STATISTICAL ANALYSIS PLAN

STUDY SPONSOR:	Buzzz Pharmaceuticals Limited 15 Main Street Raheny, Dublin 5, Ireland Tel: +44-773-310-4786
SPONSOR REPRESENTATIVE:	Dr. Deirdre O'Keeffe Head of IP and Portfolio Management
PROTOCOL TITLE:	Phase 2 Multicenter, Double-Blind, Randomized, Parallel-Group, Vehicle-Controlled Study to Evaluate the Efficacy, Safety, and Local Tolerability of G001 in Patients with Osteoarthritis (OA) of the Knee
PROTOCOL NUMBER:	2020-G001-P2 *(ClinicalTrials.gov ID: NCT05007808)*
PROTOCOL VERSION AND DATE:	Version 4.0 (December 7, 2022)
NAME OF TEST DRUG:	G001 (Celecoxib Gel 4% for Topical Administration)
PHASE:	Phase 2
METHODOLOGY:	Randomized, Double-Blind, Vehicle-Controlled, 2-Arm
ANALYSIS PLAN DATE:	30 March 2023
ANALYSIS PLAN VERSION:	Version 1.0
AUTHOR:	Julius Kirui

APPROVAL SIGNATURE PAGE

Protocol Title:	Phase 2 Multicenter, Double-Blind, Randomized, Parallel-Group, Vehicle-Controlled Study to Evaluate the Efficacy, Safety, and Local Tolerability of G001 in Patients with Osteoarthritis (OA) of the Knee
Sponsor:	Buzzz Pharmaceuticals Limited 15 Main Street, Raheny, Dublin 5, Ireland
Protocol Number:	2020-G001-P2
Document Date / Version:	30 March 2023 / Version 1.0
Veristat, Inc. Author:	134 Turnpike Road, Suite 200 Southborough, MA 01772

Approval Signature	Job Title
DocuSigned by: Julius Kirui Signer Name: Julius Kirui Signing Reason: I am the author of this document Signing Time: 30-Mar-2023 | 11:07:27 AM EDT EE88FE850D974568AD83AA0AFF1AFA3B	Biostatistician

Sponsor Signatory:	15 Main Street Raheny, Dublin 5, Ireland

Sponsor Approval

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

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

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

Approval Signature	Job Title
DocuSigned by: Deirdre O'Keeffe Signer Name: Deirdre O'Keeffe Signing Reason: I approve this document Signing Time: 31-Mar-2023 | 7:11:50 AM PDT 706E6D1324934B7285C7A10C4BF6B5DB	Head of IP and Portfolio Management

TABLE OF CONTENTS

TABLES INCLUDED IN THE TEXT

Page

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Abbreviation	Definition
ACR	American College of Rheumatology
AE	Adverse event
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
ATC	Anatomic therapeutic class
BMI	Body mass index
BL	Baseline
CI	Confidence interval
CMH	Cochran-Mantel-Haenszel
CRF	Case report form
CSR	Clinical study report
ECG	Electrocardiogram
EDC	Electronic Data Capture
EOS	End-of-study
EOT	End-of-treatment
FPFV	First patient first visit
FV	Flare visit
HR	Heart rate
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IGA	Investigator Global Assessment
IMP	Investigational medicinal product
IWRS	Interactive Web Response System
LOCF	Last observation carried forward
LPLV	Last patient last visit
LS	Least square
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified intent-to-treat
NRS	Numerical Rating Scale
NSAID	Nonsteroidal anti-inflammatory drug
OA	Osteoarthritis
PGA	Patient Global Assessment
PP	Per Protocol
PP1	Per Protocol 1
PP2	Per Protocol 2
PT	Preferred term

Abbreviation	Definition
Rel Day	Relative study day
SAP	Statistical analysis plan
SD	Standard deviation
SI	International System of Units
SOC	System organ class
TEAE	Treatment-emergent adverse event
ULN	Upper limit normal
WHO	World Health Organization
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index

1. INFORMATION FROM THE STUDY PROTOCOL

1.1. Introduction and Objectives

1.1.1. Introduction

Osteoarthritis (OA) is a degenerative joint disease that is highly prevalent, and a leading cause of disability around the globe (Cross et al. 2014; GBD 2016). Clinically, the knee is the most common site of OA (accounting for approximately 85% of the OA burden worldwide), followed by the hand and hip (Hunter and Bierma-Zeinstra, 2019).

Osteoarthritis is a heterogeneous disease with a very complex pathology that involves mechanical, inflammatory, and metabolic factors, which ultimately lead to structural destruction and failure of the synovial joint. Pain is the most disabling symptom in OA, and a major driver of clinical decision making and health service use (Hunter and Bierma-Zeinstra, 2019).

Of the pharmacological interventions for the management of OA pain, medications with the least systemic exposure (i.e., local therapy) are preferred. Strongly recommended medications for the treatment of knee OA include topical nonsteroidal anti-inflammatory drugs (NSAIDs) as first choice (to be considered prior to use of oral NSAIDs, particularly in patients with more limited disease), oral NSAIDs, and intra-articular glucocorticoid injections, whereas topical capsaicin, acetaminophen, duloxetine, and tramadol are conditionally recommended (Bannuru et al. 2019; Kolasinski et al. 2020). Topical NSAIDs have similar efficacy to that of oral NSAIDs; however, due to the low systemic absorption, topical NSAIDs have a much better safety profile (Rannou et al. 2016). Celecoxib (4-[5-(4-methylphenyl)-3-(trifluoromethyl)-1H-pyrazol-1-yl] benzenesulfonamide), a cyclooxygenase (COX)-2 selective inhibitor, belongs to the pharmacological class of NSAIDs. Celecoxib has established analgesic, anti-inflammatory, and antipyretic properties (Celebrex CPM, 2019; Celebrex USPI, 2019).

G001 is a novel gel formulation of celecoxib for topical administration being developed by Buzzz Pharmaceuticals Ltd., Ireland for the treatment of signs and symptoms of OA of the knee and hand. In a Phase 1 double-blind, randomized, parallel-group study (Study 2019-4679), G001 (dose strengths of 1.5% and 4%) was generally well tolerated when administered under single-dose (N=18) or multiple dose conditions (N=48) in adult healthy volunteers. The current Phase 2 study will provide the first evaluation of the efficacy, safety, and local tolerability of a 4-week dosing regimen of G001 (dose strength 4%) in patients with OA.

1.1.2. Study Objectives

The objective of this study is to evaluate the efficacy, safety, and local tolerability of G001 compared to vehicle in patients with symptomatic OA of the knee.

1.1.2.1. Efficacy Objectives

Primary Efficacy Objective:

To evaluate the efficacy of G001 compared to vehicle in patients with symptomatic OA of the knee, as measured by the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) Pain Subscale Score.

Secondary Efficacy Objective:

To further evaluate the efficacy of G001 compared to vehicle in patients with symptomatic OA of the knee, as measured by the WOMAC Physical Function Subscale, Stiffness Subscale, and Total scores; 11-point pain numerical rating scale (NRS) scores, patient global assessment (PGA) and investigator global assessment (IGA) of disease activity and overall treatment benefit, and rescue medication use.

1.1.2.2. Safety Objectives:

The safety objectives of this study include:

- To evaluate the overall safety of G001 in patients with symptomatic OA of the knee, as determined by adverse event (AE) reporting, vital signs, electrocardiogram (ECG) measurements, and physical examinations

- To evaluate the local tolerability of G001 in patients with symptomatic OA of the knee, as determined by AE reporting, and skin irritation test scores

1.1.3. Purpose of this Document

This statistical analysis plan (SAP) is designed to outline the methods to be used in the analysis of study data in order to answer the study objectives. Populations for analysis, data handling rules, statistical methods, and formats for data presentation are provided. The statistical analyses and summary tabulations described in this SAP will provide the basis for the results sections of the clinical study report (CSR) for this trial.

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

1.2. Study Design

1.2.1. Synopsis of Study Design

Study 2020-G001-P2 is a Phase 2 multicenter, double-blind, randomized, parallel-group, vehicle-controlled study. A total of 210 patients with primary OA of the knee and moderate OA pain are planned to be enrolled at approximately 10 centers located in Canada (Quebec and Ontario).

Following screening, prior NSAID and/or acetaminophen use will be discontinued to allow for washout (3 to 7 days) and symptom flare. Eligible patients (patients who meet all of inclusion criteria and none of the exclusion criteria) with adequate OA (flare) pain (where adequate OA pain is defined in Inclusion Criterion 10 of the protocol) in the index knee at the Baseline/Flare

Visit #1 (BL/FV1) will be randomly assigned to one of two treatment groups (G001 4.0% or Vehicle) in a 1:1 allocation ratio.

Patients will have a demonstration of the study drug application at the clinic and receive their first dose of the investigational medicinal product (IMP), either G001 or Vehicle (reference product), before being sent home. Patients will be instructed to apply 4 grams of the IMP to the index knee (i.e., knee selected for study treatment) four times daily for four weeks. Patients will also be instructed to rate their worst daytime and nighttime pain in their Treatment Diary, as well as to document the date and time of each IMP application, and any rescue medication (acetaminophen) use.

Rescue medication use is discouraged, but acetaminophen may be used, as needed, at a dose not exceeding 500 mg per dose and 2,000 mg/day to manage breakthrough pain. No rescue medication is allowed within 2 hours after each IMP application, and within 12 hours prior to an efficacy assessment.

Patients will return to the clinic twice during the treatment period for efficacy and safety assessments: at Week 2 and at Week 4/End-of-Treatment (EOT). Patients will return to the clinic 3 to 7 days after the last IMP application for a second Flare Visit (Week 5/FV2), and within approximately 2 weeks after the last IMP application (Week 6) for an end-of-study (EOS) evaluation.

Each patient's participation will be approximately 7 weeks (~1 week screening, 4 weeks of treatment, and 2 weeks of post-treatment follow-up). The overall study duration (from first patient first visit [FPFV] to last patient last visit [LPLV]) is expected to be about 6 months.

1.2.2. Randomization Methodology

Patients will be evaluated for eligibility at the Screening Visit, and at the BL/FV1 prior to randomization. Eligible patients with adequate OA pain in the index knee at BL/FV1 will be randomized. Randomization will occur centrally, based on a computer-generated randomization schedule, and using an Interactive Web Response System (IWRS) within the electronic data capture (EDC) system.

A total of 210 patients will be randomized at a 1:1 ratio to receive either G001 or Vehicle in order to achieve a total sample size of 174 evaluable patients (87 per arm), where evaluable is defined as randomized patients who have received at least one dose of study drug and provided both BL/FV1 and Week 4/EOT efficacy assessments (i.e., the modified Intent-to-Treat [mITT]) Population, as defined in Section 2.1). This sample size has been adjusted for an up to 20% drop-out rate (prior to the Week 4 efficacy assessment), which was estimated based on available literature. Patients discontinuing the trial after randomization will not be replaced due to this enrollment overage.

1.2.3. Stopping Rules and Unblinding

1.2.3.1. Stopping Rules

Reasons for permanent discontinuation of study treatment may include any of the following:

- Interruption of study treatment for >3 consecutive days (e.g., due to an AE or other reason)
- Any medical condition that prevents the patient from safely completing the study
- Investigator or Sponsor determination that treatment discontinuation is in the best interest of the patient
- Use of/requirement for prohibited medication
- Patient non-compliance with the study treatment or assessments, as determined by the Investigator or the Sponsor
- Pregnancy
- Patient's decision / withdrawal of consent

Permanent discontinuation of study treatment may be considered if any of the following set of laboratory abnormalities indicating potential severe liver injury (possible Hy's Law) is met, regardless of the causality assessment of the event:

- Alanine aminotransferase (ALT) or aspartate aminotransferase (AST) ≥3x upper limit normal (ULN) and total bilirubin ≥2x ULN (>35% direct bilirubin)
- ALT or AST ≥3x ULN and international normalized ratio (INR) >1.5

Patients have the right to voluntarily withdraw from the study at any time for any reason. In addition, the Investigator has the right to withdraw a patient from the study at any time.

Reasons for patient withdrawal from the study may include, but are not limited to, the following:

- Patient withdrawal of consent
- AE
- Loss to follow-up (a patient is considered loss to follow-up if he or she repeatedly fails to return for scheduled visits and is unable to be contacted by the study site)
- Patient non-compliance with the study treatment or assessments
- Study termination or site closure

1.2.3.2. Blinding

The Investigator and other study site personnel involved in patient assessments (safety and efficacy), as well as patients will remain blinded to treatment assignment throughout the course of the study.

Because G001 and Vehicle formulations are physically distinguishable, at each site, a third party (such as a designated pharmacist or research nurse), who is not otherwise involved in the patient assessments will be responsible for overseeing all activities that pose a potential risk of unblinding, including IMP dispensation, demonstration of the IMP application, and ensuring all

IMP applications and discussion only occur in a completely private setting not accessible/visible to other staff or patients.

The Sponsor and its agents will also be blinded to treatment assignment prior to unblinding of the treatment assignment at the study level, with the exception of select individuals who require access to patient treatment assignments to fulfil their roles during a clinical trial, including those involved in the creation and maintenance of the IWRS and drug dispensation.

1.2.4. Study Procedures

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

Table 1 Schedule of Assessments

Study Visit	**Screening Visit**	**BL/FV1 Visit**	**Week 1 Phone call**	**Week 2 Visit**	**Week 3 Phone call**	**Week 4/EOT Visit**	**Week 5/FV2 Visit**	**Week 6/EOS Visit**
Study Day	**Day -7 to -3**	**Day 1**	**Day 8±1**	**Day 15±1**	**Day 22±1**	**Day 29±2**	**3-7 Days After EOT**	**Day 43±3**
Written informed consent [a]	X							
Demographic data [b]	X							
Medical and medication history [c]	X							
OA diagnosis (ACR criteria) [d]	X							
Radiological examination [e]	X							
WOMAC Index [f]	X	X		X		X	X	
PGA of disease activity [g]	X	X		X		X	X	
IGA of disease activity [g]	X	X		X		X	X	
PGA of overall treatment benefit [g]				X		X		
IGA of overall treatment benefit [g]				X		X		
Worst daily pain 11-point NRS (24-hour recall), index knee [h]	X	X						
Worst daily pain 11-point NRS (24-hour recall), contralateral knee [i]	X	X						
Physical examination [j]	X					X		X
Height, weight, and BMI	X							
Vital signs measurements [k]	X	X		X		X		X
Standard 12-lead ECG [l]	X					X		
Safety Laboratory Testing [m]	X	X		X		X		X
Urine pregnancy test [n]	X	X				X		X
Urine drug screen	X	X		X		X		X
Review of Inclusion / Exclusion criteria	X	X						
Randomization		X						
Study drug dispensing/reconciliation		X		X		X		
Study drug application (4 times per day) [o]		X	X	X	X	X		
Adverse event recording	X	X	X [p]	X	X [p]	X	X	X
Concomitant medications & therapies	X	X	X [p]	X	X [p]	X	X	X
Skin irritation assessment [q]				X		X		X

Patient diary completion [r]	*Completed by the patient on a daily basis*							
Review of the Patient Diary [s]		X	X [p]	X	X [p]	X	X	

Abbreviations: ACR=American College of Rheumatology; BL=Baseline; BMI=body mass index; BP=blood pressure; ECG=electrocardiogram; EOS=End-of-Study; EOT=End-of-Treatment; FV1=Flare Visit #1; FV2=Flare Visit #2; IGA=Investigator Global Assessment; NRS=Numerical Rating Scale; PGA=Patient Global Assessment; WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index

[a] Written informed consent is required before performing any study-specific tests or procedures.

[b] Demographics include age, gender, and self-reported race/ethnicity.

[c] Medical history OA history (date of diagnosis, prior treatment), general medical history, surgical history, and reproductive status. Medication history includes pharmacological treatment (oral, parenteral, intra-articular, or topical) used in the past 90 days, with focus on medications used for the management of OA, and prohibited medications; see Section 6.8.1 of the protocol.

[d] Documented diagnosis of OA of the knee, meeting ACR clinical criteria for classification of idiopathic (primary) OA (see Appendix 1 of the protocol) for at least 6 months prior to Screening is required for enrolment.

[e] Patients will undergo conventional radiological (X-ray) examination of the index knee during Screening. Exemptions include patients who: (1) had an X-ray of the index knee completed within 3 months prior to the Screening visit, and (2) received no treatment since that may have altered the radiological findings, including OA grade. Radiologic evidence of OA of the knee of grade 2 (mild) or grade 3 (moderate) according to Kellgren and Lawrence Radiographic Grading (see Appendix 2 of the protocol) is required for enrolment.

[f] Patients will complete the 5-point Likert format of the WOMAC® 3.1 Index; see Appendix 3 of the protocol. To qualify for randomization, all participants need to develop a flare of pain following washout (3-7 days) of stable analgesic (NSAID and/or acetaminophen) therapy, defined as (1) an increase of ≥2 points from Screening in the WOMAC Pain Subscale score for the index knee; and (2) a score of ≥2 points on at least one of the 5 items in the WOMAC Pain Subscale (for the index knee).

[g] PGA/IGA of disease activity (or PGA/IGA of treatment benefit) evaluates OA in the in the index knee over the last 48 hours on a 5-point Likert scale, where 0=very good; 1=good; 2=fair; 3=poor; and 4=very poor.

[h] Worst daily pain (within 24 hours prior to Screening and Baseline/FV1) in the index knee needs to be between 4 and 8 on the 11-point pain NRS for enrolment.

[i] Worst daily pain (within 24 hours prior to Baseline/FV1) in the contralateral knee needs to be ≤2 on the 11-point pain NRS scale for enrolment.

[j] Complete physical examination at Screening and Week 4/EOT, and symptom-driven examination at Week 6/EOS. A complete physical examination will include a review of all body systems, with special attention to the cardiovascular, gastrointestinal, and musculoskeletal systems, as well as skin and subcutaneous tissues.

[k] Vital signs measurements will include sitting (after 3 minutes of rest) blood pressure (BP), heart rate (HR), respiratory rate (RR), and body temperature. Two consecutive BP readings (at least 1 minute apart) will be recorded, and the average of the two readings will be entered in the eCRF.

[l] Standard 12-lead ECG, taken in a supine position, after resting for at least 5 minutes.

[m] Includes: Haematology, Biochemistry, Coagulation tests, Urinalysis; see Appendix 5 of the protocol.

[n] A confirmatory serum pregnancy test must be completed in case the urine test is positive. Patients with a positive pregnancy test during Screening will be excluded from the study. Patients with a positive pregnancy test during the treatment or post-treatment follow-up period will be withdrawn from the study.

[o] The first dose will be applied at the clinic, in the presence of a qualified unblinded treatment administrator; subsequent applications will be completed by the patient at the clinic (on Day 15/Week 2 Visit) or at their home (all other applications). Accurate dosing will be ensured by using a dosing card (to be used for all study drug applications). Patients will be dispensed a two-week supply of study drug (three tubes of 100 grams) on Day 1 (Baseline/FV1) and on Day 15 (Week 2 Visit), and will be asked to return all unused study drug supplies at the next study visit.

[p] The Investigator (or qualified designee) will contact the patient by phone after one week of treatment (Week 1, study Day 8 ± 1), and after three weeks of treatment (Week 3, study Day 22 ± 1), to verify compliance with the study drug application and daily diary completion requirements, and to follow up on how the IMP is tolerated.

[q] Skin irritation will be assessed by the Investigator using the Berger/Bowman Scoring Scale; see Appendix 4 of the protocol. To the extent feasible, the same scorer should complete all three skin irritation assessments (Week 2, Week 4/EOT, and Week 6/EOS) for a patient.

[r] Patients will be given a study diary and instructed to enter the following on a daily basis, from the Screening visit through the Week 5/FV2 visit: (1) Each morning between 6 a.m. and 7 a.m. (before the first daily study drug application on treatment days), patients should rate their worst night-time pain in the index knee on the 11-point pain NRS; (2) Between 10 p.m. and 11 p.m. (before the last daily study drug application on each treatment day), patients should rate their worse daytime pain in the index knee on the 11-point pain NRS and enter any adverse event experienced, and breakthrough pain medication use during the day; (3) In addition, on treatment days, patients should enter the time of each study drug application.

[s] Non-compliance with the daily diary requirement during the Screening period (defined as <3 days of daytime or nighttime pain assessments or <3 days of rescue medication entries) will result in exclusion from the study.

1.2.5. Efficacy and Safety Parameters

1.2.5.1. Efficacy Parameters

The primary efficacy endpoint is the change from Baseline in WOMAC (version 3.1; 5-point Likert format) Index Pain Subscale score at the Week 4/EOT visit.

The secondary efficacy endpoints include the following:

- Mean worst *daytime* and mean worst *nighttime* pain severity scores (11-point pain NRS, from patient diary) at Week 4/EOT and Week 5/FV2
- Change from Baseline in mean worst *daytime* and mean worst *nighttime* pain severity scores (11-point pain NRS, from patient diary) at Week 4/EOT and Week 5/FV2 (refer to Section 3.4 for the definition of Baseline)
- Percentages of patients achieving ≥20%, ≥30%, and ≥50% reduction from Baseline in worst *daytime* pain severity scores (11-point pain NRS, from patient diary) at Week 4/EOT
- Percentages of patients achieving ≥20%, ≥30%, and ≥50% reduction from Baseline in worst *nighttime* pain severity scores (11-point pain NRS, from patient diary) at Week 4/EOT
- Change from Baseline to Week 4/EOT in WOMAC Physical Function Subscale score
- Change from Baseline to Week 4/EOT in WOMAC Total score
- Change from Baseline to Week 4/EOT in WOMAC Stiffness Subscale score
- Change from Baseline to Week 2 in WOMAC Total and each of the Subscale scores
- Change from Week 4/EOT to Week 5/FV2 in WOMAC Total and each of the Subscale scores
- Change from Baseline in PGA of disease activity over time
- Change from Baseline in IGA of disease activity over time
- PGA and IGA of overall treatment benefit at Week 2 and Week 4/EOT
- Incidence of rescue medication use, number of doses, and percentage of days with any rescue medication use during the last week of study treatment and overall, during the study treatment period (between first and last IMP dose dates)

1.2.5.2. Safety Parameters

Safety evaluations performed during the study will include physical examinations, measurement of vital signs, 12-lead ECGs, skin irritation, clinical laboratory evaluations including

hematology, serum chemistry, coagulation, and urinalysis. In addition, AEs and concomitant medications will be monitored throughout the study period.

The safety endpoints include the following:

- Frequency and severity of AEs, study drug-related AEs, serious adverse events (SAEs), and AEs leading to study drug discontinuation
- Changes in safety laboratory test results, vital signs measurements, 12-lead ECG, and physical examination findings
- Frequency and severity of application site AEs
- Skin irritation test scores at Week 2, Week 4/EOT, and Week 6/EOS

2. SUBJECT POPULATION

2.1. Population Definitions

The following patient populations will be evaluated and used for presentation and analysis of the data:

- Modified Intent-to-Treat (mITT) Population: All randomized patients who received at least one dose of IMP and provided both BL/FV1 and Week 4/EOT efficacy assessments.

- Per Protocol 1 (PP1) Population: All patients in the mITT population who completed the study without major protocol deviations/factors that could impact the assessment of efficacy.

  Reasons for exclusion from the PP1 population include:

  - Any violation of the Inclusion criteria that may affect the efficacy evaluations
    - Inclusion criteria 2 to 5, 8, and 10
  - Any violation of the Inclusion/Exclusion criteria that may affect the efficacy evaluations
    - Exclusion criteria 1 to 9, 19, 22, 25 to 36
  - Use of rescue medication other than acetaminophen, or rescue medication (acetaminophen) use within 12 hours prior to the Baseline and/or the Week4/EOT primary efficacy endpoint (WOMAC Pain Scale) assessment
  - Use of prohibited medications (as listed in Table 7 of the study protocol) during the treatment period
  - Inadequate (<80%) compliance, defined as <23 days of treatment, or <92 doses applied
  - >12 consecutive IMP doses missed
  - ≥48 hours between last IMP dose and the Week4/EOT primary efficacy endpoint (WOMAC Pain Scale) assessment

- Per Protocol 2 (PP2) Population: All patients in the PP1 population who received at least 65% of the target amount of IMP during the overall 4-week treatment period and during the last 2 weeks of study treatment. Details on the derivation of the target amount of IMP (overall and for the last 2 weeks of study treatment), and compliance based on the amount of IMP used are provided in Section 4.4.1.

- Safety Population: All patients who received any amount of IMP (G001or Vehicle).

The mITT population is the primary population for the analysis of efficacy parameters. All primary and secondary efficacy analyses will also be completed using the PP1 and PP2 populations. The Safety population is the primary population for the analysis of all safety endpoints and demographics/baseline characteristics.

For all efficacy analyses (based on the mITT, PP1 and PP2 populations), patients will be grouped according to the treatment assigned at randomization. For all safety analyses (based on the Safety population), patients will be grouped according to the treatment actually received.

2.2. Protocol Deviations

A major protocol deviation is defined as:

- Any violation of the Inclusion or Exclusion criteria
- Less than 80% treatment compliance by the number of IMP applications (<23 days of treatment, or <92 doses applied)
- >12 consecutive IMP doses missed
- Use of rescue medication (acetaminophen or a non-permitted rescue medication) within 12 hours prior to the Baseline and/or the Week4/EOT primary efficacy endpoint (WOMAC Pain Scale) assessment
- Use of prohibited medications (as listed in Table 7 of the study protocol) during the treatment period
- Any other deviations that might significantly affect the interpretation of the primary endpoint or compromise the scientific value of the trial
- Any other deviation that might significantly affect a patient's safety, well-being, or rights

Major protocol deviations will be summarized and reported. All protocol deviations will be presented in a data listing. A separate listing will be presented for all major protocol deviations.

At the discretion of the Sponsor, a blinded review of key study data will be undertaken prior to clinical database lock and unblinding at the study level, to determine the exclusions from the Per Protocol (PP) populations. The Sponsor or designee will be responsible for producing the final protocol deviation file (formatted as a Microsoft Excel). This file will include a description of each protocol deviation that warrants exclusion from each PP population (PP1 and PP2), as defined in Section 2.1. This file will be finalized prior to hard database lock.

Relevant Output

Table 14.1.5 Major Protocol Deviations Summary (Safety Population)

Listing 16.2.2.1 Inclusion/Exclusion Criteria (All Screened Patients)
Listing 16.2.2.2 Major Protocol Deviations (All Screened Patients)
Listing 16.2.2.3 Minor Protocol Deviations (All Randomized Patients)

3. GENERAL STATISTICAL METHODS

3.1. Sample Size Justification

A total sample size of 174 evaluable patients (87 patients in G001 treatment group and 87 patients in Vehicle control group) achieves 80% power to detect a statistically significant difference of 1.5 points in the mean WOMAC Likert Pain Subscale scores (range from 0 – 20) between the two treatment groups, assuming a common standard deviation (SD) of 3.5 in a two-sided t-test with Type I error rate (alpha) of 0.05.

The assumptions on change in WOMAC Pain Subscale score between the two treatment groups were based on available literature on clinically relevant differences and efficacy of celecoxib oral formulation (Williams et al. 2000; McKenna et al. 2001; Rother et al. 2007).

Based on available literature, it is estimated that up to 20% of enrolled patients may withdraw from the study prior to the Week 4/EOT efficacy assessment. Therefore, to control for possible dropouts, the sample size is increased to 210 patients, or 105 patients per treatment arm.

3.2. General Methods

All data listings that contain an evaluation date will contain a relative study day (Rel Day). Pre-treatment and on-treatment study days are numbered relative to the day of the first dose of study medication which is designated as Day 1. The preceding day is Day -1, the day before that is Day -2, etc.

All output will be incorporated into RTF and Adobe Acrobat PDF files, sorted and labeled according to the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) recommendations, and formatted to the appropriate page size(s).

Tabulations will be produced for appropriate demographic, baseline, compliance, exposure, efficacy, and safety parameters. For categorical variables, summary tabulations of the number and percentage of subjects within each category (with a category for missing data) of the parameter will be presented. For continuous variables, the number of subjects, mean, median, SD, minimum, and maximum values will be presented.

Formal statistical hypothesis testing will be performed on the primary efficacy endpoint with all tests conducted at the 2-sided, 0.05 level of significance. Summary statistics will be presented, as well as confidence intervals (CIs) on selected parameters, as described in the sections below.

Tabulations will be presented by treatment group (G001 or Vehicle) and overall.

3.3. Computing Environment

All descriptive statistical analyses will be performed using SAS statistical software Version 9.4, unless otherwise noted. Medical history and AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 23.1. Concomitant medications will be coded using the World Health Organization (WHO) Drug Dictionary Global B3 September 2020.

3.4. Baseline Definitions

For all analyses except change from baseline in worst daytime and nighttime pain severity NRS scores, baseline will be defined as the most recent measurement prior to the first application of study drug (IMP).

The baseline for worst daytime and nighttime pain severity NRS scores is defined as the mean of the 3 – 7 day pain scores prior to the Baseline Visit, as collected in the Screening diary.

3.5. Methods of Pooling Data

For purposes of summary tabulations, data will be pooled across study sites and by treatment group for all analyses. No additional pooling of data is planned.

3.6. Adjustments for Covariates

The analysis for primary and secondary efficacy endpoints includes baseline value of these endpoints as a covariate in the model. A sensitivity analysis for the primary efficacy endpoint will include percent treatment compliance based on IMP amount used (computed based on the leftover IMP returned to clinic; see Section 4.4.1), as a covariate.

3.7. Multiple Comparisons/Multiplicity

Multiplicity is not of concern for this study with a single primary efficacy endpoint.

There are multiple secondary efficacy endpoints as specified in Section 1.2.5.1. Analysis of these secondary endpoints will be considered supportive. Hence, no adjustment for multiple comparisons will be made among the secondary efficacy endpoints.

3.8. Subpopulations

Subgroup analyses of primary and select secondary efficacy endpoints may be performed based on prior OA treatment and depending on the number of patients in each category (Section 4.3.1.3).

3.9. Withdrawals, Dropouts, Loss to Follow-up

If a patient withdraws from the study prematurely for any reason or repeatedly fails to return for visits (i.e., is lost to follow-up), the Investigator should determine the primary reason for a patient's premature withdrawal from the study and record the reason in the respective patient's study documents. At the time of premature discontinuation, if feasible, an EOT visit should be conducted (within no later than 2 days after the last dose of IMP), followed by an EOS visit (14 ± 2 days after last study drug application) as shown in the Schedule of Assessment (Table 1).

If the patient withdraws consent for disclosure of future information, the Sponsor may retain and continue to use any data collected before such a withdrawal of consent.

Due to planned enrollment overage to account for potential dropouts, subjects who are withdrawn or discontinue from the study will not be replaced.

3.10. Missing, Unused, and Spurious Data

All data recorded on the case report form (CRF) will be included in data listings that will accompany the CSR.

3.10.1. Efficacy Data

Missing WOMAC assessments will be handled as instructed in the WOMAC 3.1 User Guide XI and is summarized as follows:

WOMAC Pain Subscale:

- If exactly 1 pain response (out of 5) is missing, the missing response will be imputed with the mean of the 4 non-missing pain responses.
- If more than 1 pain response (out of 5) is missing, the Pain Subscale score for that time point will be considered missing.

WOMAC Stiffness Subscale:

- If exactly 1 stiffness response (out of 2) is missing, the missing response will be imputed with the other stiffness response.
- If both stiffness responses are missing, the Stiffness Subscale score will be considered missing.

WOMAC Physical Function Subscale:

- If no more than 3 physical function responses (out of 17) are missing, the missing responses will be imputed as the mean of the non-missing physical function responses.
- If more than 3 physical function responses (out of 17) are missing, the Physical Function Subscale score will be considered missing.

WOMAC assessments that are missing completely at a visit will not be imputed. If at least one of the WOMAC subscale score is missing completely, the WOMAC Total Score will be considered missing and will not be imputed.

Furthermore, missing PGA/IGA of disease activity, PGA/IGA of overall treatment benefit, and worst daytime and nighttime pain severity scores (11-point pain NRS) will not be imputed.

3.10.2. Adverse Event Dates

When tabulating AE data, partial dates will be handled as follows to determine treatment-emergence.

- If the start date of an AE is partially or completely missing, the date will be compared as far as possible with the date of the first study drug application. The AE will be assumed

to be treatment-emergent if it cannot be definitively shown that the AE did not occur or worsen during the treatment period (worst case approach).

The following general rule will be used:

- If the start date is complete, an AE will only be excluded from being treatment-emergent if the start date and start time is before the date and time of the first study drug application.
- If the start day is missing but the start month and year are complete, an AE will only be excluded from being treatment-emergent if the start month/year is before the month/year of the first study drug application or if the stop date is before the first study drug application.
- If the start day and month are missing but the start year is complete, an AE will only be excluded from being treatment-emergent if the start year is before the year of first study drug application or if the stop date is before the first study drug application.
- If the start date is completely missing, an AE will be considered treatment-emergent unless the stop date is before the first study drug application.

3.10.3. Skin Irritation Scores

Skin irritation will be assessed by the Investigator using the Berger/Bowman Scoring Scale (Berger and Bowman, 1982; US FDA Guidance 2018b), comprising the "Dermal Response" and "Other Effects" scoring system, as outlined in Protocol Appendix 4.

If a patient was removed from treatment due to skin irritation, the patient's skin irritation score at the time of treatment discontinuation will be used as the imputed value at subsequent assessment timepoint(s) (i.e., last observation carried forward [LOCF]). If a patient was removed from treatment due to reasons other than skin irritation, missing data will not be imputed.

3.10.4. Prior and Concomitant Medication

Medications with partial or missing dates will be assumed to be concomitant unless there is clear evidence (i.e., through comparisons of partial dates as noted below) to suggest that the medication ended prior to study drug application.

Partially missing dates for concomitant medications will be imputed for determination of designation of prior vs. concomitant medication as follows:

- If the start date is unknown (i.e., complete missing start date):
    - If there is a record indicating that medication is not ongoing at baseline (although end date is unknown), the medication will be considered prior.
    - If there is no end date or any evidence that medication is not ongoing, the medication will be considered both prior and concomitant.
- If the month and the day of the start date are missing, the month and the day will be imputed to January 1 of the year specified.

- If the day of the start date is missing and there is no end date, the day will be imputed to the first day of the month specified.
- If the month and the day of the end date are missing, the month and the day will be imputed to December 31 of the year specified, unless there is evidence that the medication is not ongoing in the study (although complete end date is unknown). In such circumstances the medication will be considered as prior.
- If the day of the end date is missing, the day will be imputed to the last day of the month specified, unless there is evidence that the medication is not ongoing in the study (although complete end date is unknown). In such circumstances the medication will be considered as prior.

3.11. Visit Windows

It is expected that all visits should occur according to the protocol schedule. All data will be tabulated per the evaluation visit as recorded on the CRF even if the assessment is outside of the visit window (refer to Table 1 for visit window). If the evaluation visit is missing in the database but there are data from an unscheduled or additional visit that is inside the visit window, the data from the unscheduled or additional visit will be used in data summaries. In data listings, the relative day of all dates will be presented.

3.12. Interim Analyses

No interim analyses are planned for this study.

4. STUDY ANALYSES

4.1. Subject Disposition

Patient disposition will be tabulated and include the number screened, the number randomized, the number treated with G001/Vehicle control, and overall. The number in each patient population for analysis, the number who completed the treatment, the number who withdrew prior to completing treatment, primary reason for withdrawal from the treatment, the number who completed the study, the number who withdrew prior to completing the study and the primary reason for withdrawal from the study will also be summarized.

A by-patient data listing of study completion information including the reason for premature withdrawal from the study, if applicable, will be presented.

Relevant Output

Table 14.1.1 Subject Disposition

Listing 16.2.1.1 Subject Disposition (All Randomized Patients)
Listing 16.2.1.2 Randomization Details (All Screened Patients)
Listing 16.2.3.1 Study Populations (All Randomized Patients)

4.2. Demographics and Baseline Characteristics

Demographics and baseline characteristics will be summarized and presented by treatment group and overall for the Safety population. Age, height, weight, and body mass index (BMI) will be summarized using descriptive statistics (number of patients, mean, SD, median, minimum, and maximum). The number and percentage of patients in each sex, ethnicity and race category will be presented. The number and percentage of patients falling into each of the following age groups will also be presented: <40 years, 40 – 64 years, 65 – 74 years, ≥ 75 years.

Worst daily pain in the past 24 hours (11-point NRS) for the Screening Visit and BL/FV1 will be tabulated and summarized by the following 7 categories: 0–3, 4, 5, 6, 7, 8, 9–10, and presented separately for the index knee and contralateral knee.

General medical history and OA history of the knee will be tabulated by treatment group for the Safety population.

Demographic, baseline characteristics, OA history (date of diagnosis, American College of Rheumatology [ACR] clinical criteria, radiological examination), OA treatment history, medical history, and worst daily pain assessment results (11-point pain NRS) at Screening Visit and BL/FV1 for each patient will be provided in data listings.

Relevant Output

Table 14.1.2 Demographic and Baseline Characteristics (Safety Population)
Table 14.1.3 Medical History (Safety Population)
Table 14.1.4.1 History of Osteoarthritis of the Knee (Safety Population)

Table 14.1.4.2 Osteoarthritis Treatment History (Safety Population)
Listing 16.2.4.1 Demographics and Baseline Characteristics (All Randomized Patients)
Listing 16.2.4.2 Medical History (All Randomized Patients)
Listing 16.2.4.3 History of Osteoarthritis of the Knee (All Randomized Patients)
Listing 16.2.4.4 Osteoarthritis Treatment History (All Randomized Patients)
Listing 16.2.4.5 Worst Daily Pain 11-Point NRS (All Randomized Patients)

4.3. Efficacy Evaluation

Efficacy analyses will be conducted using the mITT population as primary analysis population, and repeated using the PP1 and PP2 populations (repeat analyses using the PP1 and PP2 populations may be limited to the primary efficacy endpoint, and select secondary efficacy endpoints). The primary and secondary efficacy endpoints are described in Section 1.2.5.1.

For all efficacy evaluations involving WOMAC ratings, the WOMAC 3.1 Index (5-point Likert format) will be used. The WOMAC OA index is a self-administered questionnaire consisting of 24 questions divided into three subscales (Bellamy et al. 1988):

- Pain Subscale: 5 questions
- Stiffness Subscale: 2 questions
- Physical Function Subscale: 17 questions

The Likert version of the WOMAC is rated on an ordinal scale of 0 to 4, where 0 represents the lowest level of symptoms or physical disability: 0 = None; 1 = Mild; 2 = Moderate; 3 = Severe; 4 = Extreme. For each subscale, the scores for individual items are summed, providing the following possible ranges: 0 to 20 for Pain Subscale, 0 to 8 for Stiffness Subscale, and 0 to 68 for Physical Function Subscale. The sum of the three subscales is the WOMAC Total score and ranges from 0 to 96. Higher scores on the WOMAC indicates worse pain, stiffness, and physical function. Imputation rules for missing WOMAC data are specified in Section 3.10.1.

Sensitivity analyses of the primary and select secondary efficacy endpoints are detailed in Section 4.3.1.1.

4.3.1. Primary Efficacy Analysis

The primary efficacy endpoint is the change from Baseline in the WOMAC Pain Subscale score in 5-point Likert format at the Week 4/EOT visit. The primary efficacy analysis will be performed using an analysis of covariance (ANCOVA) model. The model will include treatment (G001 or Vehicle) as the independent variable, change from Baseline in the WOMAC Pain Subscale score at the Week 4/EOT visit as the outcome (dependent) variable, and Baseline WOMAC Pain Subscale score as a covariate. The overall difference of adjusted means between the treatment groups will be analyzed by a partial F test at a significance level of 0.05.

The primary efficacy outcome comparison between the two treatment groups (G001 group and Vehicle control) will also be provided by comparing their Least Square (LS) means of change from Baseline WOMAC Pain Subscale scores. The LS means are interpreted as the mean change from Baseline WOMAC Pain Subscale score of the two treatment groups after adjusting for the covariate (Baseline WOMAC score) in the model. Summary tables will include LS

means, standard errors, and 95% CIs for each treatment group. The difference in LS means, as well as the associated 95% CI and the p-value from the t-test for the statistical test for the LS means difference, will be presented as well.

Refer to Section 4.3.2.1 for additional analysis of the WOMAC Pain Subscale score actual values by visit, graphical representation and by patient listing.

Relevant Output

Table 14.2.1.1.1A	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Baseline to Week 4/EOT: Pain Subscale (mITT Population: Primary Efficacy Analysis)
Table 14.2.1.1.1B	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Baseline to Week 4/EOT: Pain Subscale (Per Protocol Population 1: Supportive Primary Efficacy Analysis)
Table 14.2.1.1.1C	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Baseline to Week 4/EOT: Pain Subscale (Per Protocol Population 2: Supportive Primary Efficacy Analysis)

4.3.1.1. Sensitivity Analyses

A sensitivity analysis will be performed to evaluate the effect of IMP compliance on the primary efficacy endpoint. This sensitivity analysis will use the same ANCOVA model as described in Section 4.3.1 but with one additional covariate: percent IMP amount-based compliance. Percent IMP compliance will be computed according to leftover IMP returned to clinic by randomized patients in the mITT population, as detailed in Section 4.4.1. The overall difference of adjusted means between the treatment groups will be analyzed by a partial F test at a significance level of 0.05.

In addition, sensitivity ‘completers’ analysis on the primary (and potentially key secondary) efficacy endpoint(s) may be completed on a subset of patients who completed the Week 4/EOT efficacy assessments after a minimum 25-day overall treatment period.

Relevant Output

Table 14.2.1.1.1A1	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Baseline to Week 4/EOT: Pain Subscale (mITT Population: IMP Compliance Sensitivity Analyses for Primary Efficacy Analysis)
Table 14.2.1.1.1A2	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Baseline to Week 4/EOT: Pain Subscale (mITT Population: Completers Sensitivity Analyses for Primary Efficacy Analysis)

4.3.1.2. Dose Response Plot

To support primary efficacy analysis and the assessment of the effect of percent IMP amount-based compliance on efficacy, a dose response plot will be provided to explore the variation in

the primary efficacy endpoint as a function of percent IMP compliance (based on amount of IMP used, and number of IMP applications) in the two treatment groups. This plot will be considered supportive and no formal statistical inferences or quantitative conclusion will be drawn.

Relevant Output

Figure 14.2.1.1.1 Dose Response Scatter Plot of Change from Baseline to Week 4/EOT in Western Ontario and McMaster Universities (WOMAC) Pain Subscale Score (mITT Population: Supportive Plot for IMP Amount Base)

Figure 14.2.1.1.2 Dose Response Scatter Plot of Change from Baseline to Week 4/EOT in Western Ontario and McMaster Universities Pain Subscale Score (mITT Population: Supportive Plot for Number of IMP Applications)

4.3.1.3. Subgroup Analyses of the Primary Endpoint

Subgroup analyses of primary and select secondary efficacy endpoints may be performed based on prior OA treatment categories, including the following:

- Patients with Pre-study NSAID use for OA (with or without acetaminophen)
- Patients with Pre-study Acetaminophen use only (without an NSAID) for OA

Subgroup analysis for each category will be invoked if the minimum number of patients comprise of at least 30% of the total number of patients in mITT population and will be performed as detailed in Section 4.3.1.

Relevant Output

Table 14.2.1.1.1A3 Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Baseline to Week 4/EOT, Pain Subscale (mITT Population, Subgroup Analysis, by prior OA Medication Use)

4.3.2. Secondary Efficacy Analyses

Analyses of secondary efficacy endpoints will be considered supportive. Hence, no p-value adjustment will be made for multiple endpoints or multiple comparisons in the secondary efficacy analysis.

4.3.2.1. WOMAC Stiffness Subscale, Physical Function Subscale, and Total Scores

WOMAC Stiffness Subscale score, Physical Function Subscale score, and Total score will be analyzed via ANCOVA model similar to the primary efficacy endpoint as described in Section 4.3.1.

The following secondary efficacy endpoints will be evaluated:

- Change from Baseline to Week 4/EOT in WOMAC Stiffness Subscale score

- Change from Baseline to Week 4/EOT in WOMAC Physical Function Subscale score
- Change from Baseline to Week 4/EOT in WOMAC Total score, defined as the sum of the 3 WOMAC subscale scores (Pain, Stiffness, and Physical Function)
- Change from Baseline to Week 2 in WOMAC Pain Subscale score
- Change from Baseline to Week 2 in WOMAC Stiffness Subscale score
- Change from Baseline to Week 2 in WOMAC Physical Function Subscale score
- Change from Baseline to Week 2 in WOMAC Total score, defined as the sum of the 3 WOMAC subscale scores (Pain, Stiffness, and Physical Function)
- Change from Week 4/EOT to Week 5/FV2 in WOMAC Pain Subscale score
- Change from Week 4/EOT to Week 5/FV2 in WOMAC Stiffness Subscale score
- Change from Week 4/EOT to Week 5/FV2 in WOMAC Physical Function Subscale score
- Change from Week 4/EOT to Week 5/FV2 in WOMAC Total score, defined as the sum of the 3 WOMAC subscale scores (Pain, Stiffness, and Physical Function)

Descriptive statistics of the actual value and change from Baseline at each timepoint will be presented by treatment group.

Actual mean scores and SDs will also be displayed graphically by visit for each subscale and total score.

A by-patient data listing will be provided for each WOMAC Subscale scores (Pain, Stiffness, and Physical Function).

Relevant Output

Table 14.2.1.1.2A	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Baseline to Week 2: Pain Subscale (mITT Population: Secondary Efficacy Analysis)
Table 14.2.1.1.2B	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Baseline to Week 2: Pain Subscale (Per Protocol Population 1: Secondary Efficacy Analysis)
Table 14.2.1.1.3A	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Week 4/EOT to Week 5/FV2: Pain Subscale (mITT Population: Secondary Efficacy Analysis)
Table 14.2.1.1.3B	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Week 4/EOT to Week 5/FV2: Pain Subscale (Per Protocol Population 1: Secondary Efficacy Analysis)
Table 14.2.1.2.1A	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Baseline to Week 4/EOT: Stiffness Subscale (mITT Population)

Listing 16.2.6.4	Western Ontario and McMaster Universities (WOMAC) 3.1 Osteoarthritis Index Likert Scores: Total Scores (All Randomized Patients)
Figure 14.2.2.1A	Western Ontario and McMaster Universities (WOMAC) Mean Scores by Visit and Treatment: Pain Subscale (mITT Population: Primary Efficacy Analysis)
Figure 14.2.2.1B	Western Ontario and McMaster Universities (WOMAC) Mean Scores by Visit and Treatment: Pain Subscale (Per Protocol Population 1)
Figure 14.2.2.1C	Western Ontario and McMaster Universities (WOMAC) Mean Scores by Visit and Treatment: Pain Subscale (Per Protocol Population 2)
Figure 14.2.2.2A	Western Ontario and McMaster Universities (WOMAC) Mean Scores by Visit and Treatment: Stiffness Subscale (mITT Population)
Figure 14.2.2.2B	Western Ontario and McMaster Universities (WOMAC) Mean Scores by Visit and Treatment: Stiffness Subscale (Per Protocol Population 1)
Figure 14.2.2.3A	Western Ontario and McMaster Universities (WOMAC) Mean Scores by Visit and Treatment: Physical Function Subscale (mITT Population)
Figure 14.2.2.3B	Western Ontario and McMaster Universities (WOMAC) Mean Scores by Visit and Treatment: Physical Function Subscale (Per Protocol Population 1)
Figure 14.2.2.4A	Western Ontario and McMaster Universities (WOMAC) Mean Scores by Visit and Treatment: Total Score (mITT Population)
Figure 14.2.2.4B	Western Ontario and McMaster Universities (WOMAC) Mean Scores by Visit and Treatment: Total Score (Per Protocol Population 1)
Figure 14.2.2.4C	Western Ontario and McMaster Universities (WOMAC) Mean Scores by Visit and Treatment: Total Score (Per Protocol Population 2)

In addition, a sensitivity analysis of change from Baseline to Week 4/EOT in WOMAC Total score may also be completed using the 'completers' analysis population (see Section 4.3.1.1).

Relevant Output

Table 14.2.1.4.1A1	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Baseline to Week 4/EOT: Total Score (mITT Population: Completers Sensitivity Analyses)

Subgroup analyses of change from Baseline to Week 4/EOT in WOMAC Total score based on prior OA treatment may be performed as detailed in Section 4.3.1.3.

Relevant Output

Table 14.2.1.4.1A2	Western Ontario and McMaster Universities (WOMAC) Analysis of Covariance from Baseline to Week 4/EOT: Total Score (mITT Population: Subgroup Analysis, by prior OA Medication Use)

4.3.2.2. Global Assessment of Disease Activity

Two global assessment of disease activity scores will be recorded in this study: PGA and IGA. Both PGA and IGA of disease activity measure OA in the index knee over the last 48 hours on a 5-point Likert scale: 0 = Very Good; 1 = Good; 2 = Fair; 3 = Poor; 4 = Very Poor.

PGA and IGA of disease activity will be recorded at Screening visit, BL/FV1, Week 2, Week 4/EOT, and Week 5/FV2. At each visit post-BL/FV1, the change from Baseline PGA/IGA of disease activity scores will be calculated and the response will be classified into 3 categories according to the following rules:

- Improved – defined as either of the following:
  - For Baseline Grade ≥ 2: a reduction of ≥ 2 grades from Baseline
  - For Baseline Grade = 1: a change in grade from 1 to 0
- Worsened – defined as either of the following:
  - For Baseline Grade ≤ 2: an increase of ≥ 2 grades from Baseline
  - For Baseline Grade = 3: a change in grade from 3 to 4
- No Change – defined as a change from Baseline PGA/IGA of disease activity that is neither Improved nor Worsened

An ordinal logistic regression model will be used to analyze PGA of disease activity at the Week 2 timepoint. The model will include treatment (G001 or Vehicle) as the independent variable, change from Baseline of PGA of disease activity category (Improved, No Change, or Worsened) at the Week 2 Visit as the outcome (dependent) variable, and Baseline PGA of disease activity score as a covariate. The ordinal logistic regression model will assume a proportional odds structure and be implemented by PROC LOGISTIC in SAS. The ordering of the 3 categories will be defined as follows:

Worsened < No Change < Improved

If the final data does not support proportional odds assumption, as judged by the p-value of the Score Test being significant at a significance level of 0.05, a non-proportional odds model will be used by inserting UNEQUALSLOPES option in the PROC LOGISTIC model statement.

A summary table of the point estimates, standard errors, and p-values for the slopes and intercepts of the model will be presented. Point estimates and Wald 95% CIs for the odds ratio will also be summarized. Furthermore, a summary table for the number and percentage of patients in each of the 3 PGA disease activity categories will be presented by treatment group and overall.

The same analysis will be repeated for the change from Baseline of PGA of disease activity at Week 4/EOT and Week 5/FV2.

The change from Baseline of IGA of disease activity at the Week 2, Week 4/EOT, and Week 5/FV2 timepoint will be analyzed in similar fashion as the change from Baseline of PGA of disease activity as described above.

Descriptive statistics of the actual value and change from Baseline of PGA and IGA of disease activity at each timepoint will be presented by treatment group.

A by-patient data listing will be provided for PGA and IGA of disease activity scores.

Relevant Output

Table 14.2.3.1A	Patient Global Assessment of Disease Activity Ordinal Logistic Regression at Week 2 (mITT Population)
Table 14.2.3.1B	Patient Global Assessment of Disease Activity Ordinal Logistic Regression at Week 2 (Per Protocol Population 1)
Table 14.2.3.2A	Patient Global Assessment of Disease Activity Ordinal Logistic Regression at Week 4/EOT (mITT Population)
Table 14.2.3.2B	Patient Global Assessment of Disease Activity Ordinal Logistic Regression at Week 4/EOT (Per Protocol Population 1)
Table 14.2.3.3A	Patient Global Assessment of Disease Activity Ordinal Logistic Regression at Week 5/FV2 (mITT Population)
Table 14.2.3.3B	Patient Global Assessment of Disease Activity Ordinal Logistic Regression at Week 5/FV2 (Per Protocol Population 1)
Table 14.2.4.1A	Patient Global Assessment of Disease Activity Ordinal Scores by Visit and Treatment (mITT Population)
Table 14.2.4.1B	Patient Global Assessment of Disease Activity Ordinal Scores by Visit and Treatment (Per Protocol Population 1)
Table 14.2.5.1A	Investigator Global Assessment of Disease Activity Ordinal Logistic Regression at Week 2 (mITT Population)
Table 14.2.5.1B	Investigator Global Assessment of Disease Activity Ordinal Logistic Regression at Week 2 (Per Protocol Population 1)
Table 14.2.5.2A	Investigator Global Assessment of Disease Activity Ordinal Logistic Regression at Week 4/EOT (mITT Population)
Table 14.2.5.2B	Investigator Global Assessment of Disease Activity Ordinal Logistic Regression at Week 4/EOT (Per Protocol Population 1)
Table 14.2.5.3A	Investigator Global Assessment of Disease Activity Ordinal Logistic Regression at Week 5/FV2 (mITT Population)
Table 14.2.5.3B	Investigator Global Assessment of Disease Activity Ordinal Logistic Regression at Week 5/FV2 (Per Protocol Population 1)
Table 14.2.6.1A	Investigator Global Assessment of Disease Activity Ordinal Scores by Visit and Treatment (mITT Population)
Table 14.2.6.1B	Investigator Global Assessment of Disease Activity Ordinal Scores by Visit and Treatment (Per Protocol Population 1)
Table 14.2.7.1A	Investigator and Patient Global Assessments of Disease Activity Actual Scores and Change from Baseline by Visit and Treatment (mITT Population)
Table 14.2.7.1B	Investigator and Patient Global Assessments of Disease Activity Actual Scores and Change from Baseline by Visit and Treatment (Per Protocol Population 1)

Listing 16.2.6.5 Patient Global Assessment (All Randomized Patients)
Listing 16.2.6.6 Investigator Global Assessment (All Randomized Patients)

4.3.2.3. Global Assessment of Overall Treatment Benefit

The PGA and IGA of overall treatment benefit are secondary efficacy endpoints and similar to the PGA/IGA of disease activity, are each evaluated on a 5-point Likert scale: 0 = Very Good; 1 = Good; 2 = Fair; 3 = Poor; 4 = Very Poor.

The mean of the PGA or IGA of overall treatment benefit at Week 2 from the two treatment groups (G001 and Vehicle) will be compared using a t-test and a 2-sided p-value will be presented. A similar analysis will be performed for the mean of the PGA or IGA of overall treatment benefit at Week 4/EOT. No formal statistical conclusion will be drawn based on these tests.

Descriptive statistics of the actual value of PGA and IGA of overall treatment benefit scores at Week 2 and at Week 4/EOT will be presented by treatment group.

A by-patient data listing will be provided for PGA and IGA of overall treatment benefit scores.

Relevant Output

Table 14.2.8.1A Investigator and Patient Global Assessments of Overall Treatment Benefit Actual Scores by Visit and Treatment (mITT Population)
Table 14.2.8.1B Investigator and Patient Global Assessments of Overall Treatment Benefit Actual Scores by Visit and Treatment (Per Protocol Population 1)

4.3.2.4. Weekly Mean Worst Daytime and Nighttime Pain NRS Scores

The mean worst **daytime** pain NRS score at the specified visit is computed based on patient diary entries as follows:

- Baseline mean worst daytime pain NRS scores = sum of worst daytime pain NRS scores reported divided by the number of worst daytime pain assessments reported from 1 day before the date of Baseline/FV1 Visit up to 7 days before the date of Baseline/FV1 Visit
- Week 2 Visit mean worst daytime pain NRS scores = sum of worst daytime pain NRS scores reported divided by the number of worst daytime pain assessments reported during the 7-day period from 1 day before the date of Week 2 Visit to 7 days before the date of Week 2 Visit
- Week 4 Visit mean worst daytime pain NRS scores = sum of worst daytime pain NRS scores reported divided by the number of worst daytime pain assessments reported during the 7-day period from 1 day before the date of Week 4 Visit to 7 days before the date of Week 4 Visit
- Week 5 Visit mean worst daytime pain NRS scores = sum of worst daytime pain NRS scores reported divided by the number of worst daytime pain assessments reported from 1 day before the date of Week 5 Visit up to the later of 7 days prior to the date of Week 5 Visit or the date of Week 4

The mean worst **nighttime** pain NRS score at the specified visit is computed based on patient diary entries as follows:

- Baseline mean worst nighttime pain NRS scores = sum of worst nighttime pain NRS scores reported divided by the number of worst nighttime pain assessments reported from the date of Baseline/FV1 Visit up to 6 days before the date of Baseline/FV1 Visit
- Week 2 Visit mean worst nighttime pain NRS scores = sum of worst nighttime pain NRS scores reported divided by the number of worst nighttime pain assessments reported during the 7-day period from the date of Week 2 Visit to 6 days before the date of Week 2 Visit
- Week 4 Visit mean worst nighttime pain NRS scores = sum of worst nighttime pain NRS scores reported divided by the number of worst nighttime pain assessments reported during the 7-day period from the date of Week 4 Visit to 6 days before the date of Week 4 Visit
- Week 5 Visit mean worst nighttime pain NRS scores = sum of worst nighttime pain NRS scores reported divided by the number of worst nighttime pain assessments reported from the date of Week 5 Visit up to the later of 6 days prior to the date of Week 5 Visit or the date of Week 4 +1

Weekly mean worst *daytime* and *nighttime* pain NRS scores will be presented by treatment group. Changes from baseline in mean worst *daytime* and in mean worst *nighttime* pain severity scores at Week 4/EOT in the two treatment groups will be compared using a t-test. Percentages of patients achieving ≥20%, ≥30% and ≥50% reduction from Baseline in worst *daytime* pain and in worst *nighttime* pain severity scores (11-point pain NRS) at Week 4/EOT will be descriptively summarized, including the 95% exact CI. Cochran-Mantel-Haenszel (CMH) test will be used to determine the difference between treatment groups. The p-value from the general association test of CMH will be presented.

A by-patient data listing will be provided for worst daytime and nighttime pain NRS scores.

NRS scores will also be displayed graphically.

Relevant Output

Listing 16.2.6.7	Worst Daytime/Nighttime Pain 11-Point NRS Diary (All Randomized Patients)
Table 14.2.9.1A	Weekly Mean Worst Daytime Pain Severity Numerical Rating Scale (NRS) by Visit (mITT Population)
Table 14.2.9.1B	Weekly Mean Worst Daytime Pain Severity Numerical Rating Scale (NRS) by Visit (Per Protocol Population 1)
Table 14.2.9.1C	Weekly Mean Worst Daytime Pain Severity Numerical Rating Scale (NRS) by Visit (Per Protocol Population 2)
Table 14.2.9.2A	Weekly Mean Worst Nighttime Pain Severity Numerical Rating Scale (NRS) by Visit (mITT Population)
Table 14.2.9.2B	Weekly Mean Worst Nighttime Pain Severity Numerical Rating Scale (NRS) by Visit (Per Protocol Population 1)

Table 14.2.9.2C	Weekly Mean Worst Nighttime Pain Severity Numerical Rating Scale (NRS) by Visit (Per Protocol Population 2)
Table 14.2.9.3A	Percent Reduction from Baseline to Week 4/EOT in Worst Daytime and Worst Nighttime Pain (mITT Population)
Table 14.2.9.3B	Percent Reduction from Baseline to Week 4/EOT in Worst Daytime and Worst Nighttime Pain (Per Protocol Population 1)
Table 14.2.9.3C	Percent Reduction from Baseline to Week 4/EOT in Worst Daytime and Worst Nighttime Pain (Per Protocol Population 2)
Figure 14.2.9.1A	Weekly Mean Worst Daytime Pain Severity Numerical Rating Scale (NRS) Actual Scores by Visit (mITT Population)
Figure 14.2.9.1B	Weekly Mean Worst Daytime Pain Severity Numerical Rating Scale (NRS) Actual Scores by Visit (Per Protocol Population 1)
Figure 14.2.9.1C	Weekly Mean Worst Daytime Pain Severity Numerical Rating Scale (NRS) Actual Scores by Visit (Per Protocol Population 2)
Figure 14.2.9.2A	Weekly Mean Worst Nighttime Pain Severity Numerical Rating Scale (NRS) Actual Scores by Visit (mITT Population)
Figure 14.2.9.2B	Weekly Mean Worst Nighttime Pain Severity Numerical Rating Scale (NRS) Actual Scores by Visit (Per Protocol Population 1)
Figure 14.2.9.2C	Weekly Mean Worst Nighttime Pain Severity Numerical Rating Scale (NRS) Actual Scores by Visit (Per Protocol Population 2)

In addition, a sensitivity analysis may also be completed using the 'completers' analysis population as detailed in Section 4.3.1.1.

Sensitivity Analysis outputs:

Table 14.2.9.1A1	Weekly Mean Worst Daytime Pain Severity Numerical Rating Scale (NRS) by Visit (mITT Population: Completers Sensitivity Analyses)
Table 14.2.9.2A1	Weekly Mean Worst Nighttime Pain Severity Numerical Rating Scale (NRS) by Visit (mITT Population: Completers Sensitivity Analyses)
Table 14.2.9.3A1	Percent Reduction from Baseline to Week 4/EOT in Worst Daytime Pain (mITT Population: Completers Sensitivity Analyses)

Subgroup analyses based on prior OA treatment may be performed for the following secondary endpoints as detailed above.

- Mean worst daytime and mean worst nighttime pain severity scores (11-point pain NRS, from patient diary) at Week 4/EOT, including mean changes from Baseline
- Percentages of patients achieving ≥20%, ≥30%, and ≥50% reduction from Baseline in worst daytime pain severity scores (11-point pain NRS, from patient diary) at Week 4/EOT
- Percentages of patients achieving ≥20%, ≥30%, and ≥50% reduction from Baseline in worst nighttime pain severity scores (11-point pain NRS, from patient diary) at Week 4/EOT

Relevant Outputs:

Table 14.2.9.1A2 Weekly Mean Worst Daytime Pain Severity Numerical Rating Scale (NRS) by Visit (mITT Population: Subgroup Analysis, by prior OA Medication Use)

Table 14.2.9.2A2 Weekly Mean Worst Nighttime Pain Severity Numerical Rating Scale (NRS) by Visit (mITT Population: Subgroup Analysis, by prior OA Medication Use)

Table 14.2.9.3A2 Percent Reduction from Baseline to Week 4/EOT in Worst Daytime and Worst Nighttime Pain (mITT Population: Subgroup Analysis, by prior OA Medication Use)

4.3.2.5. Incidence of Rescue Medication Use

The number and percentage of patients with at least 1 rescue medication (acetaminophen) use will be presented by treatment group for:

- Patients with at least one rescue medication use during the study
- Patients with at least one rescue medication use during treatment period (day 1 through week 4/EOT)
- Patients with at least one rescue medication use within 12 hours prior to primary efficacy endpoint assessment (WOMAC Pain Scale) at week 4/EOT visit
- Rescue medication doses (mg) within 12 hours prior to primary efficacy endpoint assessment (WOMAC Pain Scale) at week 4/EOT visit
- Number of rescue medication doses during the study
- Number of rescue medication doses during treatment period (day 1 through week 4/EOT)
- Average daily rescue medication dose (mg/day) during the study
- Average daily rescue medication dose (mg/day) during treatment period (day 1 through week 4/EOT)
- Percentage of days with any rescue medication use during the study
- Percentage of days with any rescue medication use during treatment period (day 1 through week 4/EOT)
- The percentage of days with any rescue medication use during the last week (from 1 day before the date of Week 4/EOT Visit to 7 days before the date of Week 4/EOT Visit) of study treatment

Relevant Output

Table 14.2.10.1A Rescue Medication Use by Treatment (mITT Population)
Table 14.2.10.1B Rescue Medication Use by Treatment (Per Protocol Population 1)

Listing 16.2.6.8 Rescue Medication Diary (All Randomized Patients)

4.4. Safety Analyses

Safety analyses will be conducted using the Safety population.

4.4.1. Study Drug Exposure and Compliance

The first IMP dose will be applied at the clinic, in the presence of a qualified unblinded treatment administrator; subsequent applications will be completed by the patient at the clinic (on Day 15/Week 2 Visit) or at their home (all other applications). Accurate dosing will be ensured by using a dosing card (to be used for all study drug applications). Patients will be dispensed a two-week supply of study drug (three tubes of 100 grams) on Day 1 (Baseline/FV1) and on Day 15 (Week 2 Visit) and will be asked to return all unused study drug supplies at the next study visit.

The following study drug exposure and compliance parameters will be summarized using descriptive statistics for Safety population and mITT population, respectively.

- **Number of missed IMP applications** evaluated based on overall target number of IMP applications (the greater of overall duration of exposure up to 30 days OR 26 days multiplied by 4 IMP applications per day) will be calculated by subtracting the number of IMP applications from overall target number of IMP applications. The number and percentage of patients with at least one missed IMP application, 5 to 8 consecutive missed IMP applications, 9 to 12 consecutive missed IMP applications or >12 consecutive missed IMP applications will be provided.
- **Overall treatment period duration of exposure (days)** will be derived as date of last IMP application minus date of first IMP application + 1 (Data source: Patient diary).
    - **Week 1 + 2 duration of exposure (days)** will be derived as last date of IMP application before the date of Week 2 Visit minus date of first IMP application + 1.
    - **Week 3 + 4 (last 2 weeks of study treatment) duration of exposure (days)** will be derived as date of last IMP application minus date of Week 2 Visit + 1.
- **Overall treatment period number of IMP applications (doses)** will be derived by counting the number of times where IMP was applied regardless of days or frequency of applications (Data source: patient diary).
    - **Week 3 + 4 (last 2 weeks of study treatment) number of IMP Applications (doses)** will be derived by counting the number of times where IMP was applied

regardless of days or frequency of applications in week 3 and 4 (starting from date of week 2 visit to the date of last IMP application).

- **Overall treatment period target number of IMP applications** will be derived as (the greater of overall duration of exposure up to 30 days OR 26 days) multiplied by 4 IMP applications per day (Data source: patient diary).
    - **Week 3 + 4 (last 2 weeks of study treatment) target number of IMP applications** will be derived as (the greater of Week 3 + 4 duration of exposure up to 16 days OR 13 days) multiplied by 4 IMP applications per day.
- **Overall treatment period percent compliance (%) based on the number of applications** will be calculated as (overall number of IMP applications divided by Overall treatment period target number of IMP applications) x 100.
    - **Week 3 + 4 (last 2 weeks of study treatment) percent compliance (%) based on the number of applications** will be calculated as (Week 3 + 4 number of

IMP applications divided by Week 3 + 4 target number of IMP applications) x 100.

- **Overall treatment period amount of IMP used (g)** will be derived as the total amount of IMP dispensed (g) minus the total amount of IMP returned (g) (Data source: patient Investigational product dispensing logs).
    - **Week 3 + 4 (last 2 weeks of study treatment) amount of IMP used (g)** will be derived as the amount of IMP dispensed (g) on Week 2 Visit (last kit number) minus the amount of IMP returned (g) on the same kit number.
- **Overall treatment period target amount of IMP used (g)** will be derived as the overall treatment period target number of IMP applications multiplied by 4 g (Note: Overall treatment period target number of IMP applications derived above).
    - **Week 3 + 4 (last 2 weeks of study treatment) target amount of IMP used (g)** will be derived as the Week 3 + 4 target number of IMP applications multiplied by 4 g (Note: Week 3 + 4 target number of IMP applications derived above).
- **Overall treatment period percent compliance (%) based on the amount of IMP used (g)** will be calculated as the amount of IMP used (g) divided by the target amount of IMP used (g) x 100.
    - **Week 3 + 4 (last 2 weeks of study treatment) percent compliance (%) based on the amount of IMP used (g)** will be calculated as the week 3 + 4 amount of IMP used (g) divided by week 3 + 4 target amount of IMP used (g) x 100.
- **Number of Days of Exposure (Overall treatment period)** will be calculated as the number of distinct days where a patient has applied IMP at least once (Data source: patient diary).
- **Average daily exposure (g/day)** will be calculated as overall treatment period amount of IMP used (g) divided by number of days of exposure.
- **Average exposure per dose (g/dose)** will be calculated as the overall treatment period amount of IMP used (g) divided by the overall treatment period number of IMP applications (doses).

All exposure data, including date and time of each IMP application, will be presented in a by-patient data listing.

Relevant Output

Table 14.3.5.1.1	Treatment Compliance Based on Number of Applications (mITT Population)
Table 14.3.5.1.2	Treatment Compliance Based on Amount of IMP Used (mITT Population)
Table 14.3.5.1.3	Treatment Exposure (Safety Population)
Listing 16.2.5.1.1	Study Drug Dispensation (All Randomized Patients)
Listing 16.2.5.1.2	Study Drug Application Diary (Safety Population)
Listing 16.2.5.1.3	Study Drug Administration Summary Parameters - Based on Number of IMP Applications (Safety Population)
Listing 16.2.5.1.4	Study Drug Administration Summary Parameters - Based on IMP Weight (Safety Population)

4.4.2. Adverse Events

All AEs will be coded using the MedDRA coding system (MedDRA version 23.1) and displayed in tables and data listings using system organ class (SOC) and preferred term (PT).

Analyses of AEs will be performed for those events that are considered treatment-emergent, where treatment-emergent is defined as any AE with onset during or after the first application of IMP through the end of the study (Week 6/EOS visit), or any event that was present at baseline but worsened in severity during or after the first application of IMP.

4.4.2.1. Summary of Adverse Events

The following AE subsets will be summarized by the total number of events and the number and percentage of patients with:

- At least 1 treatment-emergent AE (TEAE)
- At least 1 treatment-related TEAE
- At least 1 serious TEAE
- At least 1 treatment-related serious TEAE
- At least 1 TEAE, by maximum severity
- At least 1 TEAE leading to study drug discontinuation
- At least 1 TEAE leading to withdrawal from the study
- At least 1 TEAE leading to death
- At least 1 TEAE related to application site (See Section 7.1 Table 2 for a list of applicable MedDRA PTs)

In addition to the overall summary, each of the above categories will also be summarized and tabulated for each treatment group by SOC and PT. In these tabulations, each subject will contribute only once to the count for a given AE (SOC or PT), regardless of the number of episodes. The total number of AE events will also be summarized.

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

All AEs occurring on-study will be listed in by-patient data listings. By-patient listings also will be provided for serious AEs, AEs leading to study drug discontinuation and withdrawal from the study, and AEs leading to death.

Relevant Output

Table 14.3.1.1 Treatment-Emergent Adverse Events Summary (Safety Population)
Table 14.3.1.2 Treatment-Emergent Adverse Events by System Organ Class and Preferred Term (Safety Population)
Table 14.3.1.3 Treatment-Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term (Safety Population)

Table 14.3.1.4	Serious Treatment-Emergent Adverse Events by System Organ Class and Preferred Term (Safety Population)
Table 14.3.1.5	Serious Treatment-Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term (Safety Population)
Table 14.3.1.6	Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, and Maximum Severity (Safety Population)
Table 14.3.1.7	Treatment-Emergent Adverse Events Leading to Study Drug Discontinuation by System Organ Class and Preferred Term (Safety Population)
Table 14.3.1.8	Treatment-Emergent Adverse Events Leading to Death by System Organ Class and Preferred Term (Safety Population)
Table 14.3.1.9	Application Site Treatment-Emergent Adverse Events by System Organ Class and Preferred Term (Safety Population)
Table 14.3.2.1	Listing of Serious Adverse Events
Table 14.3.2.2	Listing of Adverse Events Leading to Treatment Withdrawal
Table 14.3.2.3	Listing of Adverse Events Leading to Withdrawal from the Study
Table 14.3.2.4	Listing of Adverse Events Leading to Death
Listing 16.2.7.1	Listing of Adverse Events

4.4.3. Laboratory Data

Clinical laboratory values will be expressed using the International System of Units (SI).

The actual value and change from Baseline to each on-study evaluation will be summarized for each clinical laboratory parameter, including hematology, clinical chemistry, and coagulation. In the event of repeat values, the last non-missing value per study day/time will be used.

Shift tables of change from Baseline in level categories (normal, low, high, missing) will be presented by treatment group across all visits.

Urine drug screen results will be summarized and presented by treatment group, including the number and percentage of patients with at least one positive result. Urine drug screen will include amphetamines, barbiturates, benzodiazepines, cannabinoids, cocaine, methadone, opiates, phencyclidine, and a panel of tricyclic antidepressants.

All laboratory data will be provided in data listings, including urine pregnancy test results, urine drug screen results, and serology test results at screening. Data listings will also be presented for all abnormal laboratory values, with a flag indicating clinical significance.

Relevant Output

Table 14.3.5.2.1	Clinical Laboratory Data Summary of Actual Value and Change from Baseline: Hematology (Safety Population)
Table 14.3.5.2.2	Clinical Laboratory Data Summary of Actual Value and Change from Baseline: Chemistry (Safety Population)
Table 14.3.5.2.3	Clinical Laboratory Data Summary of Actual Value and Change from Baseline: Coagulation (Safety Population)
Table 14.3.5.2.4	Urine Drug Screen Results by Visit (Safety Population)

Table 14.3.5.3.1	Clinical Laboratory Shifts from Baseline to Each On-Study Visit: Hematology (Safety Population)
Table 14.3.5.3.2	Clinical Laboratory Shifts from Baseline to Each On-Study Visit: Chemistry (Safety Population)
Table 14.3.5.3.3	Clinical Laboratory Shifts from Baseline to Each On-Study Visit: Coagulation (Safety Population)
Listing 16.2.8.1	Clinical Laboratory Results: Hematology (Safety Population)
Listing 16.2.8.2	Clinical Laboratory Results: Chemistry (Safety Population)
Listing 16.2.8.3	Clinical Laboratory Results: Coagulation (Safety Population)
Listing 16.2.8.4	Clinical Laboratory Results: Urinalysis (Safety Population)
Listing 16.2.8.5	Clinical Laboratory Results: Serology (Safety Population)
Listing 16.2.8.6	Urine Drug Screen (Safety Population)
Listing 16.2.8.7	Pregnancy Test Results (Safety Population)
Listing 16.2.8.8	Listing of Abnormal Clinical Laboratory Parameters (Safety Population)

4.4.4. Vital Signs and Physical Examination

The actual value and change from Baseline to each on-study evaluation will be summarized for vital signs. Vital signs include oral temperature, pulse rate, respiratory rate, systolic blood pressure, and diastolic blood pressure.

Physical examination results (normal; abnormal, not clinically significant; abnormal, clinically significant; not done) at each time point will be summarized.

All vital sign measurements and physical examination findings will be presented for each patient in a data listing.

Relevant Output

Table 14.3.5.4	Summary of Actual Value and Change from Baseline for Vital Signs (Safety Population)
Table 14.3.5.5	Physical Examination (Safety Population)
Listing 16.2.9.1	Vital Signs (Safety Population)
Listing 16.2.9.2	Physical Examinations (Safety Population)

4.4.5. Electrocardiogram

Descriptive statistics (including actual value and change from baseline) will be provided for heart rate (HR), PR interval, QRS interval, QT interval, and QTc interval at Screening Visit (baseline) and Week 4/EOT Visit. QTc intervals calculated and reported on the electronic CRF using methods other than the Bazett method will be calculated using the following formula:

$$QTc = QT / \sqrt{RR}$$

Number and percentage of subjects with normal, abnormal, and clinically significant abnormal results at Screening Visit and Week 4/EOT Visit for overall ECG interpretation will be provided.

Electrocardiogram data for each subject will be provided in a data listing.

Relevant Output

Table 14.3.5.6 Summary of Actual Value and Change from Baseline for Electrocardiogram Results (Safety Population)
Table 14.3.5.7 Electrocardiogram Interpretation by Visit (Safety Population)
Listing 16.2.9.3 12-Lead ECG (Safety Population)

4.4.6. Analysis of Skin Irritation

Skin irritation will be assessed by the Investigator using the Berger/Bowman Scoring Scale (Berger and Bowman, 1982; US FDA Guidance 2018b), which comprises the "Dermal Response" and "Other Effects" scoring system, as outlined in Protocol Appendix 4. The Dermal Response score ranges from 0 (no evidence of irritation) to 7 (strong reaction spreading beyond the application site). The Other Effect score is a combination of a letter and a number with 7 possible ordinal categories: A(0), B(1), C(2), F(3), G(3), H(3).

The Total Skin Irritation Score will be calculated as the sum of the "Dermal Response" score and the numeric equivalent of the "Other Effect".

Total Skin Irritation Score, Dermal Response score, and Other Effect score (in "letter (number)" format) will be summarized descriptively by treatment group and visit.

Missing Dermal Response and Other Effect scores will be imputed using LOCF as described in Section 3.10.3.

Relevant Output

Table 14.3.5.8 Skin Irritation Assessment Results by Visit (Safety Population)
Listing 16.2.9.4 Skin Irritation Assessment

4.4.7. Concomitant Medications

Concomitant medications will be coded using the WHO Drug Dictionary (WHO Drug Global B3 September 2020). Results will be tabulated by anatomic therapeutic class (ATC) Level 3 and PT. In any tabulation, a patient contributes only once to the count for a given ATC class or PT.

Concomitant medications will be tabulated by treatment group, where any medications that did not end prior to first IMP application will be included. Missing and partial dates for concomitant determination will be handled as described in Section 3.10.4.

The use of concomitant medications will be included in a by-subject data listing. A concomitant medication that is used as a rescue medication will be flagged in the listing.

Relevant Output

Table 14.3.5.9 Concomitant Medications (Safety Population)

Listing 16.2.9.5 Prior and Concomitant Medications (All Randomized Patients)

5. CHANGES TO PLANNED ANALYSES

The following are changes between the protocol-defined statistical analyses and those presented in this SAP:

- Section 6.4 of the final study protocol version 4.0) defined inadequate compliance as ≤80% of the planned doses (i.e., <23 out of 28) applied during each week of study treatment. In addition, inadequate compliance requiring permanent treatment discontinuation was defined as >3 days of consecutive study treatment missed. In addition to these criteria, compliance based on the *amount* of IMP applied relative to the target amount of IMP is also addressed on this SAP; see Section 4.4.1.
- Compared to the final study protocol, this SAP (Section 2.1) includes an expanded definition of the PP1 Population, and provisions for a second Per Protocol Population (i.e., PP2) that also includes criteria for minimum compliance based on the *amount* of IMP used by study patients.
- The analysis method for the percentages of patients achieving ≥20%, ≥30% and ≥50% reduction from Baseline in worst daytime pain and in worst nighttime detailed in Section 4.3.2.4 of this SAP was not specified in the final protocol.

6. REFERENCES

Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage. 2019 Nov;27(11):1578-1589. doi: 10.1016/j.joca.2019.06.011.

Bellamy N, Buchanan WW, Goldsmith CH, et al. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988 Dec;15(12):1833-40.

Berger RS and Bowman JP. A Reappraisal of the 21-Day Cumulative Irritation Test in Man. Journal of Toxicology: Cutaneous and Ocular Toxicology. 1982;1(2):109-115. DOI: 10.3109/15569528209051516.

CELEBREX® (celecoxib) capsules - Canadian Product Monograph. Pfizer Canada ULC, Kirkland, Quebec. Date of Revision: March 19, 2019.

CELEBREX® (celecoxib) capsules, for oral use - U.S. Prescribing Information. New York, NY: G.D. Searle, LLC. Revised 5/2019.

Cross M, Smith E, Hoy D, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis. 2014;73(7):1323-1330. doi:10.1136/annrheumdis-2013-204763.

GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016 [published correction appears in Lancet. 2017 Oct 28;390(10106):e38]. Lancet. 2017;390(10100):1211-1259. doi:10.1016/S0140-6736(17)32154-2.

Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet. 2019 Apr 27;393(10182):1745-1759. doi: 10.1016/S0140-6736(19)30417-9.

Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee. Arthritis Rheumatol. 2020 Jan 6. doi: 10.1002/art.41142.

McKenna F, Borenstein D, Wendt H, et al. Celecoxib versus diclofenac in the management of osteoarthritis of the knee: a placebo-controlled, randomised, double-blind comparison. Scandinavian Journal of Rheumatology 2001;30(1):11–8.

Rannou F, Pelletier JP, Martel-Pelletier J. Efficacy and safety of topical NSAIDs in the management of osteoarthritis: Evidence from real-life setting trials and surveys. Semin Arthritis Rheum. 2016 Feb;45(4 Suppl):S18-21. doi: 10.1016/j.semarthrit.2015.11.007.

Rother M, Lavins BJ, Kneer W, et al. Efficacy and safety of epicutaneous ketoprofen in Transfersome (IDEA-033) versus oral celecoxib and placebo in osteoarthritis of the knee: multicentre randomised controlled trial. Annals of the Rheumatic Diseases 2007;66(9):1178–83.

US Food and Drug Administration. Guidance for Industry (Draft): Assessing the Irritation and Sensitization Potential of Transdermal and Topical Delivery Systems for ANDAs. October 2018b. Available at: https://www.fda.gov/media/117569/download

Williams GW, Ettlinger RE, Ruderman EM, Hubbard RC, Lonien ME, Yu SS, et al. Treatment of osteoarthritis with a once-daily dosing regimen of celecoxib: a randomized, controlled trial. Journal of Clinical Rheumatology 2000;6(2):65–74.

7. APPENDIX

7.1. Adverse Events Related to Application Site

The PTs listed in Table 2 will be used to summarize all TEAEs related to study drug application site.

Table 2 MedDRA Preferred Terms for Application Site Adverse Events

System Organ Class Code	System Organ Class	Preferred Term Code	Preferred Term
10018065	General disorders and administration site conditions	10003036	Application site dermatitis
10018065	General disorders and administration site conditions	10003050	Application site oedema
10018065	General disorders and administration site conditions	10065331	Application site fissure
10018065	General disorders and administration site conditions	10049043	Application site papules
10018065	General disorders and administration site conditions	10058730	Application site photosensitivity reaction
10018065	General disorders and administration site conditions	10066209	Application site scab
10018065	General disorders and administration site conditions	10052162	Application site discharge
10018065	General disorders and administration site conditions	10065577	Application site erosion
10018065	General disorders and administration site conditions	10003046	Application site irritation

System Organ Class Code	System Organ Class	Preferred Term Code	Preferred Term
10018065	General disorders and administration site conditions	10064578	Application site exfoliation
10018065	General disorders and administration site conditions	10003054	Application site rash
10018065	General disorders and administration site conditions	10050104	Application site urticaria
10018065	General disorders and administration site conditions	10048943	Application site dryness
10018065	General disorders and administration site conditions	10003041	Application site erythema
10018065	General disorders and administration site conditions	10003053	Application site pruritus
10018065	General disorders and administration site conditions	10072694	Application site haemorrhage
10018065	General disorders and administration site conditions	10003055	Application site reaction
10018065	General disorders and administration site conditions	10048941	Application site vesicles