Statistical Analysis Plan

Sponsor:	Eupraxia Pharmaceuticals Inc.
Protocol No:	EP-104IAR-201
Protocol Title:	A Phase 2, Randomized, Double-Blind, Vehicle-Controlled Study Evaluating the Safety, Efficacy and Pharmacokinetics of Single Dose EP-104IAR for 24 Weeks in Patients with Osteoarthritis of the Knee
Document Date:	14-Jun-2023
Document Revision	Final 1.0

NCT04120402

APPROVAL SIGNATURES

[redacted]

Eupraxia Pharmaceuticals Inc. Personnel

[redacted]

Chief Scientific Officer	Amanda Malone	Amanda Malone Signer Name: Amanda Malone Signing Reason: I approve this document Signing Time: 14-Jun-2023 | 11:57:36 AM PDT E4C8A4AAEB934BBB93BDEDB030DC713A	14-Jun-2023

DOCUMENT HISTORY

Rev No	Date	Description
1.0	14-Jun-2023	Original document

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Definition
ACTH	Adrenocorticotropic hormone
AE	Adverse event
AI	Adrenal insufficiency
AUC_{0-t}	Area under the concentration-time curve from time zero to time of last quantifiable concentration
AUC_{0-24}	Area under the concentration-time curve from time zero to 24 weeks post-dose
$AUC_{0-\infty}$	Area under the concentration-time curve from time 0 extrapolated to infinity
BLQ	Below the lower limit of quantification
BMI	Body mass index
BOCF	Baseline observation carried forward
BP	Blood pressure
CFB	Change from baseline
C_{max}	Maximum concentration
C_{last}	Last quantifiable plasma concentration
CL/F	Apparent clearance
CM	Concomitant medication
C_{ss}	Average concentration at steady state
CSR	Clinical Study Report
CV	Coefficient of Variation
eCRF	Electronic case report form
EDC	Electronic data capture
ePRO	Electronic patient-reported outcome
EP-104IAR	Long-Acting Fluticasone Propionate for Intra-Articular Injection
FAS	Full analysis set
FP	Fluticasone propionate
HbA1c	Hemoglobin A1c
HPA	Hypothalamic pituitary adrenal axis
IA	Intra-articular
ICF	Informed consent form
ID	Identification

IMP	Investigational Medicinal Product (also referred to as study-drug)
IRB	Institutional Review Board
ITT	Intention-to-treat
KL	Kellgren and Lawrence system for classifying osteoarthritis severity
LOCF	Last observation carried forward
MDGA	Physician's Global Assessment of Arthritis
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed-effects model for repeated measures
NPRS	Numeric Pain Rating Scale
NRS	Numeric rating scale
NSAID	Non-steroidal anti-inflammatory drug
OA	Osteoarthritis
OMERACT-OARSI	Outcome Measures in Rheumatology-Osteoarthritis Research Society International
PK	Pharmacokinetic(s)
PPS	Per-protocol analysis set
PT	Preferred Term
PtGA	Patient's Global Assessment of Arthritis
PtPain	Patient pain scores (11-point numerical rating scale)
PVA	Polyvinyl alcohol
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SOC	System Organ Class
SF-36	Short Form 36 Health Survey
SRC	Safety Review Committee
SRM	Safety Review Meeting
TEAE	Treatment-emergent adverse event
$t_{1/2}$	Terminal phase half-life
t_{max}	Time to maximum observed concentration
t_{last}	Time to reach the last quantifiable plasma concentration
V_z/F	Apparent volume of distribution associated with the terminal phase after oral administration
WHO-Drug	World Health Organization drug dictionary
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index

λ_z	Apparent terminal phase rate constant

1. INTRODUCTION

1.1 Background

The investigational product EP-104IAR is a long-acting formulation of fluticasone propionate (FP) for intra-articular (IA) injection being developed by Eupraxia Pharmaceuticals Inc. for the treatment of pain associated with osteoarthritis (OA) of the knee.

FP is a synthetic trifluorinated corticosteroid with potent anti-inflammatory activity. Relative to other corticosteroids, FP has a high affinity for the glucocorticoid receptor, low solubility, a low rate of dissociation, and a comparatively long half-life. These characteristics make FP an excellent candidate for prolonged anti-inflammatory effects. FP has a well-established systemic safety record in humans given its several decades of global market presence in the form of widely used inhaled and intranasal brands such as Flovent®, Flonase® and Advair®, as well as a topical agent under other brand names. In humans, FP has shown to be locally active with virtually no oral bioavailability (systemic exposure), and what little is absorbed, is rapidly metabolized.

EP-104IAR contains FP crystals that are coated with polyvinyl alcohol (PVA) and heat-treated to form particles that release FP slowly over several months. PVA is used extensively in the medical industry and has a 30-year safety record of use in various human tissues. It has numerous biomedical applications and has been safely used as an implantable orthopedic medical device for meniscus/cartilage tissue replacement.

The technology used for manufacturing EP-104IAR differs substantially from traditional coacervation technology. In microspheres produced by coacervation technologies, the drug is dispersed in the volume of biodegradable polymer. Such drug delivery systems rely on degradation of the polymer to release the drug from drug/polymer microspheres and consequently have a high polymer to drug ratio. In contrast, EP-104IAR relies on the diffusion of low solubility FP across a very thin (4 μm) crosslinked PVA membrane comprising approximately 6% of the drug product. The combination of a highly potent and low solubility corticosteroid with low levels of crosslinked polymer is expected to translate into prolonged and stable drug delivery, with substantially less polymer injected into the knee.

EP-104IAR has shown a favorable profile in non-clinical studies, suggesting it will be a safe and effective locally administered therapeutic in OA. The safety, pharmacokinetics (PK) and preliminary efficacy of EP-104IAR have also been studied in a Phase 1 randomized, double-blind, vehicle-controlled clinical study in 32 subjects with knee OA. A single IA injection of EP-104IAR

15 mg was found to be safe and well-tolerated. Although this small study was not powered to evaluate efficacy, pain relief onset was rapid and numerical separation from placebo was maintained for up to 12 weeks post-dose [as measured via patient pain scores (PtPain, 11-point numerical rating scale) and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)] (Clinical Study Report EP-104IAR-101).

1.2 Study Design

This is a Phase 2, multi-center, randomized, double-blind, vehicle-controlled, parallel-group study.

Approximately 300 subjects with diagnosed knee OA will be randomized and followed for up to 24 weeks to evaluate the safety, efficacy, and PK of a single IA dose of EP-104IAR compared to vehicle control.

Subjects may have unilateral or bilateral knee OA. In subjects where both knees are potentially eligible Index knee selection will be based on WOMAC Pain scores collected during the Washout and Baseline period and other eligibility criteria such as Kellgren-Lawrence grade.

Safety assessments include AEs, vital signs, hematology and clinical chemistry assessments, and safety urinalysis testing. Local safety will be assessed via physical examinations of the treated knee. Evaluations of EP-104IAR's impact on the hypothalamic pituitary adrenal (HPA)-axis include morning serum cortisol collection (all subjects) and the adrenocorticotropic hormone (ACTH) stimulation test (selected subjects).

The primary means of efficacy data collection is via an electronic patient-reported outcome (ePRO) device which will be used throughout the study to collect daily assessments of pain using the Numeric Pain Rating Scale (NPRS), rescue medication and non-drug rescue therapy use, and activity levels; weekly WOMAC Pain subscale measurements; and monthly WOMAC Pain, Function and Stiffness. The MDGA, PtGA, and SF-36 health survey will also be collected at site visits.

In addition to the main study, an exploratory imaging sub-study is being performed in a subset of the total subjects, with those subjects being followed to 52 weeks post-dose (c.f. Study Protocol Version 7.0). The main study will be completed and analyzed prior to completion of the imaging sub-study, after all randomized patients have either discontinued or completed Visit 9 / Week 24; the imaging sub-study will be analyzed and reported separately. As such, no analyses of the imaging sub-study are included in this analysis plan; and data collected as part of the sub-study

(for example, adverse events or concomitant medications reported after Visit 9 / Week 24) will not be included in the analyses presented in this analysis plan. The one exception is that a summary of MRI-related treatment-emergent adverse events will be included.

1.3 Rationale

The use of injectable corticosteroids for the management of pain and stiffness associated with inflammation in patients diagnosed with OA of the knee has proven to be safe and effective. However, currently available injectable corticosteroids are suboptimal in the treatment of OA of the knee due to their limited duration of activity and risk for systemic side effects.

EP-104IAR has been developed to maximize local residency time in the knee joint and is expected to provide greater clinical benefit than currently approved injectable corticosteroids. In addition, by decreasing the peak-to-trough ratio of plasma concentrations it is anticipated that this greater duration of efficacy will be achieved with a reduced risk of systemic side effects.

Safety and efficacy data for EP-104IAR have been generated in a single clinical study in 32 subjects with knee OA (Study EP-104IAR-101). A dose of 15 mg EP-104IAR was safe and well-tolerated and all efficacy measures showed an immediate improvement that was sustained for between 8- and 12-weeks post-dose. A higher dose of EP-104IAR is anticipated to sustain efficacy for a longer duration.

2. STUDY OBJECTIVES AND ENDPOINTS

Primary Objective and Endpoint(s)
To evaluate the efficacy of a single intra-articular (IA) injection of EP-104IAR in patients with osteoarthritis (OA) of the knee for up to 24 weeks. Primary efficacy endpoint: • Change from Baseline in WOMAC Pain score at Week 12 Key secondary efficacy endpoints: • Difference in change from baseline between EP-104IAR and vehicle in WOMAC Function score at Week 12 • Difference between EP-104IAR and vehicle in the area under the WOMAC Pain-time curve to 12 weeks post-dose

- Difference in change from baseline between EP-104IAR and vehicle in WOMAC Pain score at Week 24
- Difference between EP-104IAR and vehicle in Outcome Measures in Rheumatology-Osteoarthritis Research Society International (OMERACT-OARSI) strict responders at Week 12.

Additional secondary efficacy endpoints:

- Difference in change from baseline between treatment groups in WOMAC Pain at Weeks 1-11 and 13−23
- Difference in change from baseline between treatment groups in WOMAC Pain Q1 (pain on walking) at Weeks 1-24
- Difference in change from baseline between treatment groups in average daily NPRS at Weeks 1-24
- Difference in change from baseline between treatment groups in WOMAC Function at Weeks 4, 8, 16, 20 and 24
- Difference in change from baseline between treatment groups in WOMAC Stiffness scores at Weeks 4, 8, 12, 16, 20 and 24
- Difference in change from baseline between treatment groups in WOMAC Total scores at Weeks 4, 8, 12, 16, 20 and 24
- Difference between treatment groups in time to return to baseline WOMAC Pain
- Difference between treatment groups in area under the curve (AUC) of WOMAC Pain at Weeks 1−24
- Difference between treatment groups in AUC of average daily NPRS at Weeks 1-24
- Difference between treatment groups in AUC of WOMAC Function, Stiffness and Total scores at Weeks 4, 8, 12, 16, 20 and 24
- Difference between treatment groups in OMERACT-OARSI strict responders at Weeks 4, 8, 16, 20 and 24
- Difference between treatment groups in PtGA at Weeks 4, 8, 12, 18 and 24
- Difference between treatment groups in MDGA at Weeks 4, 8, 12, 18 and 24
- Difference between treatment groups in SF-36 at Weeks 12 and 24
- Difference between treatment groups in average rescue medication usage at Weeks 1-24
- Difference between treatment groups in average activity levels at Weeks 1-24.

Secondary Objectives and Endpoints

Exploratory efficacy endpoints:

- Difference between treatment groups in patient beliefs about treatment received
- Difference between treatment groups in physician beliefs about treatment received
- Difference in the proportion of patients achieving a 30, 50 and 70% reduction in WOMAC pain at weeks 12, 18 and 24
- Subgroup analyses will be conducted to explore treatment differences within baseline characteristics age, gender, BMI, baseline pain, symptom duration, Kellgren-Lawrence (KL) grade, and disease type.

To evaluate the safety of a single IA injection of EP-104IAR in patients with OA of the knee for up to 24 weeks.

Safety will be assessed using the following:

- Treatment-emergent adverse events (TEAEs)
- Safety laboratory assessments (including serum cortisol)
- Vital signs
- Physical examinations (including knee examinations)
- ACTH stimulation tests.

To evaluate the pharmacokinetics (PK) of a single IA injection of EP-104IAR for up to 24 weeks.

Plasma concentrations of fluticasone propionate (FP) will be assessed, and the following parameters will be evaluated:

- The area under the concentration-time curve from zero to infinity ($AUC_{0\text{-}inf}$)
- The area under the concentration-time curve from time zero to 24 weeks post-dose ($AUC_{0\text{-}24wks}$)
- The maximum concentration (C_{max})
- The last quantifiable plasma concentration (C_{last})
- The time to reach the last quantifiable plasma concentration (t_{last})
- the area under the concentration-time curve from time zero to the time of last quantifiable concentration ($AUC_{0\text{-}t}$)
- The percentage of the area under the concentration-time curve extrapolated to infinity ($AUC_{\%extrap}$)
- The time to reach maximum plasma concentration (t_{max})

- The terminal disposition rate constant (λ_z) with the respective half-life ($t_{1/2}$)
- Other parameters, including V_z/F, CL/F, and dose adjusted parameters, may be determined.

3. STATISTICAL METHODOLOGY

3.1 General Principles

All collected study data will be presented in listings. All derivations and statistical analyses will be performed using SAS® software Version 9.4 (or higher) or Phoenix® WinNonlin® Version 8.3 (or higher) (Certara USA, Inc., Princeton, NJ). Prior and concomitant medications/procedures will be coded using World Health Organization (WHO) Drug Dictionary Enhanced March 1, 2021 (or most recent version prior to data base lock). Adverse events will be coded using Medical Dictionary for Regulatory Activities (MedDRA) Version 24.0 (or most recent version prior to data base lock).

Unless stated otherwise, descriptive summaries for continuous variables will include n (number of subjects with non-missing results), mean, standard deviation (SD), median, Q_1, Q_3, minimum, and maximum. For categorical variables, n and percent will be provided. Percentages will be calculated out of the number of subjects in the given sample with a non-missing result.

The primary means of efficacy data collection will be via an electronic patient-reported outcome (ePRO) device. Subjects will be provided with a hand-held device programmed to prompt the subject to record data as required throughout the study. The primary means of non-efficacy data collection will be via the electronic case report forms (eCRFs).

Summaries presented by collection day and/or by time point will be based on scheduled assessments as planned in the protocol except for all WOMAC data (See Section 4.1). Unscheduled assessments will be presented in listings and will not be included in the analyses except for: determination of baseline as described in Section 4.3, ACTH stimulation test summaries (See Section 4.4), and for the derivation of PK parameters.

3.2 Randomization

Eligible subjects will be randomly assigned to treatment and vehicle in a 1:1 allocation ratio.

3.3 Sample Size Considerations

This study has been sized making the following assumptions:

- Inferential testing is at the (two-sided) 5% level, and with a desired power of 80%.
- The primary endpoint is the difference between treatment groups in change from baseline WOMAC Pain (average of 5 pain items) at 12 weeks post-dose, as measured using a 0-10 scale
- The standard deviation of the endpoint is 2.2, as assessed from review of similar studies from clinicaltrials.gov
- The desired difference to detect is 0.8.

Based on these assumptions, a sample size of approximately 120 subjects per treatment group is required. Assuming a drop-out rate of approximately 20%, 150 subjects will be randomized and dosed in each treatment arm.

3.4 Analysis Populations

Evaluation of the primary and secondary efficacy endpoints will be performed using the Intention-to-Treat (ITT) Population. In addition, if the Per-protocol Set (PPS) Population markedly differs from the ITT population, as discussed in Section 3.7, identical analyses for the PPS population will be performed for the primary and key secondary endpoints. Safety endpoints will be analyzed using the Safety Population. Pharmacokinetic (PK) data will be analyzed using the Safety Population.

3.4.1 Intention-to-Treat Population

The intention-to-treat (ITT) population consists of all subjects who are both randomized to and receive treatment. Analysis will be conducted allocating subjects to the treatment to which they were randomized, irrespective of what was actually received.

3.4.2 Per-protocol Set Population

The per-protocol analysis set (PPS) population consists of all randomized subjects who received a dose of randomized treatment, had at least 1 post-baseline assessment, and had no major protocol violations potentially impactful of efficacy data. The subjects in this group will be analyzed based on the treatment they received.

Major protocol violations leading to exclusion from the per-protocol set population will be finalized after the blind review process and will be documented in a memo prior to unblinding of treatment assignments and database lock. All protocol deviations will be listed by subject.

3.4.3 Safety Population

The safety population consists of all randomized subjects who were administered a dose of IMP/vehicle control. The subjects in this group will be analyzed based on the treatment they received.

3.4.4 MRI Sub-Study Population

The MRI sub-study consists of all randomized subjects who entered into the exploratory imaging sub-study. The subjects in this group will be analyzed based on the treatment they received.

3.5 Subject Disposition

The number of subjects screened will be summarized overall and by site. The number of subjects randomized and included within each of the analysis sets will be summarized by treatment, site, and overall, where applicable.

The number and percentage of subjects who completed the study will be presented by treatment and overall, for each of the analysis populations. Frequency and percentage of subjects who withdrew or discontinued from the study, and the primary reason for withdrawal, will also be summarized by treatment and overall.

3.6 Subject Accounting and Baseline Characteristics

Demographic data and subject characteristics at Baseline will be summarized descriptively. Medical history, detailed OA history and Knee OA Criteria (including duration of symptoms in the index knee in years and time since diagnosis in the index knee in years, where each is defined as the number of days from the date of symptom onset or diagnosis to the date of dosing divided by 365.25) will be summarized.

Dose administration details, exposure and compliance with IMP, characterized by; volume of synovial fluid collected prior to injection, lidocaine injection, knee of injection, route of injection, injected suspension volume, residual amount of FP in the injection kit, and actual dose of FP administered, will be summarized descriptively. Incidence of prior and concomitant medications and procedures will be summarized by WHO Drug dictionary coded terms – Anatomical

Therapeutic Chemical (ATC) classification and preferred name (PN). Prior and concomitant medications/procedures are classified as follows.

- A prior medication/procedure is any medication/procedure that started prior to date of the dose of study drug. Medications with an unknown start date will also be considered prior medications.
- A concomitant medication/procedure is any medication/procedure used on or after the date of the dose of study drug. Ongoing medications/procedures or medications/procedures with an unknown end date will also be considered concomitant medications.

A medication/procedure may be classified as both prior and concomitant.

3.7 Primary Endpoint Analysis

The primary efficacy endpoint is the difference in change from baseline between EP-104IAR and vehicle in WOMAC Pain at Week 12, in the ITT population. Baseline WOMAC Pain will be calculated at the individual level and is defined as the average of no more than three pre-dose measurements including the three scheduled measurements: 2 measurements during the Washout and Baseline Period and 1 at Visit 2 (pre-dose). If the subject has less than three pre-dose measurements, the average of all available pre-dose measurements will serve as the subject's baseline value. If a subject has more than three pre-dose measurements, then the three measurements collected immediately prior to pre-dose will be used to derive baseline WOMAC pain and earlier measurements will not contribute.

Change from individual pre-dose baseline will then be calculated for each post-dosing weekly assessment from Week 1 through Week 24. A mixed-effects model for repeated measures (MMRM) model will then be fit to the change from baseline data for all subjects in the ITT population, including fixed effects terms for: site (potentially grouped, see Section 4.2 for more details); individual baseline WOMAC Pain; and the treatment-by-week interaction, where treatment and week are both defined as categorical variables. A random per-patient intercept will also be included. An unstructured covariance matrix at the patient level will initially be fit; if this model fails to converge numerically, a compound symmetry structure will be used.

No imputation of missing data will be performed for this analysis; the MMRM model allows handling of missing data under the assumption that data is missing at random (MAR). To assess robustness to the assumption that data is missing at random. , a multiple imputation approach will be used Under this approach, outcome data will be imputed according to the following logic:

- If missing data is due to lack of efficacy (as captured by the Study Termination page on eCRF), data will be imputed as if subjects received placebo.
- If missing data is due to other reasons, data will be imputed as if subjects received their actual treatment.

The imputation procedure will begin by generating 100 imputed datasets with SEED=1867004294, excluding subjects that discontinued due to lack of efficacy. The 100 imputed datasets for the WOMAC Pain scores will be generated separately for each randomized treatment with a fully conditional specification (FCS) model using the regression method. Following these initial imputations, 100 copies of the non-missing WOMAC Pain scores for subjects that discontinued due to lack of efficacy will be generated. One of these copies will be added to each of the 100 imputed datasets generated for the placebo group. A single imputation will then be generated for each of these 100 combined sets of data using the same FCS model as above. This will result in 100 imputed datasets for all subjects.

Change from baseline to Week 12 in WOMAC Pain from each of the 100 imputed datasets will be analyzed separately using an ANCOVA model with effects treatment, site (possibly grouped; see Section 4.2) and baseline WOMAC Pain. For each comparison of interest, estimates and their associated standard errors from the ANCOVA models will be combined to produce a single overall estimate and standard error. Associated confidence intervals and p-values will then be generated as appropriate.

The SAS® MI procedure will be utilized for imputing data. The first imputation step is described above.

The imputation step for subjects randomized to EP-104IAR that discontinued due to lack of efficacy will be similar to the step above except with SEED=454120304, NIMPUTE=1, and without "by trt" since imputation will be based on a single treatment group.

The results of the 100 analyses will be transformed into a normal statistic and combined into a single analysis using PROC MIANALYZE.

In addition to the robustness analysis described above, if the PPS population markedly differs from the ITT population, identical analyses for the PPS population will be performed. Subject to finalization after the blind review, the PPS population will be considered significantly different if it is less than 85% of the ITT population.

3.8 Secondary Endpoint Analyses

3.8.1 Key Secondary Efficacy Endpoints

The following secondary endpoints will be assessed in a hierarchical manner, comparing EP-104IAR versus vehicle. If the primary endpoint is statistically significant at the 5% level, then the first secondary endpoint may be formally assessed for statistical significance. If the first secondary endpoint is statistically significant at the 5% level, then the second secondary endpoint may be formally tested, and so forth. Upon failure of any inferential test at the 5% level, no subsequent secondary endpoints may be declared to be statistically significant. Analysis of these secondary endpoints will be based on the ITT population. If the PPS population markedly differs from the ITT population as described above, identical analyses for the PPS population will be performed.

The endpoints are as follows:

- Difference in change from baseline between EP-104IAR and vehicle in WOMAC Function at Week 12
  - Analysis will be identical to that described for the primary endpoint.
- Difference between EP-104IAR and vehicle in the area under the WOMAC Pain-time curve to 12 weeks post-dose
  - Area under the curve in change from baseline WOMAC Pain from Week 0 to Week 12 will be calculated on a per-individual basis using the linear trapezoidal rule. AUC will be normalized to account for subjects whose actual days and nominal days at Week 12 differ. Treatments will be compared using an ANCOVA model containing site (possibly grouped, see Section 4.2); individual baseline WOMAC Pain; and treatment group as covariates.
- Difference in change from baseline between EP-104IAR and vehicle in WOMAC Pain at Week 24
  - Analysis will be identical to that described for the primary endpoint.
- Difference between EP-104IAR and vehicle in Outcome Measures in Rheumatology-Osteoarthritis Research Society International (OMERACT-OARSI) strict responders at Week 12
  - The proportion of patients meeting the strict responder definition at Week 12 will be compared (EP-104IAR versus placebo) using Fisher's exact test. Patients who discontinue prior to Week 12 will be considered non-responders.
  - Strict responders, under the OMERACT-OARSI criteria, are defined as $\geq$ 50% improvement in WOMAC Pain or Function score and an absolute change $\geq$ 2 in the respective score (scores scaled 0 – 10).

3.8.2 Additional Secondary Efficacy Endpoints

All other efficacy data (WOMAC subscales, total WOMAC, NPRS, PtGA, MDGA, SF-36) will be listed and summarized in tables using standard summary statistics by treatment group. Exploratory analyses will be performed to identify key features of the efficacy profile. Analyses of these additional secondary endpoints will be based on the ITT population.

The following additional secondary endpoints will be assessed:

- Difference in change from baseline between treatment groups in WOMAC Pain at Weeks 1-11 and 13−23
    - This analysis will be identical to the primary analysis.
- Difference in change from baseline between treatment groups in WOMAC Pain Q1 (pain on walking) at Weeks 1-24
    - This analysis will be identical to the primary analysis.
- Difference in change from baseline between treatment groups in average daily NPRS at Weeks 1-24
    - This analysis will be identical to the primary analysis.
    - Average daily NRPS at Week 1 is defined as the average of non-missing NPRS pain scores from Day 1 through 8. For subsequent Weeks, average daily NPRS is defined as the average of the 7 days of non-missing daily NPRS data, allowing for the 8 day Week 1 (for example, Week 2 would be Days 9 through 15).
    - Baseline NPRS is defined as the average of the seven non-missing days immediately prior to dosing (Day 1).
- Difference in change from baseline between treatment groups in WOMAC Function at Weeks 4, 8, 16, 20 and 24
    - This analysis will be identical to the primary analysis.
- Difference in change from baseline between treatment groups in WOMAC Stiffness scores at Weeks 4, 8, 12, 16, 20 and 24
    - This analysis will be identical to the primary analysis.
- Difference in change from baseline between treatment groups in WOMAC Total scores at Weeks 4, 8, 12, 16, 20 and 24
    - This analysis will be identical to the primary analysis.
    - WOMAC Total is defined as the average of the three subscale scores (Pain, Stiffness, and Function).
- Difference between treatment groups in time to return to baseline WOMAC Pain

- This will be analyzed using a Cox proportional hazards regression with treatment, site (possibly grouped; see Section 4.2), and baseline WOMAC Pain as covariates. Kaplan-Meier median and quartiles will also be displayed.
- Time to return to baseline is defined as the first week, of two consecutive non-missing weeks, where the WOMAC Pain score is >= Baseline. For example, if a subject's WOMAC Pain scores are 5.0 (Baseline) and then:
  - 5.0 (Week 1), 4.8 (Week 2), 5.6 (Week 3), 5.0 (Week 4), 3.0 (Week 5), and 2.8 (Week 6), the return to baseline is Week 3 for this subject.
  - 5.6 (Week 1), 4.4 (Week 2), 4.2 (Week 3), 4.0 (Week 4), 3.0 (Week 5), and 2.0 (Week 6) (and then discontinued), the time to baseline pain for this subject is censored at Week 6.
  - 5.6 (Week 1), 6.0 (Week 2), 5.4 (Week 3), 4.6 (Week 4), and 4.4 (Week 5), the return to baseline is Week 1 for this subject.
  - 4.4 (Week 1), 5.2 (Week 2), Missing (Week 3), 5.6 (Week 4), and 3.2 (Week 5), the return to baseline is Week 2 for this subject.
  - Has 23 consecutive weeks of 2.0, and then 5.2 at Week 24, the time to baseline pain is censored at Week 24.

- Difference between treatment groups in area under the curve (AUC) of WOMAC Pain at Weeks 1−24
  - This analysis will be identical to that described for the key secondary endpoint for AUC of WOMAC Pain.
- Difference between treatment groups in AUC of average daily NPRS at Weeks 1-24
  - This analysis will be identical to that described for the key secondary endpoint for AUC of WOMAC Pain, however no missing data imputation will be done for this endpoint.
  - Average daily NPRS at Week 1 is defined as the average of non-missing NPRS pain scores from Day 1 through 8. For subsequent Weeks, average daily NPRS is defined as the average of the 7 days of non-missing daily NPRS data, allowing for the 8-day Week 1 (for example, Week 2 would be Days 9 through 15).
- Difference between treatment groups in AUC of WOMAC Function, Stiffness and Total scores at Weeks 4, 8, 12, 16, 20 and 24
  - This analysis will be identical to that described for the key secondary endpoint for AUC of WOMAC Pain, however no missing data imputation will be done for these endpoints.
- Difference between treatment groups in OMERACT-OARSI strict responders at Weeks 4, 8, 16, 20 and 24

  - This analysis will be identical to that described for the key secondary endpoint for OMERACT-OARSI strict responders at Week 12.
- Difference between treatment groups in PtGA at Weeks 4, 8, 12, 18 and 24
  - This analysis will be identical to the primary analysis.
- Difference between treatment groups in MDGA at Weeks 4, 8, 12, 18 and 24
  - This analysis will be identical to the primary analysis.
- Difference between treatment groups in SF-36 at Weeks 12 and 24
  - Each of the 8 SF-36 domains, as well as the two component scores (Mental and Physical), will be modeled separately using an ANCOVA model with treatment, site (possibly grouped; see Section 4.2), and covariate baseline score as covariates.
  - SF-36 scoring will be directly obtained from the QualityMetric PRO CoRE software.
- Difference between treatment groups in average rescue medication usage at Weeks 1-24
  - The weekly average frequency of rescue medication use will be based on the ePRO diary data. Rescue medication will be considered to be taken on each day where it is recorded that rescue medication was consumed for index knee pain. Weekly averages will be calculated based on actual days in the study. At least 4 non-missing rescue medication diary entries in a week must be available for the weekly average at that week to be calculated. Section 4.3 describes the imputation strategy for instances when less than 4 rescue medication entries are available at a given week. The weekly average frequency of rescue medication use will be analyzed at actual weeks with an ANCOVA model with site (possibly grouped; see Section 4.2), treatment, and baseline (pre-dose) weekly average frequency of rescue medication use as covariates.

    Depending on the distribution of the observed data, the weekly average frequency of rescue medication use may be transformed appropriately prior to fitting the model.
  - The average daily dosage (mg) of rescue medication will be based on information from the rescue medication accountability eCRF pages only. Average daily dosage (mg) of rescue medication will be calculated between each observed scheduled visit. This will be calculated by multiplying 500 mg by the number of tablets within a bottle consumed divided by the number of days between visits associated with the bottle. Note that at each visit, rescue medication is returned and accounted for from the previous observed visit and new medication is potentially both dispensed and re-dispensed. It will be assumed that rescue medication taken on the day of return/dispensing will be from the newly dispensed/re-dispensed medication. For a

given observed visit, the number of days between visits will be calculated as: start date of current visit – start date of previous visit. If an interim visit is missing but dispensing and return information is provided at the previous and later visits, then rescue medication usage at the later visit will be calculated according to the period between the two observed visits. The number of tablets consumed will be calculated from the eCRF data as: number of tablets dispensed/re-dispensed – number of tablets in returned kit – number of tablets not accounted for. If a subject marks "unknown" for number of not accounted for tablets, then the number of consumed tablets cannot be calculated and, therefore, the average daily dose of rescue medication for that period will be considered missing. Rescue medication may be calculated for following visits if all dispensing and return information is provided for those periods. If no medication is dispensed at a given visit, then no medication is assumed to have been taken between that visit and the next or until new rescue medication is dispensed, if any. If this occurs, then the daily average dosage (mg) will be considered 0 while the subject is in the study.

The average daily dosage (mg) will be analyzed at each post-baseline scheduled visit with an ANCOVA model with site (possibly grouped; see Section 4.2, treatment, and baseline average daily dosage (mg) as covariates.

- Difference between treatment groups in average activity levels at Weeks 1-24
  - This analysis will be analogous to the average rescue medication endpoint above.
  - Average daily activity level will be transformed to a 0-1 scale, where 0=Low, 0.5=Moderate and 1=High.
- Difference between treatment groups in WOMAC Pain responder categories at Week 4, Week 8, Week 12, Week 18, and Week 24
  - Subjects will be considered to be an X% responder if they have at least an X% decrease from baseline WOMAC Pain.
  - Separate summaries for 30% response, 50% response and 70% response will be presented. Each summary will be done by time point and treatment, and will include the number of subjects with data, the number and percentage of responders as well as the P-value from Fisher's exact test comparing proportions of responders between treatment and vehicle.

3.8.3 Exploratory Efficacy Analyses

Subgroup analyses will be conducted to explore treatment differences within baseline characteristics age, gender, BMI, baseline pain, symptom duration, Kellgren-Lawrence (KL)

grade, and disease type. If >10% of the ITT population received less than 20mg of EP-104, an additional subgroup analysis of at least, or less than 20mg EP-104 will be included. An ANCOVA model with treatment and baseline WOMAC pain as covariates will be estimated for each subgroup factor at the time points Week 4, Week 8, Week 12, Week 18, and Week 24, and these analyses will be based on the ITT population. Subgroups will consist of the following:

- Age: >= or < 65 years
- Gender: Male or female
- BMI: >= or < 30 kg/m2
- Baseline pain: Split by median WOMAC Pain
- Symptom duration: Split by median years
- KL Grade: Grade 2 or Grade 3
- Disease type: Unilateral or bilateral OA
- Dose received: <20mg or >=20mg (only performed if the <20mg group contains greater than 10% of the total ITT population)

Perceptions of treatment received/administered will be analyzed in the following way:

- Difference between treatment groups in patient beliefs about treatment received
    - A summary table will be used to explore the proportion of patients, within each responder category, that believe they received EP-104IAR vs. placebo. Responder categories will be based on Week 12 change from baseline WOMAC Pain as described in Section 3.8.2.
    - A summary table with reasons for belief will also be provided.
- Difference between treatment groups in physician beliefs about treatment received
    - This summary will be identical to that above, with responder categories instead based on Week 12 change from baseline MDGA. Subjects will be considered to be an X% responder if they have at least an X% decrease from baseline MDGA. Those with at least 30% decrease, at least 50% decrease and at least 70% decrease will be presented in the summary table.

3.8.4 Safety Analyses

Safety data (safety labs, cortisol, physical exam, knee exams, vital signs, ACTH stimulation test results, etc.) will be summarized for the safety population using standard summary statistics, by

treatment group. Both by time point data and change from baseline, where appropriate, will be assessed for continuous parameters. Shifts from baseline, where appropriate, will be summarized for categorical parameters. Figures will be used to display the time profiles of serum cortisol and glucose. No prospective hypothesis testing will be performed on safety parameters.

3.8.4.1 Adverse Events

An adverse event (AE) is any untoward medical occurrence in a subject administered a study product and which does not necessarily have a causal relationship with this treatment. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not related to the medicinal (investigational) product.

The AE reporting period begins when the subject signs the informed consent form and ends at Visit 9 (Week 24/End-of-Study/Early Exit Visit). All AEs will be captured in the eCRF. This applies to all AEs regardless of the presumed relationship to the study treatment. The verbatim term for adverse events will be classified by system organ class (SOC) and preferred term (PT). Pre-treatment AEs will be summarized overall by count and percentage. Treatment-emergent adverse events (TEAEs) will be summarized and analyzed for safety evaluations. All adverse events, including non-treatment-emergent, will be included in listings.

The summary of TEAEs will include frequencies and percentages and will be presented by treatment for the following:

- Overview of AEs
- All TEAEs by SOC and PT
- Serious TEAEs by SOC and PT
- TEAEs related to EP-104IAR by SOC and PT
- TEAEs by SOC, PT, and severity
- TEAEs of special interest by SOC and PT
- TEAEs leading to study discontinuation by SOC and PT
- TEAEs leading to death by SOC and PT.
- MRI-Related TEAEs by SOC and PT

An adverse event will be considered related to study drug if the relationship is “possible”, “probable/likely” or “definite”. The severity of an adverse event will be rated on the CRF as “Mild”, “Moderate”, or “Severe”. For the summary by severity, subjects who have multiple

occurrences of the same AE will be classified according to the worst reported severity of the AE. Similarly, for the summary by relationship to EP-104IAR, the AE will be classified according to the most related relationship.

Potential adverse events of special interest will be assessed as part of an initial blinded review process as well as an unblinded medical monitor review process with a final designation of AESIs made following database lock and a Sponsor review of study data.

3.8.4.2 Clinical Laboratory Assessments

Clinical laboratory assessments will be summarized using standard summary statistics for continuous parameters and using frequencies and percentages for categorial parameters. Results will be listed for each subject and flagged high or low relative to the normal range where appropriate. Clinical laboratory test results will be listed in SI units (Metric). Concentration values below the limit of quantification (BLQ) will be analyzed at the lower limit of quantification (LLOQ) when calculating summary statistics.

Clinical Chemistry and Hematology

Clinical chemistry and hematology values will be presented at Screening, Baseline, Week 1, Week 2, Week 4, Week 12 and Week 24 (End of Study) or Early Termination. HbA1c will be presented at Screening and at Week 24 (End of Study) or Early Termination visit.

Serum Cortisol

Serum cortisol will be presented at Screening, Baseline, Week 1, Week 2, Week 4, Week 8, Week 12, Week 18, and Week 24 (End of Study) or Early Termination.

ACTH Stimulation Tests

Results for ACTH stimulation tests will be summarized by frequency and percentage of positive tests, and listed for each subject.

Urine Pregnancy and Urinalysis Dipstick Tests

Urinalysis dipstick and safety, and urine pregnancy test results will be presented in data listings and the urinalysis drug and safety test results will be summarized in shifts from baseline tables.

3.8.4.3 Vital Signs

Vital sign measurements will be summarized at Screening, Baseline and 1-hour post-dose, Week 2, Week 4, Week 8, Week 12, Week 18, and Week 24 (End of Study) or Early Termination. Vital signs include systolic and diastolic blood pressure (SBP and DBP), heart rate (HR), respiration rate (RR) and body temperature. Weight will also be summarized per its collection schedule at Screening, Baseline, Week 12, and Week 24 (End of Study) or Early Termination. Vital signs will be summarized as absolute values and change from baseline using standard summary statistics.

3.8.4.4 Physical Examination

Physical examinations will be presented at Screening, Baseline, Week 12, and Week 24 (End of Study) or Early Termination. Physical examinations include assessments of general appearance, skin, HEENT (head, ears, eyes, nose and throat), chest and lungs, cardiovascular, abdomen, musculoskeletal and extremities, neurological, psychiatric and mental status.

Knee examinations will be done at Screening, Baseline, Week 2, Week 4, Week 8, Week 12, Week 18, and Week 24 (End of Study) or Early Termination. Knee examinations will be done for both knees at Screening and only for the Index Knee for subsequent visits. Knee examinations include range of motion, presence of an effusion (stroke test), crepitus on palpation with range of motion, stability of collateral ligaments, stability of cruciate ligaments, warmth on palpation of the joint, tenderness on palpation of the joint, gait, and skin appearance.

Both physical examination and knee examination parameters will be listed for each subject and summarized by frequency and percentage, as well as by shift from baseline.

3.9 Analysis of Plasma Pharmacokinetic Concentrations

The pharmacokinetics of EP-104IAR will be assessed by analysis of plasma levels of fluticasone propionate (FP). Blood samples for determination of FP plasma concentrations will be collected at following time points relative to study drug administration: Pre-Dose, 2 hours Post-Dose, Week 1, Week 2, Week 4, Week 8, Week 12, Week 18 and Week 24 (End of Study). PK data will be analyzed using the Safety Population.

Where estimable, the following plasma PK parameters calculated for FP will be calculated using non-compartmental methods:

- the maximum quantifiable plasma concentration (C_{max})
- the time to reach maximum quantifiable plasma concentration (t_{max})

- the last quantifiable plasma concentration (C_{last})
- the time to reach the last quantifiable plasma concentration (t_{last})
- the terminal disposition rate constant (λ_z) with the respective half-life ($t_{1/2}$)
- the area under the concentration-time curve from time zero to the time of last quantifiable concentration (AUC_{0-t})
- the area under the concentration-time curve from zero to infinity ($AUC_{0-\infty}$)
- the area under the concentration-time curve from time zero to 24 weeks post-dose ($AUC_{0-24wks}$)
- the percentage of the area under the concentration-time curve extrapolated to infinity ($AUC_{\%extrap}$)
- the apparent clearance (CL/F)
- the apparent volume of distribution (V_z/F).

The individual PK parameters of FP will be presented in data listings and summarized for subjects receiving EP-104IAR using descriptive statistics (number of subjects, mean, SD, geometric mean, geometric mean CV, median, Q1, Q3, minimum, and maximum). Geometric mean versus nominal time profiles will be presented in figures on both linear and semi-logarithmic scales (base 10). Concentration values that are BLQ of the assay will be treated as half the LLOQ when calculating summary statistics for plasma concentrations. Where all values are BLQ, the mean, geometric mean, median, minimum, and maximum concentration will be presented as BLQ and the SD will be reported as not applicable. Concentration values reported as BLQ will be listed with the LLOQ in parentheses. Missing values will not be imputed when calculating summary statistics for plasma drug concentrations and for calculating PK parameters.

For the derivation of PK parameters, pre-dose plasma concentration results will be analyzed as time 0 (hours post-dose) and all post-dose plasma concentration results will be analyzed according to their actual sampling times relative to dosing. Concentration data will be used as received from the bioanalytical laboratory without rounding.

Residual EP-104IAR in used injection kits will be used to calculate actual doses received. From this, dose proportionality for PK parameters (AUC and C_{max}) will be assessed using the power model: $P = \beta_0 \times Dose^{\beta_1}$, where P represents the parameter result. The natural log of the power model will be used to fit the data using an analysis of variance with effect dose: $ln(P) = ln(\beta_0) + \beta_1 \times \ln(dose)$. An estimate of β_1=1 would indicate perfect dose proportionality. The slope parameter β_1 and its associated confidence interval will be estimated and summarized. Additionally, the PK parameter (AUC and C_{max}) multiplication factor associated with a 2-fold

increase in dose along with its corresponding confidence interval will be included in summary table. The multiplication factor and associated confidence interval will be estimated as $2^{\beta 1}$ and ($2^{\text{lower confidence limit of } \beta 1}$, $2^{\text{upper confidence limit of } \beta 1}$), respectively.

PK parameters will be derived as follows. PK parameters containing a time component for their unit of measure are presented below using hours (h). However, a different unit of time (e.g., days, weeks) may be utilized instead depending on the scale of the parameter.

C_{max} (pg/mL)

C_{max} will be the maximum quantifiable plasma concentration post-dose of study drug.

t_{max} (h)

t_{max} will be the earliest actual time in hours post-dose of study drug associated with C_{max}.

C_{last} (pg/mL)

C_{last} will be the last quantifiable plasma concentration post-dose of study drug.

t_{last} (h)

T_{last} will be the actual time in hours post-dose of study drug associated with C_{last}.

λ_z (/h)

The terminal phase elimination rate constant will be calculated using concentrations from the elimination phase (i.e., after t_{max}). Only quantifiable measurements occurring after C_{max}, but not including C_{max}, will be considered. If any of these quantifiable concentrations occur after an elimination phase concentration that was below the LLOQ, the concentration will be examined by the pharmacokineticist who will determine if the concentration should be included in the calculation. Any such case will be documented in the Clinical Study Report (CSR).

The terminal phase elimination rate constant will be considered unknown if two or fewer concentrations are available. If three or more concentrations are available, then slopes will be estimated for natural log-concentrations over time. The linear regression slope will be calculated

using the last three quantifiable measurements, again for the last four measurements, again for the last five, and so on. The absolute value of the slope associated with the fitted line having the largest adjusted R^2 value will be used as the terminal elimination rate constant unless there exists an adjusted R^2 value within 0.0001 of the largest, but utilizes a larger number of points, in which case, the regression with the larger number of points is used. If the unadjusted R^2 value is less than 0.8, then λ_z will be considered unreliable and the λ_z value as well as the associated pharmacokinetic parameters $t_{1/2}$, $AUC_{0-\infty}$, $AUC_{\%extrap}$, CL/F, and V_z/F will be reported in the listings but excluded from descriptive and statistical analyses.

$t_{1/2}$ (h)

The terminal elimination half-life will be calculated as ln (2)/λ_z, where ln is the natural logarithm.

AUC_{0-t} (h•pg/mL)

AUC_{0-t} will be calculated using the linear trapezoidal method. Pre-dose concentrations will be analyzed as time 0. Concentrations below the LLOQ occurring before the first quantifiable concentration will be analyzed as half the LLOQ. Concentrations below the LLOQ occurring after the first quantifiable concentration will be considered missing. Similar to the calculation of λ_z, if a quantifiable measurement during the elimination phase occurs after an elimination phase concentration that is below the LLOQ, the concentration will be examined by the pharmacokineticist who will determine if the concentration should be included in the calculation.

$AUC_{0-24wks}$ (h•pg/mL)

$AUC_{0-24wks}$ will be calculated using the linear trapezoidal method. Pre-dose concentrations will be analyzed as time 0. Concentrations below the LLOQ occurring before the first quantifiable concentration will be analyzed as half the LLOQ. Concentrations below the LLOQ occurring after the first quantifiable concentration will be considered missing. If there is no Week 24 collection, the Week 24 collection does not occur at precisely 24 weeks post-dose, or the Week 24 result is below the LLOQ, the Week 24 plasma concentration level may be estimated as follows:

- If there are quantifiable plasma concentrations before and after Week 24, then the Week 24 concentration will be estimated using linear interpolation.

- If there are no quantifiable plasma concentrations after Week 24, but λ_z is estimable with the unadjusted R^2 being greater than or equal to 0.8, then the Week 24 plasma concentration will be estimated by:

  Week 24 plasma concentration = C_{last}*exp[-λ_z*(24*7*24 - T_{last})]

  Where C_{last} is the last concentration used in the calculation of AUC_{0-t}. and T_{last} is its associated time (h) post-dose.

- If there are no quantifiable plasma concentrations after Week 24 and λ_z was not estimable or reliable, then $AUC_{0-24wks}$ will be analyzed using the result for AUC_{0-t} if C_{last}*(24*7*24 - T_{last}) is less than 1% of AUC_{0-t}.
- If none of the above criteria were met, then the Week 24 concentration and $AUC_{0-24wks}$ will not be estimated.

$AUC_{0-\infty}$ (h•pg/mL)

$AUC_{0-\infty}$ will be calculated as:

$$AUC_{0-\infty} = AUC_{0-t} + \frac{C_{last}}{\lambda_z}$$

Where C_{last} is the last concentration used in the calculation of AUC_{0-t}.

$AUC_{\%extrap}$ (%)

$AUC_{\%extrap}$ will be calculated as:

$$AUC_{\%extrap} = 100 * \frac{AUC_{0-\infty} - AUC_{0-t}}{AUC_{0-\infty}}$$

CL/F (L/h)

CL/F will be calculated as:

$$CL/F = 1000^2 * \frac{Dosage\ in\ mg}{AUC_{0-\infty}}$$

Where dosage is the actual dosage calculated from the residual EP-104IAR in the used injection kits.

V_z/F (L)

V_z/F will be calculated as:

$$V_z/F = \frac{CL/F}{\lambda_z}$$

4. DATA HANDLING

4.1 Screening, Baseline, Pre-Dose, and Study Visits

Except for PK and WOMAC data, which will be analyzed at the actual sampling times, all other analyses will be presented based on the planned schedule of assessments in the protocol and recorded on the eCRFs and ePRO devices.

WOMAC pain post-baseline data (including subscale and associated questions) will be analyzed based on actual week from dosing per the mapping rules below:

Actual Week	Target Day	Study Day Range
Week 1	Day 8	7-13
Week 2	Day 15	14-20
Week 3	Day 22	21-27
Week 4	Day 29	28-34
Week 5	Day 36	35-41
Week 6	Day 43	42-48
Week 7	Day 50	49-55
Week 8	Day 57	56-62
Week 9	Day 64	63-69
Week 10	Day 71	70-76
Week 11	Day 78	77-83
Week 12	Day 85	84-90
Week 13	Day 92	91-97
Week 14	Day 99	98-104

Week 15	Day 106	105-111
Week 16	Day 113	112-118
Week 17	Day 120	119-125
Week 18	Day 127	126-132
Week 19	Day 134	133-139
Week 20	Day 141	140-146
Week 21	Day 148	147-153
Week 22	Day 155	154-160
Week 23	Day 162	161-164
Week 24	Day 169	165-174

WOMAC stiffness, WOMAC function, and WOMAC total post-baseline data (including subscales and associated questions) will be analyzed based actual week from dosing per the mapping rules below:

Actual Week	Target Day	Study Day Range
Week 4	Day 29	7-43
Week 8	Day 57	44-71
Week 12	Day 85	72-99
Week 16	Day 113	100-127
Week 20	Day 141	128-155
Week 24	Day 169	156-183

If more than one WOMAC data point is mapped to the same actual week, the one data point that is closest to the target day at the time of dosing will be used. Any data points that fall outside the windowing study day range will not be used.

Unless stated otherwise, baseline will be the last non-missing value, including unscheduled, prior to study drug administration. See Section 3.7 for the definition of baseline WOMAC pain. For efficacy data that are collected daily (NPRS scores, rescue medication usage, and physical activity levels), baseline is defined as the average of the available 7 days prior to dosing. Screening, when summarized, will be the last non-missing value, including unscheduled, that is prior to the day of study drug administration.

4.2 Site Grouping Strategy

In some analyses site will be used as a covariate, and in these cases, sites will be pooled within country by size, smallest to largest, until there are at least 16 subjects from within the ITT population included in each pooled group.

4.3 Missing Data

Handling of missing data for assessing the primary and secondary endpoints is discussed in Sections 3.7, 3.8 and 4.5. Handling of plasma/serum values that are BLQ with respect to the pharmacokinetic analyses are described in Section 3.9. Additional considerations include:

- For the derivation of duration of symptoms in the index knee (years) and time since diagnosis in the index knee (years), partial dates will be imputed with the following rules:
    - When only the year is available (both month and day are missing), then the month will be imputed as July and the day will be imputed as 1.
    - When the month and year are available (only day is missing), then the day will be imputed as 15.
    - When only the year is available and it is the same year as administration of study drug, the duration in days will be calculated as the day of the year study drug was administered divided by two.

- Intermittent missing ePRO-collected diary data for rescue medication, activity levels and daily pain (NPRS) for post-baseline weeks will be imputed in instances when less than 4 entries are available at a given week. Last observation carried forward (LOCF) will be applied to carry forward the most recent post-baseline diary entries so that 4 entries are made available at a week. For example, if a subject has only 1 entry at a given week, the 3 most recent entries from previous post-baseline weeks will be carried forward to compute a weekly average out of 4 days. Similarly, if a subject has only 3 entries at a given week, only the 1 most recent entry from previous weeks will be carried forward to compute a weekly average out of 4 days.
- For the SF-36 survey, PRO CoRE will be used to impute missing values via proration, whereby the missing responses will be replaced with the mean of the observed scores of the completed questions within the same scale (per subject and timepoint). PRO CoRE will only be used in instances where at least ½ items in each of the eight scales are non-missing. If less than ½ items are completed within a scale or the complete scale or survey is missing, no imputation will be applied.

4.4 Unscheduled Data

Except for ACTH stimulation test results as described below, data obtained from unscheduled assessments (which also includes early exit visits) will only be considered for baseline determination; otherwise, unscheduled data will not be used in summary tables or included in the analyses. All unscheduled data will be presented in the listings.

ACTH stimulation test results will be mapped via the windowing rules in the below table. When more than one result for a subject occurs within one visit window, the worst result (i.e., Fail) will be mapped to that visit.

Visit (Mapped)	Target Day	Study Day Range
Screening		<1
Week 1	Day 3	1-9
Week 2	Day 15	10-22
Week 4	Day 29	23-43
Week 8	Day 57	44-71
Week 12	Day 85	72-106
Week 18	Day 127	107-148
Week 24	Day 169	149+

4.5 WOMAC Score Derivations

The WOMAC subscale scores are going to be calculated as follows:

Pain Subscale Score: It is defined as the average of 5 pain related questions (labeled in data as Section A). The pain score will be calculated if there are at least 4 available data points, and the score will be based on the average of the available data points. If there are 2 or more of the data points are missing, then the pain score will be missing.

Stiffness Subscale Score: It is defined as the average of 2 stiffness related questions (labeled in data as Section B). The stiffness score will be calculated if there is at least 1 available data point, and the score will equal the available data point. If both data points are missing, then the stiffness score will be missing.

Function Subscale Score: It is defined as the average of 17 function related questions (labeled in data as Section C). The function score will be calculated if there are at least 14 available data points, and the score will be based on the average of the available data points. If there are 4 or more of the data points are missing, then the function score will be missing.

Total Score: If any of the subscale scores are missing, then the total score will also be missing. Otherwise, total score will be calculated as the average of the 3 subscale scores. If any of the subscales were calculated in the presence of the missing data point, then the missing data point will first be imputed as the average of the rest of the questions within that scale, and then the total score will be calculated.

4.6 Presentation Conventions

For all parameters, where applicable:

- n: no decimal place
- CV% geometric mean will be displayed to one decimal place
- Geometric mean ratios and associated confidence intervals to two decimal places.

For the following PK parameters, where applicable:

- Arithmetic mean, geometric mean, SD, median, minimum, maximum, and listings:
  - t_{last}, t_{max}: two decimal places (may be changed depending on the unit of time)
  - λ_z: three significant figures
  - $t_{1/2}$, CL/F, V_z/F: three significant figures
  - C_{max}, C_{last}, $AUC_{0-\infty}$, AUC_{0-t}, $AUC_{\%extrap}$: three significant figures.

All other parameters that are reported or derived at a fixed number of decimal points (e.g., vital sign measurements, including body mass index (BMI), clinical laboratory test parameters, and plasma/serum drug concentrations), will be reported as follows:

- Mean, median, Q_1 and Q_3 will be displayed to one more decimal than the maximum number of decimal places reported for the original data
- SD will be reported to two more decimal places than the maximum number of decimal places reported for the original data
- Minimum and maximum will be reported to the maximum number of decimal places reported for the original data.

4.6.1 Significant Figures

When descriptive statistics are to be presented to a specific number of significant figures, the following conventions will be utilized:

- Results of "0" will be displayed as "0".

- Results that cannot be displayed to the exact specified number of significant figures (in Section 3.5) without the use of scientific notation will be displayed using the maximum number of significant figures that is less than the specified number of significant figures. E.g., "1000" and "1295" cannot be displayed to three significant figures without the use of scientific notation, so "1000" and "1300" would be presented, respectively.
- All other results will be displayed to the specified number of significant figures in Section 3.5. Examples of results displayed to three significant figures are presented below:

Original Result	Presentation Result (3 Significant Figures)
12.34	12.3
0.01234	0.0123
0.12	0.120
10	10.0
1234	1230

5. DEVIATIONS FROM THE PROTOCOL

The Full Analysis Set (FAS) Population is defined in the protocol but is omitted from this document, as it is not used for any of the analyses.

Two additional exploratory efficacy endpoints are included in this document but are not in the protocol:

- Difference between treatment groups in patient beliefs about treatment received
- Difference between treatment groups in physician beliefs about treatment received.

Three additional secondary efficacy analyses were added whereby WOMAC Pain responder levels (30, 50, 70) are summarized at Weeks 12, 18 and 24.

The analyses of perceptions of treatment received/administered are expanded to consider summaries by response type.

The intention-to-treat (ITT) population definition is updated to all subjects who are both randomized to and receive treatment. Analysis will be conducted allocating subjects to the treatment to which they were randomized, irrespective of what was actually received.

An MRI sub-study population was added as a study population to reflect the sub-study for the imaging.