visit will be presented.

16.3.3 WOMAC LK3.1 OA (STIFFNESS) CHANGE FROM BASELINE AT WEEKS 1, 2, 4, AND 8

The change from baseline will be analyzed using a MMRM analysis. As described in section 16.1.3 for the WOMAC stiffness score, the same MMRM model will be applied.

A plot of the LS Mean change from baseline by visit will be presented.

16.3.4 WOMAC LK3.1 OA (COMPOSITE SCORE) CHANGE FROM BASELINE AT WEEKS 1, 2, 4, AND 8

The change from baseline will be analyzed using a MMRM analysis. As described in section 16.1.3 for the WOMAC composite score, the same MMRM model will be applied, with no adjustment for intercurrent events. The WOMAC Composite score is calculated by finding the total of the scores for the 3 subscales, WOMAC pain, WOMAC physical function and WOMAC stiffness, and then dividing by 96. Higher scores on the WOMAC indicate worse pain, stiffness, and functional limitations.

16.3.5 CHANGE FROM BASELINE IN PAIN INTENSITY ASSESSED ON AN 11-POINT NRS AT WEEKS 1, 2, 4, 8, AND 12

Change from baseline in pain intensity will be tabulated on an 11-point NRS at Weeks 1, 2, 4, 8, and 12 (visits 4, 5, 6, 7, and 8) average and worst score between baseline and each week. The NRS pain score for assessment at Weeks 1, 2, 4, 8, and 12 (visits 4, 5, 6, 7, and 8) is the mean of the NRS pain scores reported for pain [redacted] using the 11-point NRS average pain intensity score and NRS worst pain intensity score. Baseline is [redacted]

Average pain intensity:

Change from baseline in average pain intensity assessed on an 11-point NRS at Weeks 1, 2, 4, 8, and 12 (visits 4, 5, 6, 7, and 8) will be presented descriptively and analyzed similar to the primary analysis of the change from baseline as the repeated variable, with effects for treatment (HP-5000 and placebo), visit, treatment-by-visit interaction, and the baseline average pain intensity score as a covariate. With no adjustment for intercurrent events.

The average pain is collected through eDiary in clinic.

Worst pain intensity:

Change from baseline in worst pain intensity assessed on an 11-point NRS at Weeks 1, 2, 4, 8, and 12 (visits 4, 5, 6, 7, and 8) will be presented descriptively and analyzed similar to the primary analysis of the primary efficacy variable (refer to Section 16.1.3). A MMRM model will be used with change from baseline as the repeated variable, with effects for treatment (HP-5000 and placebo), visit, treatment-by-visit interaction, and the baseline worst pain intensity score as a covariate. With no adjustment for intercurrent events.

The worst pain is collected through eDiary in clinic.

A figure plotting the 3-day average of the average and worst NRS pain scores by visit will be generated.

16.3.6 Change from Baseline in pain intensity assessed on an 11-point NRS of a weekly average of all available daily pain scores at each week from Week 1 through 12

Change from baseline to each week (Week 1 through 12) in pain intensity will be assessed on an 11-point NRS of a weekly average of all available daily pain scores at each week.

All available scores of daily NRS scores evaluated when the subjects replace the patch and the NRS scores evaluated before any rescue medication use will be included. Weekly pain would be the mean of average NRS pain and mean of worst NRS pain over 7 days (e.g., Week 1 = average pain scores on Days 1 through 7). Baseline is defined [redacted]

A weekly average of all available daily pain scores will be calculated from subject's diaries. MMRM model will be used with change from baseline as the repeated variable, with effects for treatment (HP-5000 and placebo), visit, treatment-by-visit interaction, and the baseline 3-day average NRS score as a covariate. With no adjustment for intercurrent events.

A figure plotting the weekly average of the average and worst NRS pain scores by week will be generated.

16.3.7 Change from Baseline in Patient Global Assessment (PGA) at Weeks 1, 2, 4, 8, and 12

Subject's will be asked to complete the following statement: "How would you rate your osteoarthritis condition over the last 24 hours?" using a 5-point scale from 0 (Very Good) to 4 (Very Poor).

Change from baseline in PGA at Weeks 1, 2, 4, 8, and 12 (visits 4, 5, 6, 7, and 8) will be analyzed similar to the primary analyses (refer to section 16.1.3). An MMRM model will be used with change from baseline in PGA as the repeated dependent variable, with effects for treatment (HP-5000 and placebo), visit, treatment-by-visit interaction, and the baseline PGA score as a covariate. With no adjustment for intercurrent events. The observed value and change from baseline will also be summarized descriptively by visit using summary statistics in addition to the above noted model.

The responders are defined as subjects achieving a score of

0: Very Good or

1: Good

Non-responders are defined as subjects achieving a score of

2: Moderate or

3: Poor or

4: Very Poor

Subjects not assessed will be considered as having missing data.

The proportion of PGA responders and non-responders will be summarized by visit. Treatment comparisons will be made using a logit model adding PGA baseline as the covariate. This model will present an estimate of the odds ratio, Wald 95% CI, and p-value. The sample SAS code used for the Logit model for the PGA analysis can be found in Appendix 2.

16.3.8 Patient Global Impression of Change (PGIC) at Weeks 1, 2, 4, 8, and 12

Subject's will be asked the following question: "How would you rate your overall improvement with treatment during the clinical trial?" using a 7-point scale from 1 (Very Much Improved) to 7 (Very Much Worse).

PGIC at Weeks 1, 2, 4, 8, and 12 (visits 4, 5, 6, 7, and 8) will be analyzed using logit model as

below.

The responders are defined as subjects achieving a score of

1: Very much improved or

2: Much improved

Non-responders are defined as subjects achieving a score of

3: Minimally improved

4: No change or

5: Minimally worse or

6: Much worse or

7: Very much worse

Subjects achieving a score of 0: Not assessed will be considered as having missing data.

The proportion of PGIC responders and non-responders will be summarized by visit. Treatment comparisons will be made using a logit model with treatment effect as the only term in the model analyzing the proportion of PGI responders. For each comparison, an estimate of the odds ratio, corresponding Wald 95% CI, and p-value will be presented. The observed values will also be summarized descriptively by visit using summary statistics.

16.3.9 Treatment Response: Reduction in the WOMAC LK3.1 OA Index (pain)

The number and percentage of subjects with reduction in the WOMAC pain score of ≥30%, ≥50%, ≥70%, and ≥90% at Weeks 1, 2, 4, 8, and 12 (visits 4, 5, 6, 7, and 8) from baseline will be summarized by treatment groups.

The responders are defined as subjects with a reduction in the WOMAC pain score of ≥30%, ≥50%, ≥70%, and ≥90%, and non-responders are defined as subjects with a reduction in the WOMAC pain score of <30%, <50%, <70%, and <90%, respectively. Treatment comparisons will be made using a logit model with treatment effect as the only term in the model analyzing the proportion of WOMAC pain score responders. For each comparison, an estimate of the odds ratio, corresponding Wald 95% CI, and p-value will be presented.

16.3.10 Treatment Response: Reduction in the WOMAC LK3.1 OA Index (physical function)

The number and percentage of subjects with reduction in the WOMAC physical function score of ≥30%, ≥50%, ≥70%, and ≥90% at Weeks 1, 2, 4, 8, and 12 (visits 4, 5, 6, 7, and 8) from baseline will be summarized by treatment groups.

The responders are defined as subjects with a reduction in the WOMAC physical function score of ≥30%, ≥50%, ≥70%, and ≥90%, and non-responders are defined as subjects with a reduction in the WOMAC physical function score of <30%, <50%, <70%, and <90%, respectively. Treatment comparisons will be made using a logit model with treatment effect as the only term in the model analyzing the proportion of WOMAC physical function score responders. For each comparison, an estimate of the odds ratio, corresponding Wald 95% CI, and p-value will be presented.

16.3.11 Treatment Response: Reduction in the WOMAC LK3.1 OA Index (stiffness)

The number and percentage of subjects with reduction in the WOMAC stiffness score of ≥30%, ≥50%, ≥70%, and ≥90% at Weeks 1, 2, 4, 8, and 12 (visit 4, 5, 6, 7, and 8) from baseline will be summarized in a table by treatment groups.

The responders are defined as subjects with a reduction in the WOMAC stiffness score of ≥30%, ≥50%, ≥70%, and ≥90%, and non-responders are defined as subjects with a reduction in the WOMAC stiffness score of <30%, <50%, <70%, and <90%, respectively. Treatment comparisons will be made using a logit model with treatment effect as the only term in the model analyzing the proportion of WOMAC stiffness score responders. For each comparison, an estimate of the odds ratio, corresponding Wald 95% CI, and p-value will be presented.

16.3.12 Proportion of subjects, the number of days, and the total number of rescue medication tablets used during the treatment phase

Proportion of subjects who used rescue medication at any time during the entire treatment period and descriptive statistics for the number of days these subjects used rescue medication and the total number of rescue medication tablets used during the treatment phase will be presented by treatment arm.

The number and percentage of subjects using rescue medication will be summarized by treatment and overall. A logit model will be used to compare treatment against placebo. This model will not contain any baseline covariates but will present an estimate of the odds ratio, Wald 95% confidence interval, and p-value for use of rescue medication. A by-subject listing of

all efficacy assessments collected will be presented.

For number of days of rescue medication and number of tablets of rescue medication during the treatment phase, an independent two sample t-test will be used to compare treatment against placebo. An estimate of the difference, the 95% CI, and the p-value will be presented. If the normality assumption is not supported by the data, a Wilcoxon Rank Sum Test will be used.

16.3.13 The onset of the effect: Pain intensity assessed on an 11-point NRS change from Baseline

Onset of the effect of HP-5000 will be defined as the first significant difference in NRS pain reduction between HP-5000 vs Placebo.

16.3.14 Mini-OA Knee and Hip QOL (Mini-OAKHQOL) scores change from Baseline at Weeks 4 and 12

The Mini-OA Knee and Hip QOL (Mini-OAKHQOL) questionnaire contains 20 items that describes QOL in six domains: physical activities (item numbers 1, 2, 3, 4, 5, 7, and 11), mental health (item numbers 8, 9, 17, and 18), pain (item numbers 12, 13, and 16), social support (item numbers 19 and 20), social functioning (item numbers 14 and 15), and Independent Items (item 6 and 10). Each item is scored on a scale of 0 to 10. For each specific domain, a score is obtained by computing the mean of the items in that domain, for each subject at baseline and subsequent visits. The derived scores at baseline and post-baseline visits will be summarized as a change from baseline to the post-baseline visit. The baseline, post-baseline, and change from baseline will be tabulated for each domain, except for the Independent Items domain, by treatment group and overall.

Mini-OAKHQOL for each domain will be analyzed using MMRM with change from baseline as the repeated dependent variable, with effects for treatment (HP-5000 and placebo), visit, treatment-by-visit interaction, and the baseline score as the covariate. With no adjustment for intercurrent events.

16.3.15 Work Productivity and Activity Impairment Questionnaire for Osteoarthritis (WPAI: OA) impairment scores change from Baseline at Weeks 4 and 12

Work Productivity and Activity Impairment Questionnaire for Osteoarthritis (WPAI: OA) impairment scores change from baseline at Weeks 4 and 12 will be tabulated. The score for each of the domains (work time missed, impairment at work, overall work productivity loss, activity impairment) from the 6-item questionnaire (Q1 through Q6) will be used to analyze the

four endpoints using the formulas below.

Endpoint	Formula
1.Percent Activity Impairment due to OA	Q6*10
2.Percent Impairment while working due to OA	Q5*10
3. Percent Overall Work Impairment due to OA	100*{Q2/(Q2+Q4) + [1– Q2/(Q2+Q4)]*(Q5/10)}
4.Percent work time missed due to OA	100*Q2/(Q2+Q4)

The sub-scores are expressed as an impairment percentage (range from 0 to 100), with higher numbers indicating greater impairment and less productivity.

Percent Activity Impairment for each endpoint will be analyzed using MMRM with change from baseline as the repeated dependent variable, with effects for treatment (HP-5000 and placebo), visit, treatment-by-visit interaction, and the baseline score as a covariate. With no adjustment for intercurrent events.

16.3.16 Responders and Non-Responders Analysis

The proportion of subjects who are responders for WOMAC pain score, WOMAC physical function score, stiffness score, composite score, NRS, and users of rescue medication will be based on ≥ X% improvement from baseline[6]. At each time point, subjects having X% or greater improvement from baseline in WOMAC pain score will be defined as “responders”. The percentage change from baseline is defined as (Value at Week X – Baseline value) * 100 / (Baseline value). In other words, the responder indicator will be set to 1 if the percentage change is less than or equal to -X% and set to 0 if the percentage change is greater than -X% or missing. The different levels of required percent improvement from baseline are X=5% to 100% with 5% increments. These different thresholds for improvement will be summarized by visit. Results will be reported by means of a graph with response threshold on the X-axis and proportion of subjects who are responders in each treatment group on the Y-axis, this figure will only be completed for the WOMAC pain score. Corresponding statistics will be presented in a summary table. The same statistics will be provided for WOMAC stiffness score, physical function score, and composite score.

17. Quality of Life Analysis

Quality of life analysis is as described in other secondary efficacy outcomes in section 16.3.14.

18. SAFETY OUTCOMES

Safety analyses will be conducted using data from the SAF population. Safety variables include AEs, clinical laboratory test results, body weight, vital signs, ECG readings, physical examination results, and dermal safety results. No formal inferential analyses will be conducted for safety variables, unless otherwise noted.

18.1 ADVERSE EVENTS

AEs will be coded using MedDRA version in effect at the time of final database lock.

TEAEs are defined as:

AEs that begin or pre-existing conditions that worsen at the time of or following the start of treatment with study drug through the Follow-up Visit or Early Termination Visit, whichever occurs first.

See Appendix 2 for handling of partial or completely missing dates for AEs. In the case where it is not possible to define an AE as treatment emergent or not, the AE will be classified as the worst case, i.e., treatment emergent. Adverse event listings will display start and end dates as collected (even if partial) and they will be sorted by treatment group, subject ID, then by SOC/PT in alphabetical order.

The AE summaries that will be provided are described below in sub-sections 18.1.1 and 18.1.2. Additional details are included in the relevant table and listing shells.

Listings will include all adverse events.

18.1.1 ALL TEAEs

The incidence of TEAEs will be presented by SOC and PT and also broken down further by severity and relationship to study medication.

The number and percentage of subjects with at least one TEAE and the number of TEAEs the subjects had will be presented for each SOC and each PT within SOC by treatment group and all treated subjects. An additional summary of TEAEs will be provided for PTs occurring in at least 1% of subjects in HP-5000 arm and having a higher percentage of subjects in the HP-5000 Arm than the Placebo Arm, this summary will be presented by SOC and PT within SOC. Summaries will be presented by treatment group and in decreasing frequency. AEs which occur during the follow-up period after the last patch removal will be summarized separately by SOC and PT

within SOC and treatment group.

18.1.1.1 Severity

The summary of TEAEs by SOC and PT will be expanded to show the number and percent of subjects and number of events by severity level. Severity is classified as “mild”, “moderate”, “severe”, or “life-threatening”. TEAEs with missing severity will be classified as “not specified”. If a subject reports a TEAE more than once within the same PT and SOC, the event with the worst-case severity will be used in the corresponding severity summaries.

18.1.1.2 Relationship to Study Medication

Relationship, as indicated by the Investigator, is classed as “not related”, “unlikely related”, “possibly related”, or “definitely related”. A “related” TEAE is defined as a TEAE with a relationship to study medication as “possibly related”, or “definitely related” to study medication. TEAEs with missing relationship to study medication will be classified as “related”. If a subject reports the same AE more than once within the same PT and SOC, the AE with the worst-case relationship to study medication will be used in the corresponding relationship summaries. A summary of related vs. non-related TEAEs by SOC and PT will be presented.

The number and percentage of subjects with at least one “related” AE (Adverse reaction by investigator) and the number of TEAEs the subjects had will be presented for each SOC and each PT within SOC by treatment group and all treated subjects. An additional summary of related AEs will be provided for PTs occurring in at least 1% of subjects in any treatment group. Summaries will be presented by treatment group and in decreasing frequency.

18.1.2 TEAEs Leading to Discontinuation of Study Medication and Discontinuation From Study

AEs leading to early withdrawal or discontinuation of study medication will be identified by using the Adverse Events page of the eCRF, where item ‘Action taken with study treatment’ indicates permanent discontinuation of study medication, i.e., “Drug withdrawn”. AEs leading to early study withdrawal will be identified by using the Adverse Events page of the eCRF, where the answer to item ‘Did the AE cause the subject to discontinue from the study?’ indicates “Yes”.

For TEAEs leading to early study withdrawal or discontinuation of study medication, summaries of incidence rates (frequencies and percentages) by SOC and PT will be prepared. The listings of those AEs will be provided.

18.1.3 Serious Adverse Events and Death

Serious adverse events (SAEs) are those events recorded as "Serious" on the Adverse Events page of the eCRF. A summary of SAEs by SOC and PT will be prepared.

18.1.4 Adverse events Leading to Death

AEs leading to death are those events which are recorded as "Fatal" on the Adverse Events page of the eCRF. Deaths may also be recorded on the End of Study page of eCRF. A listing of all deaths will be prepared based on these sources.

18.1.5 TEAEs at Application Site

A summary table with TEAEs at application site will be presented by SOC and PT. This data will be collected from adverse event eCRF form. If the question "is the event at the application site?" is marked "yes".

18.2 Deaths

A listing of all deaths will be presented based on the section 18.1.4.

18.3 Laboratory Evaluations

The list of laboratory assessments to be included in the outputs is as in protocol section 10.3.2.1.1. Results from the central laboratory for hematology, clinical chemistry, and urinalysis will be included in the reporting of this study. If central laboratory tests could not be performed, local laboratory results may be collected. All laboratory test results will be presented in SI units and listed using the sort order subject number, clinic visit number and relative study day (See the protocol section 10.2 and SAP section 6.1 for definitions.)

Quantitative laboratory measurements reported as "< X", i.e. below the limit of quantitation (BLQ), or "> X", i.e. above the upper limit of quantitation (ULQ), will be converted to X for the purpose of quantitative summaries, but will be presented as recorded, i.e. as "< X" or "> X" in the listings.

The following summaries will be provided for laboratory data for each parameter by treatment group and study visit:

Actual values and change from baseline (for quantitative measurements);

Shifts from baseline according to normal/reference range criteria (for quantitative and categorical measurements).

18.4 ECG EVALUATIONS

The following ECG parameters will be reported for this study:

PR Interval (msec);

QRS Interval (msec);

QT Interval (msec);

RR Interval (msec);

QTcF Interval (msec) [derived by central ECG];

HR (bpm);

Overall assessment of ECG.

Should ECGs be collected in triplicate or multiple assessments on the same day, they will be analyzed as an average of the non-missing measurements on the same day. For the overall assessment, there are three possible results from the central cardiologist: 'Abnormal, Not Clinically Significant', 'Normal', and 'Abnormal, Clinically Significant'. All assessments will be summarized according to the worst non-missing assessment and categorized in the following order, "Normal", "Abnormal, Not Clinically Significant", "Abnormal, Clinically Significant".

The following summaries will be provided for ECG data for each parameter by treatment group and study visit:

Actual and change from baseline at Visit 8 (for quantitative measurements);

Shifts from baseline to worst overall ECG result.

18.5 VITAL SIGNS

The following vital sign measurements will be reported for this study:

Systolic blood pressure (mmHg);

Diastolic blood pressure (mmHg);

Pulse rate (bpm);

Respiratory rate (breaths/min);

Body temperature;

Weight (kg);

BMI (kg/m2) [derived].

The following summaries will be provided for vital signs data for each parameter by treatment group and study visit:

Actual and change from baseline

BMI Weight Status Categories[7]

BMI (kg/m^2)	Weight Status
< 18.5	Underweight
18.5 ≤ BMI < 25.0	Normal
25.0 ≤ BMI < 30	Overweight
30.0 ≤ BMI	Obese

18.7 Other Safety Assessments

18.7.1 Dermal evaluations at Clinic

18.7.1.1 Patch Adhesion

Results of patch adhesion assessed at clinic and collected in the *Dermal Evaluations* eCRF form will be tabulated by visit and overall, for each treatment group. The frequency and percentage of patches in each category will be summarized, per the 5-point numerical scale described below, for each visit and overall. Also, for each patch adhesion score category, the number and percent of subjects with worst score in the specific category is presented by treatment group.

The following 5-point numerical scale is used to assess the patch adhesion:

0 = ≥90% adhered (essentially no lift off the skin).

1 = ≥75% to <90% adhered (some edges only lifting off of the skin).

2 = ≥50% to <75% adhered (less than half of the system lifting off of the skin).

3 = <50% adhered but not detached (more than half the system lifting off of the skin without falling off).

4= patch detached (patch completely off the skin).

18.7.1.2 Discomfort

Discomfort reported by the subjects will is collected in clinic on the *Dermal Evaluations* eCRF form. Number and percentage of subjects who report mild, moderate, severe, and no discomfort will be presented for each treatment group by visit and overall. The overall summary will show the number and percent of subjects with worst patch discomfort score in the specific category by treatment group. The scale used for assessment is as below:

0 = No discomfort

1 = Mild discomfort

2 = Moderate but tolerable discomfort

3 = Severe, intolerable discomfort

4 = Patch not present

Types of discomforts (itching, burning, pain, stinging, soreness, dryness, other) will be summarized and listed.

18.7.1.3 Irritation

Results of irritation assessment in clinic after patch removal as collected in *Dermal Evaluations* eCRF form using 2-part scales (Dermal Response and Other Effects) will be tabulated for each visit as a concatenated and combined score.

The following Berger and Bowman numerical scale is used to assess the Dermal Response:

- 0 = no evidence of irritation
- 1 = minimal erythema, barely perceptible
- 2 = definite erythema, readily visible; minimal edema or minimal papular response
- 3 = erythema and papules
- 4 = definite edema

- 5 = erythema, edema, and papules
- 6 = vesicular eruption
- 7 = strong reaction spreading beyond the application site

The Other Effects will be scored via a letter as below:

- A (0)= slightly glazed appearance
- B (1) = marked glazed appearance
- C (2) = glazing with peeling and cracking
- F (3) = glazing with fissures
- G (3) = film of dried serous exudates covering all or part of the patch site
- H (3) = small petechial erosions and/or scabs
- N (0) = no other observations

The combined irritation score will be calculated as a numerical total; i.e., numerical "Dermal Response" score + numeric equivalent for the "Other Effects" lettered score (note that the "Other Notations" do not have numeric equivalents and do not contribute to the numeric Combined irritation score); e.g., 2N, 2A, and 3G will be reported as 2, 2, and 6 combined irritations scores respectively.

The frequency and percentage of subjects with patches in each combined and concatenated category will be summarized from the above scales at each visit and overall.

18.7.2 Dermal evaluations at home in e-diary

At home, subjects will complete a questionnaire about adhesion, irritation, discomfort, and adhesive residue (on forms scheduled patch diary, unscheduled patch diary) in the eDiary. Data will be captured in the Scheduled Patch Diary Form and Unscheduled Patch Diary eCRF form. A by-subject listing of this data will be presented in addition to summary tables for the number of events for each of the dermal assessments such as adhesion, irritation, discomfort, and adhesive residue.

Additionally, the data regarding the adhesive residue at Visit 8 will be summarized based on the eCRF form Patch Removal at the site at the final visit.

19. TOPLINE AND FINAL DATABASE TRANSFER

After database lock, the topline results that contains only five tables will be delivered to the study sponsor before the complete set of outputs. The topline tables will include:

1) Table 14.1.1.1 Subject Disposition
2) Table 14.2.2.2.1 Change from Baseline in WOMAC Pain Score by Visit
3) Table 14.2.2.2.2 Change from Baseline in WOMAC Physical Function Score by Visit
4) Table 14.2.2.2.3 Change from Baseline in WOMAC Stiffness Score by Visit
5) Table 14.3.1.2 Summary of Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

After finalization of the analysis and database lock, the following will be transferred to the study sponsor:

- Randomization List
- Statistical Analysis Plan
- Annotated CRF
- All SAS programs used in the project for statistical analysis, report tables generation, and analysis dataset creation (including all source code for any macros that may be used)
- Tables, listings, and figures as included in the clinical study report
- Final SAS datasets, with full audit trail from initial data entry through final accepted version
- ADaM datasets that are necessary for analysis will be programmed in addition to SDTM
  (to be modeled in accordance with the Analysis Data Model version 2.1: Implementation Guide version 1.1, and the CDISC Study Data Tabulation Model version 1.4 or latest: Implementation Guide version 3.2 or latest)
- Define.xml for ADaM and SDTM data sets (to be modeled in accordance with the CDISC Define-XML Specification Version 2.0 or latest)
- Study data and analysis data reviewer guides (to be modeled in accordance with the Analysis Data Reviewer's Guide Completion Guidelines Version 1.1, and the Study Data
- Reviewer's Guide Completion Guidelines Version 1.2) for any unavoidable CDISC checker errors or warnings (found using Pinnacle21)
- Pinnicle21 reports with explanation for errors/warnings if any
- Relevant Correspondence

Database archiving at [redacted] is performed in accordance with [redacted] SOPs, which ensure security, integrity, disaster recovery, and adequate backup.

REFERENCES

1. PERMUTT T. SENSITIVITY ANALYSIS FOR MISSING DATA IN REGULATORY SUBMISSIONS. STAT MED. 2016;35(17):2876–2879
2. NATIONAL RESEARCH COUNCIL. THE PREVENTION AND TREATMENT OF MISSING DATA IN CLINICAL TRIALS. WASHINGTON, DC: NATIONAL ACADEMIES PRESS; 2010
3. KOCH GG, WIENNER LE. COMMENTARY FOR THE MISSING DATA WORKING GROUP'S PERSPECTIVE FOR REGULATORY CLINICAL TRIALS, ESTIMANDS, AND SENSITIVITY ANALYSES. STAT MED. 2016;35:2887–2893
4. https://database.ich.org/sites/default/files/E9-R1_EWG_Step2_TrainingMaterial.pdf
5. https://www.ecfr.gov/cgi-bin/text-idx?SID=3ee286332416f26a91d9e6d786a604ab&mc=true&tpl=/ecfrbrowse/Title21/21tab_02.tpl
6. FDA Guidance: Analgesic Indications: Developing Drug and Biological Products; CDER February 2014
7. https://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmicalc.htm

APPENDIX 1. PARTIAL DATE CONVENTIONS

Imputed dates will NOT be presented in the listings. The below tables in Appendix 1 represents the study medication start date as patch application and study medication end date as patch removal.

ALGORITHM FOR TREATMENT EMERGENCE OF ADVERSE EVENTS

START DATE	STOP DATE	ACTION
Known	Known	If start date < study medication start date or start date>study medication end date + 1 day, then not TEAE If study medication start date ≤start date ≤ study medication end date + 1 day, then TEAE
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If start date < study med start date or start date>study med end date + 1 day, then not TEAE If study med start date ≤start date ≤ study med end date + 1 day, then TEAE
	Missing	If start date < study med start date or start date>study med end date + 1 day, then not TEAE If study med start date ≤start date ≤ study med end date + 1 day, then TEAE
Partial, but known components show that it cannot be on or after study medication start date or cannot be	Known	Not TEAE

START DATE	STOP DATE	ACTION
before study medication end date		
	Partial	Not TEAE
	Missing	Not TEAE
Partial, could be on or after study medication start date but before study med end date	Known	If stop date < study med start date, then not TEAE If stop date ≥ study med start date, then TEAE
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If stop date < study med start date, then not TEAE If stop date ≥ study med start date, then TEAE
	Missing	Assumed TEAE
Missing	Known	If stop date < study med start date, then not TEAE If stop date ≥ study med start date, then TEAE
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If stop date < study med start date, then not TEAE If stop date ≥ study med start date, then TEAE

START DATE	STOP DATE	ACTION
	Missing	Assumed TEAE

ALGORITHM FOR PRIOR / CONCOMITANT MEDICATIONS, SURGERIES, PROCEDURES, AND THERAPIES

START DATE	STOP DATE	ACTION
Known	Known	If stop date < study med start date, assign as prior If stop date ≥ study med start date and start date < study med start date, assign as prior concomitant If stop date ≥ study med start date and start date ≤ end of treatment, assign as concomitant If stop date ≥ study med start date and start date > end of treatment, assign as post study
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If stop date < study med start date, assign as prior If stop date ≥ study med start date and start date < study med start date, assign as prior concomitant If stop date ≥ study med start date and start date ≤ end of treatment, assign as concomitant If stop date ≥ study med start date and start date > end of treatment, assign as post study
	Missing	If stop date is missing could never be assumed a prior medication If start date ≤ end of treatment, assign as concomitant If start date > end of treatment, assign as post treatment

START DATE	STOP DATE	ACTION
Partial	Known	Impute start date as earliest possible date (i.e. first day of month if day unknown or 1st January if day and month are unknown), then: If stop date < study med start date, assign as prior If stop date ≥ study med start date and start date < study med start date, assign as prior concomitant If stop date ≥ study med start date and start date ≤ end of treatment, assign as concomitant If stop date ≥ study med start date and start date > end of treatment, assign as post study
	Partial	Impute start date as earliest possible date (i.e. first day of month if day unknown or 1st January if day and month are unknown) and impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If stop date < study med start date, assign as prior If stop date ≥ study med start date and start date < study med start date, assign as prior concomitant If stop date ≥ study med start date and start date ≤ end of treatment, assign as concomitant If stop date ≥ study med start date and start date > end of treatment, assign as post study
	Missing	Impute start date as earliest possible date (i.e. first day of month if day unknown or 1st January if day and month are unknown), then: If stop date is missing could never be assumed a prior medication If start date ≤ end of treatment, assign as concomitant If start date > end of treatment, assign as post treatment
Missing	Known	If stop date < study med start date, assign as prior If stop date ≥ study med start date, assign as concomitant Cannot be assigned as 'post treatment'

START DATE	STOP DATE	ACTION
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If stop date < study med start date, assign as prior If stop date ≥ study med start date, assign as concomitant Cannot be assigned as 'post treatment'
	Missing	Assign as concomitant

DocuSign Envelope ID: D2BB5456-E458-4E7C-8650-F63E715880F5

APPENDIX 2. **SAS CODE**

PROC MIXED:

ODS OUTPUT LSMEANS=LSEST DIFFS=LSDIFF COVPARMS=COVS;
PROC MIXED DATA = TEST *COVTEST*;
CLASS TREATMENT VISIT SUBJECT;
MODEL change in womac pain score = treatment visit treatment*visit baseline womac pain score;
LSMEANS TRTPN*AVISITN / DIFF CL;
REPEATED VISIT / TYPE = UN SUBJECT = SUBJECT;
RUN;

PROC LOGISTIC:

proc logistic data="Test" descending;
where visit variable = a given number (e.g., 4)
class treatment / param=ref ;
model respond = treatment PGA baseline value;
run;

APPENDIX 3. IMPLEMENTATION OF SENSITIVITY ANALYSES AND ANALYSES FOR OTHER RELEVANT ESTIMANDS

Sensitivity analyses of the primary and key secondary efficacy endpoint will be performed to assess the impact of assumptions about missing data patterns on the primary inferences during the trial.

The change from baseline to Week 12 summary statistics will be presented for both imputation approaches described below. The dropout reason, shift value, treatment group, LS Mean difference values along with 95% CIs will be presented. Nominal p-values comparing the HP-5000 treatment group versus placebo will also be displayed.

a. Multiple Imputation: Pattern Mixture Model (PMM)

Step 0: Missing data will be imputed using a multiple imputation technique assuming MAR for all treatment groups.

Step 1: A shift (Δ1) will be added to all imputed data in Step 0 in the Placebo group to worsen imputed scores. The range of values Δ1 can be calculated from Δ1=k1 × the upper bound (UB1) where k1=0, 0.33, 0.67, and 1 × the UB1 for Δ1. The UB1 will be based on the end point, UB1 for WOMAC Pain is 4, UB1 for WOMAC Physical Functioning is 12, and UB1 for WOMAC Stiffness is 2. The UB1's for Δ1 are selected to be reasonably higher than the SD estimated from the blinded data to provide a sufficiently wide range for the imputation. Note for WOMAC Pain Δ1=0, 1.3333, 2.6667, 4 for imputation purpose. For WOMAC Physical Functioning Δ1=0, 4, 8, 12. For WOMAC Stiffness Δ1=0, .6667, 1.3333, 2.

Step 2: A shift (Δ2) where Δ2=Δ1+ k2 × LS mean difference (LSMD) is also added to the imputed data in Step 0 in the active treatment group to worsen their imputed scores. For each value of Δ1, the value of Δ2 will be calculated using k2 (k2=0, 0.2, 0.4, 0.6, 0.8 and 1) × LSMD between the active treatment group vs placebo based on the MMRM primary analysis. It is important to note that if k2=0, then Δ2=Δ1, and if k2=1, then Δ2=Δ1+ LSMD; therefore, the considered range for Δ2 is from Δ1 to Δ1+ LSMD. A total of 24 combinations of shifts will be added to the imputed values.

Step 3: The primary MMRM model is then applied to each of the complete datasets created from Step 0 through Step 2.

Step 4: Step 0 through Step 3 are then repeated m=100 times and PROC MIANALYZE is then applied to combine the estimates using Rubin's combination rules.

Justification of the ranges of Δ1 and Δ2 values to cover the plausible domain of the missing values is supported by the following, using WOMAC Pain UB1. The Δ1 will assume a value ranging from 0 to 4. This UB1 of 4 for Δ1 is selected to be reasonably higher than 1 SD of the primary endpoint of the study and is used to cover wider range of imputations. For the case of the lower boundary of Δ1=0, the missing data in the placebo group are considered as MAR. For the case of upper boundary of the range of Δ1=4, the missing data in the placebo group are considered MNAR with a fixed worsening shift of 4 from the imputed values assuming MAR of all missing WOMAC data. Such a magnitude represents a reasonable upper bound for the range of deviation from MAR imputation. The boundaries of Δ2 are selected to start from Δ2=Δ1 to Δ2=Δ1+LSMD. With the starting values of Δ2=Δ1, MNAR data for discontinued subjects either from the active treatment group or the placebo group are worsened by the same shift from their imputed WOMAC total score based on MAR. With the upper boundary of Δ2 defined by Δ2=Δ1+LSMD, the MNAR data for discontinued subjects from the active treatment group will have their WOMAC imputed scores based on MAR worsened by larger amounts that are equal to the whole LSMD rather than a fraction of the LSMD (in addition to the Δ1).

b. Multiple Imputation: Copy Reference (Placebo) Approach

Step 0: Missing data in the placebo group are imputed using a multiple imputation technique assuming MAR.

Step 1: Missing data in the active treatment group are imputed based on the conditional distribution estimated from the placebo group using a multiple imputation technique.

Step 2: A range of common shifts (Δ=0, 1.3333, 2.6667, 4 units – WOMAC Pain) are added to the imputed data in Step 0 for the placebo group and in Step 1 for the active treatment group to worsen imputed scores. (Δ=0, 4, 8, 12 units – WOMAC Physical Functioning) (Δ=0, 0.6667, 1.3333, 2 units – WOMAC Stiffness)

Step 3: The primary MMRM model is then applied to each of the complete datasets created from Step 0 through Step 2.

Step 4: Step 0 through Step 3 are repeated m=100 times and PROC MIANALYZE is then applied to combine the estimates using Rubin’s combination rules.

APPENDIX 4. STUDY MEDICATION EXPOSURE DERIVATIONS

The date of 1st patch application will be collected in the Patch Application at site at Visit 3 eCRF page.

The last patch application date will be the latest date collected in the *Scheduled Patch Diary* or *Unscheduled Patch Diary* eCRF pages.

The last patch removal date will be the latest date collected in the *Scheduled Patch Diary, Unscheduled Patch Diary,* as well as in the *Patch Removal at the site at final visit* eCRF page.