Mylan

Statistical Analysis Plan	
Study code	MYL-1601N-3002
Version	Version 2 Final
Date	*11/Jan/2018*

A RANDOMIZED, DOUBLE BLIND, THREE-ARM, PARALLEL, PLACEBO-CONTROLLED, CLINICAL STUDY TO EVALUATE THE BIOEQUIVALENCE USING CLINICAL ENDPOINT OF DICLOFENAC SODIUM GEL, 1% (MYLAN INC.) TO VOLTAREN® GEL (DICLOFENAC SODIUM TOPICAL GEL) 1% (NOVARTIS CONSUMER HEALTH, INC.) IN PATIENTS WITH OSTEOARTHRITIS (OA) OF THE KNEE

Study Statistician	[redacted]	JAN 25, 2018 14:18 EET Date
Biostatistics Representative	[redacted]	Jan 25, 2018 8:30 EST Date
Clinical Representative(s)	[redacted]	25 Jan 2018 19:09 IST Date
		25 Jan 2018, 18:49 IST Date
Medical writing Representativ	[redacted]	25 Jan 2018, 21:00 IST Date

Table of Contents

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse Event
ALT	Alanine transaminase
ANCOVA	Analysis of Covariance
AST	Aspartate transaminase
BMI	Body Mass Index
CI	Confidence Interval
eCRF	Electronic Case Report Form
CRO	Contract Research Organization
ECG	Electrocardiogram
FDA	Food and Drug Administration
LOCF	Last Observation Carried Forward
LS	Least Squares
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent to Treat
OA	Osteoarthritis
OTC	Over the Counter
POM	Pain On Movement
PP	Per Protocol
PSRM	Product Safety and Risk Management
RLD	Reference Listed Drug
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SD	Standard Deviation
SE	Standard Error
SGOT	Serum Glutamic Oxaloacetic Transaminase
SGPT	Serum Glutamate-Pyruvate Transaminase
TEAE	Treatment-emergent Adverse Event
US	United States
VAS	Visual Analog Scale
WHO	World Health Organization
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index

CHANGES FROM VERSION 1 TO VERSION 2

1. Additional subgroup descriptive summaries of WOMAC Pain Subscale score (for subjects with baseline WOMAC Pain Subscale score <= 10 and by Kellgren-Lawrence grade)
2. Additional summary of percentage of patients that used acetaminophen (paracetamol) at least once during the trial.
3. Average daily dose of acetaminophen is calculated only over days when acetaminophen was used.
4. Added forest plot for test/reference ration in various subgroups.

1. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol "A Randomized, Double blind, Three-arm, Parallel, Placebo-controlled, Clinical Study to Evaluate the Bioequivalence using Clinical Endpoint of Diclofenac Sodium Gel, 1% (Mylan Inc.) to Voltaren® Gel (Diclofenac Sodium Topical Gel) 1% (Novartis Consumer Health, Inc.) in Patients with Osteoarthritis (OA) of the Knee" version 1.1 dated May 3, 2017.

Mylan is developing a generic diclofenac sodium gel to the Reference Listed Drug (RLD) Voltaren® gel. This randomized, double-blind, three-arm, placebo-controlled, bioequivalence study with clinical endpoint has been designed to establish clinical equivalence and safety of Mylan's diclofenac gel in the symptomatic treatment of osteoarthritis of knee compared to Voltaren® gel and to establish superiority in efficacy of both compared to a placebo (vehicle) gel.

This SAP describes the study endpoints, derived variables, anticipated data transformations and manipulations, and other details of the analyses not provided in the study protocol. The SAP also outlines the statistical programming specifications for the tables, listings and figures.

2. OBJECTIVES

The primary objectives of this study are:

1. To determine the clinical equivalence of the Test Drug (Diclofenac sodium gel of 1% of Mylan Inc.) with the Reference Listed Drug (RLD) (Voltaren® Gel 1% of Novartis Consumer Health, Inc.) as measured by mean change from baseline in the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) Likert (version 3.1) pain score.
2. To establish superiority of the Test Drug and Reference Listed Drug (RLD) over placebo in mean change from baseline in the WOMAC Likert (version 3.1) pain score.

The secondary objectives of this study are:

1. To assess the safety and tolerability of study treatments.
2. To determine the clinical equivalence of the Test Drug (Diclofenac Sodium gel) with the Reference Listed Drug (RLD) (Voltaren® Gel) as measured by mean change from

baseline in subjects whose baseline WOMAC Likert (version 3.1) pain scores are less than or equal to 10.

3. STUDY OVERVIEW

3.1 Study Design

This is a double-blind, multiple-site, randomized, parallel-design study to investigate the bioequivalence with clinical endpoint of Mylan's Diclofenac Sodium Topical Gel, 1% to Voltaren® Gel, 1% (Novartis US) and to investigate that both the active products are superior to placebo.

Subjects will have to visit the clinic for the following visits during the study:

- Visit 1 – Screening visit (within 28 days prior to randomization)
- Visit 2 – Start of Placebo Run-in period (Day -11 to Day -7)
- Visit 3 – Randomization (baseline) visit (Day 0)
- Visit 4 – Interim visit (Week 2/Day 14 ± 4)
- Visit 5– End of study (Week 4/Day 28 ± 4)/Early termination visit

After informed consent process and completion of all screening assessments, the eligible subjects will be requested to stop the currently ongoing osteoarthritis drug therapy. The wash out period (run-in period) of 7 days or ≥ 5 half-life of previous osteoarthritis drug therapy, whichever is longer, will be maintained before the baseline visit. Subjects will undergo at least 7 days of placebo run-in period. In case subject extends run-in period beyond 11 days, continuation of the subject will be decided in consultation with Medical monitor.

The Pain on Movement (POM) score on a Visual Analog Scale (VAS) and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC Likert version 3.1) pain score will be recorded at the start of run-in Period and at the end of run-in Period (i.e. immediately prior to randomization on the day of baseline visit). Recording of the VAS will be done for both knees and recording of the WOMAC pain score will be done for the target knee. Target knee is defined as the knee that is more painful and will be identified during the screening visit.

During the washout period, subjects will undergo a single-blind placebo run-in period where the subjects will not be aware of the placebo nature of the drug. Subjects will be provided with

placebo gel for the administration. Approximately 4 g dose of placebo gel will be applied to the target knee four times daily for at least 7 days. First administration of gel will be done at the study center.

At the baseline visit, subjects will be randomly assigned in 1:1:1 ratio to Mylan's Diclofenac Sodium Topical Gel 1%, Voltaren® Gel 1% (Reference Listed Drug) and Placebo. All the subjects will be instructed to apply the gel next day (Day 1) onwards. All subjects will be applying approximately 4 g dose of either Mylan's Diclofenac Sodium Topical Gel 1% or Reference Listed Drug (RLD) or Placebo gel to the arthritic target knee four times daily for 4 weeks.

The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC Likert version 3.1) pain score for the target knee and pain on movement (POM) on a Visual Analog Scale (VAS) for both the knees will be recorded on Week 2 and Week 4 from the start of the treatment.

Acetaminophen (paracetamol) tablets will be given to the subjects for use as a rescue medication or for the treatment of aches and pains unrelated to knee pain such as headache during treatment period (except 3 days prior to Visit 4 and Visit 5). Maximum allowed dose of acetaminophen (paracetamol) will be 2 g/day. Consumption of rescue medications in terms of number of tablets consumed will be monitored.

Subject diary will be provided to subjects to record study medication gel application, rescue medication, side effects, and concomitant medication details. Subject diary will be used to evaluate drug application compliance.

Physical examination will be conducted during each visit (except at the start of run-in period visit). Application site reaction assessment will be conducted on each visit after start of run-in period.

A standard 12-lead ECG will be recorded after 5 minutes of rest in the supine position on screening visit and on end of study visit. Vital signs will be measured after 5 minutes of rest (sitting) on each visit.

Serum pregnancy test for females of childbearing potential will be performed at screening and end of treatment visit. The urine pregnancy test can be performed if clinically indicated during other visits. If urine pregnancy is found to be positive, serum pregnancy will be drawn and analyzed for confirmation. If serum pregnancy test will found positive then patient will be excluded from the study and followed up.

Laboratory assessments like hematology and biochemistry will be performed during screening and end of study. The hematology and biochemistry at screening is to be performed within 14 days of baseline visit. If the gap between lab assessments and baseline visit will be more than 14 days, the lab assessments will be repeated at baseline visit. Laboratory investigations (if clinically significant) can be repeated once as per investigator's discretion.

Principal investigator or designee will contact the subjects approximately midway between the visit 3-4 and visit 4-5 for assessment of the well-being, adverse event, concomitant medication, and treatment compliance. Additional telephonic contact to reiterate the instructions regarding the use of rescue medications during 3 days prior to visits 4 and 5 will be done. Use of rescue medications by subjects during this period for intolerable pain will require rescheduling the visits in such a way that visits happen after at least 3 days of stopping rescue medication and subjects will be required to inform the center about the rescue medication use.

Study design is represented graphically in Figure 1 Study Design.

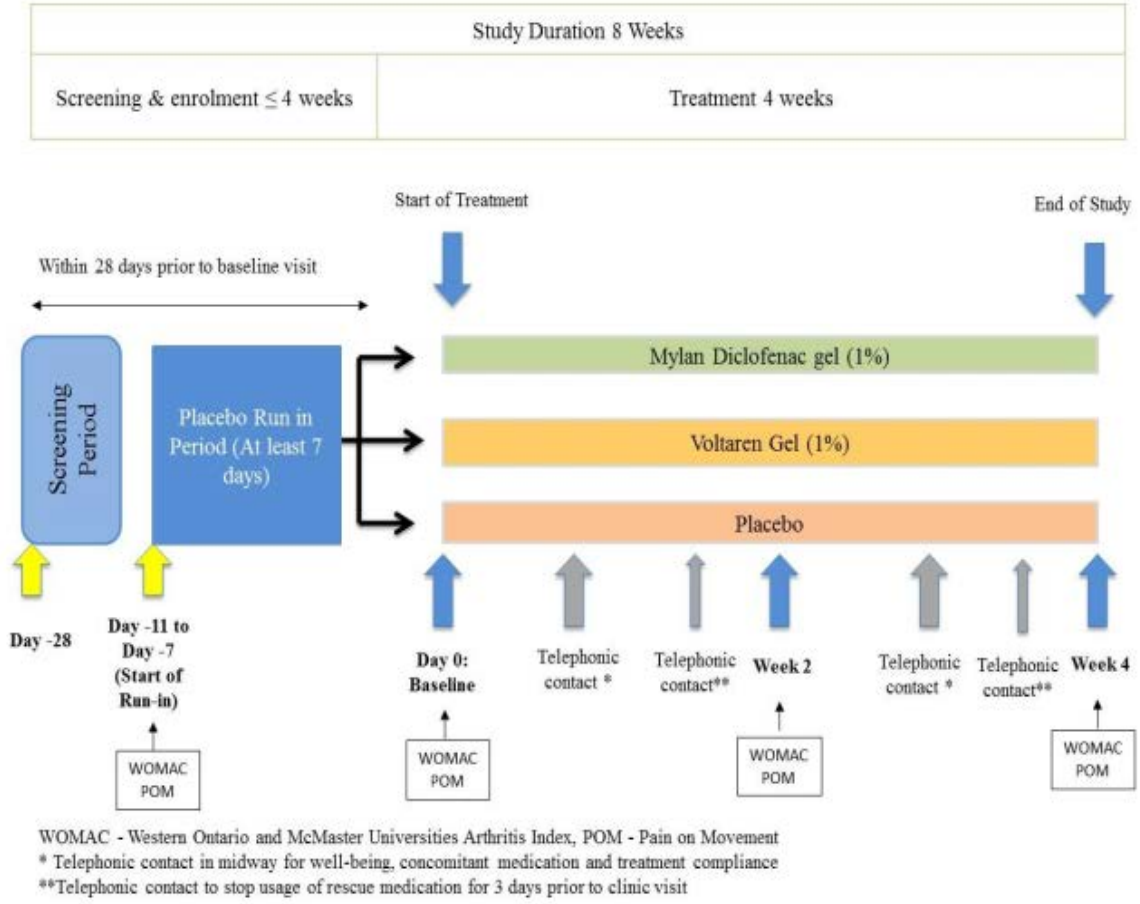

Figure 1 Study Design

3.2 Sample Size

The study protocol, MYL-1601N-3002, is designed to demonstrate the Bioequivalence of Diclofenac Sodium Gel to Voltaren® Gel and the superiority of Diclofenac Sodium Gel and/or Voltaren® Gel over Placebo by using a clinical endpoint in patients with Osteoarthritis (OA) of the Knee. The clinical endpoint under consideration is the change from baseline Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) Likert (version 3.1) pain score. The framework under consideration and the sample size requirements were fine-tuned to satisfy the objectives of bioequivalence of test drug with the reference drug as well as superiority of the test drug to placebo. The sample size estimation is done to ensure 90% statistical power to test two hypotheses of bioequivalence and superiority, the hypotheses are briefly described below:

1. Null Hypothesis (Equivalence) is that the ratio of mean change from baseline in WOMAC Likert pain score between test and reference drug is less than or equal to 0.80 or greater than or equal to 1.25 Vs. the alternative hypothesis of ratio of mean change

from baseline in WOMAC Likert pain score between test and reference drug lies within the limits of 0.80 to 1.25, where the equivalence limit of 0.80 to 1.25 is adapted.

2. Null Hypothesis (Superiority): The mean change from baseline WOMAC Likert pain score of test or reference drug is greater than or equal to Placebo vs the alternative hypothesis of mean change from baseline WOMAC Likert pain score of test or reference drug is less than the placebo group.

To meet the requirement for the two hypotheses, three tests need to be performed to meet overall objectives:

1) Bioequivalence between Diclofenac Sodium Gel and Voltaren®
2) Superiority of Diclofenac Sodium Gel over Placebo
3) Superiority of Voltaren® over Placebo

Sample size calculations are performed under the following assumptions:

1. Mean (± Standard Deviation) reduction from baseline in WOMAC pain score is 5.85 (±4.23) for both active treatments.
2. Mean (± Standard Deviation) reduction from baseline in WOMAC pain score is 4.68 (±4.89) for placebo.

3.2.1 Equivalence between test and reference product

At least 261 subjects in Test Drug and 261 subjects in Reference Drug group will be required to show bioequivalence between Test and Reference product with 90% power and 5% level of significance assuming the following estimates:

1) 5% difference between Test and Reference of mean change in WOMAC score
2) 75 % inter subject coefficient of variation for Reference in WOMAC Score
3) Equivalence margin of 80.00-125.00%

3.2.2 Superiority of test and reference product over placebo

At least 363 subjects will be required in each arm to prove the superiority of test/reference drug to placebo with 90% power and 5% level of significance.

Considering 10% dropout rate, around 1212 subjects will be required to enroll in the study. All subjects will be randomized in 1:1:1 ratio.

Overall, to satisfy both hypotheses of equivalence and superiority, the highest sample size will be used in the study: 1212 subjects equally randomized to Test, Reference and Placebo treatments:

1. Diclofenac Sodium Gel: 404 subjects.

2. Voltaren® Gel : 404 subjects
3. Placebo: 404 subjects

3.3 Randomization and Unblinding Procedures

A computerized randomization will be generated by an independent statistician (not involved in the packaging and labelling of the study medication).

Subjects will be randomly assigned in 1:1:1 ratio to three groups. Treatment allocation can be Mylan's Diclofenac Sodium Topical Gel 1%, Voltaren® Gel 1% (Reference Listed Drug), or Placebo. A balanced randomization schedule will be generated using SAS® software (Version: 9.3 or higher; SAS Institute Inc., USA). The randomization code will be held by an independent statistician throughout the conduct of the study in order to minimize bias.

A sealed copy of the randomization code for each kit provided under separate cover inside the block will be retained at investigational center. It should be available to FDA investigators at the time of center inspection to allow for verification of the treatment identity of each subject.

Placebo run-in period will be single blinded, in which subjects will not be aware of the treatment given.

Treatment period will be double blinded, in which all the investigators, subjects, center staff, CRO team, study monitors, laboratory personnel and/or other designated individuals will be blinded to the medication codes. The formulation of test, reference and placebo investigational products will be identical in appearance to make any difference in treatment less obvious to the subjects and to maintain adequate blinding of evaluators.

Planned unblinding will occur after the database lock.

Unplanned unblinding (prior to database lock) will be performed by a designated person in case of SAE or other significant events. Mylan Product Safety and Risk Management (PSRM) can unblind the treatment assignment for any subject with a SAE which qualifies for the unblinding based on country specific regulatory requirements.

The event of unplanned unblinding prior to database lock shall be documented in the source notes. Reasons for treatment unplanned unblinding must be clearly explained and justified in the

eCRF. The date on which the code was broken together with the identity of the person responsible must also be documented.

If a subject's treatment assignment is unblinded prior to database lock for any reason then the subject should be withdrawn from the study.

4. STUDY ENDPOINTS/OUTCOMES

Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline to Week 4 in total WOMAC Pain Subscale score for the Target Knee.

Secondary Efficacy Endpoints

The secondary efficacy endpoints are:

- The change from baseline to Week 2 in total WOMAC Pain Subscale score for the Target Knee.
- The change from baseline to Week 4 in total WOMAC Pain Subscale score for the Target Knee for the subgroup of subjects whose baseline WOMAC pain score is less than or equal to 10.
- The change from baseline to Week 4 in POM VAS score for the Target Knee.
- Number of days of acetaminophen (paracetamol) consumption during the trial.
- Average daily dose of acetaminophen (paracetamol) during the trial

Safety Endpoints

1. The incidence of treatment-emergent adverse events
2. Application site reaction evaluation

5. HYPOTHESES TESTING

Hypothesis of Equivalence (Primary Endpoint)

The null hypothesis to be tested is that the ratio of mean change from baseline in WOMAC Likert pain score between test and reference drug is less than or equal to 0.80 or greater than or equal to 1.25 vs the alternative hypothesis that this ratio is within the interval from 0.80 to 1.25:

H_0: $\mu_T / \mu_R \leq \theta_1$ or $\mu_T / \mu_R \geq \theta_2$ versus H_A : $\theta_1 < \mu_T / \mu_R < \theta_2$

Where μ_T = mean of test treatment, and μ_R = mean of reference treatment

Typically, H_0 will be rejected with a type I error $\alpha = 0.05$ (two 1-sided tests), if the 90% confidence interval for the ratio of means between test and reference products (μ_T / μ_R) is contained within the interval $[\theta_1, \theta_2]$, where $\theta_1 = 0.80$ and $\theta_2 = 1.25$.

A two-sided, 90% confidence interval on the test/reference ratio for mean change from baseline to Week 4 in total WOMAC Pain Subscale score for the Target Knee will be constructed using Fieller's method. The estimates of treatment means and standard errors will be obtained from Analysis of Covariance (ANCOVA) of the Test and Reference results, using a statistical model containing fixed effects for treatment and center and baseline WOMAC Pain Subscale score as covariate.

Bioequivalence will be established if the 90% confidence interval for the ratio of test/reference means is contained within the interval [0.80, 1.25] for Per Protocol population.

Hypothesis of Superiority (Primary Endpoint)

The null hypothesis to be tested is that there is no difference in the mean change from baseline to Week 4 in total WOMAC Pain Subscale score for the Target Knee between the active treatment and the Placebo treatment vs the alternative hypothesis that the reduction in the WOMAC Pain Subscale score is greater for both active treatments than for placebo:

H_0: $\mu_T = \mu_P$ and $\mu_R = \mu_P$ versus H_A: $\mu_T < \mu_P$ and $\mu_R < \mu_P$

Where μ_T = mean of test treatment, μ_R = mean of reference treatment, μ_P = mean of Placebo treatment.

The hypothesis testing will be performed separately for the Test treatment versus the Placebo treatment and for the Reference treatment versus the Placebo treatment using Analysis of Covariance (ANCOVA) with a statistical model containing fixed effects for treatment and center and baseline WOMAC Pain Subscale score as covariate.

Superiority will be established if the mean change from baseline for each active treatment is less than, and statistically significantly different from ($p<0.05$), that for the Placebo using the

modified intent-to-treat (mITT) population and with Last Observation Carried Forward (LOCF) approach.

Hypotheses of Equivalence and Superiority (Secondary Endpoints)

The secondary endpoints will be evaluated in the same way as the primary endpoint, but will be considered as supportive information.

6. ANALYSIS SUBSETS

6.1 Run-in Population

The run-in population includes all subjects who make at least one application of placebo gel in the run-in period of the study.

6.2 Randomized Population

The randomized population consists of all subjects who were randomized.

6.3 Safety Population

A safety population subject is any individual who was randomized into the study and has evidence of usage of at least one dose of the randomized investigational product according to subject's diary. For subjects who were randomized but fail to return their diary and for whom no evidence of study drug use is thus available, the following conservative approach will be applied:

- if some safety or efficacy data for these patients is available after Visit 3, they will be included in the safety population;
- if no safety or efficacy data for these patients is available after Visit 3, they will be considered lost to follow-up and excluded from the safety population.

The safety population will be the primary population for the safety analysis. Analyses on the safety population will be based on actual treatment received, as determined by kit schedule.

6.4 Intent to Treat Population (ITT Population)

Following the intent-to-treat principle, the ITT population includes all randomized subjects. Analyses on the ITT population will be based on randomized treatment, as determined by randomization schedule. This population will be used for sensitivity analysis of both equivalence and superiority, LOCF method will be used for imputing missing data.

6.5 Modified Intent to Treat Population (mITT Population)

The mITT population includes all Safety population subjects who met all inclusion/exclusion criteria and return for at least one post-baseline evaluation of WOMAC Pain Subscale.

Subjects who discontinued early and have at least one post-baseline efficacy evaluation will be included in the mITT population using LOCF. This population will be used for testing superiority. Analyses on the mITT population will be based on randomized received, as determined by randomization schedule.

6.6 Per Protocol Population (PP Population)

The Per Protocol (PP) population includes all mITT subjects who:

- met all inclusion/exclusion criteria
- apply 75% to 125% (both inclusive) of the scheduled applications of the assigned product
- return for the 4 week evaluation of WOMAC Pain Subscale score within the visit window (+/- 4 days; i.e., study Day 24-32) or discontinue from the study due to treatment failure (included as treatment failures)
- have no major protocol deviations impacting the primary efficacy outcome measure. These major protocol deviations are described in section 7.9. The list of protocol deviations which lead to exclusion of subjects from the PP population will be finalized at the Blind Data Review (BDR) meeting to be held prior to database lock.

This population will be used for testing the clinical equivalence.

7. STATISTICAL METHODS OF ANALYSIS

7.1 General Principles

If not otherwise specified, statistical significance is defined as $p<0.05$ and is two-tailed. No adjustment will be made for multiplicity. Data will be summarized with respect to demographic and baseline characteristics, efficacy variables and safety variables.

For categorical variables, the number and percent of each category within a parameter will be calculated for non-missing data. For continuous variables, statistics will include number of observations, mean, standard deviation, median, minimum and maximum values.

All statistical analyses will be conducted using SAS®, Version 9.3 or higher. Data will be summarized using descriptive statistics for continuous variables. Unless otherwise specified, descriptive statistics will include the number of subjects, mean, standard deviation, minimum, median and maximum. A subset of these statistics may be presented in some tables. The minimum and maximum statistics will be presented to the same number of decimal places as the original data. The mean and median will be presented to one more decimal place than the original data. The standard deviation will be presented to two more decimal places than the original data.

In by-visit summary tables only scheduled visits will be presented. All visits, both scheduled and unscheduled, will be presented in listings.

Baseline will be defined as the last assessment, scheduled or unscheduled, made prior to the first application of the randomized study drug. For subjects who are randomized, but do not receive the randomized study drug, baseline will be defined as the last assessment, scheduled or unscheduled, made prior to the randomization.

7.1.1 Derivation of dates

The start of the run-in period will be taken from the "Run-In Period Day 1 Date" field collected at the "Run-In Period Dosing Log" CRF page.

The end of the run-in period will be derived as the date corresponding to the maximum day recorded at the "Run-In Period Dosing Log" CRF page when gel was applied.

The start of the treatment period, i.e. the date of the first application of the randomized study drug, will be taken from the "Treatment Period Day 1 Date" field collected at the "Treatment Period Dosing Log" CRF page.

For subjects who started the treatment period, the end of the treatment period, i.e. the date of the last application of the randomized study drug, will be taken from the "Date of Last Gel Application" field collected at the "End of Study" CRF page. It must correspond to the maximum day recorded at the "Treatment Period Dosing Log" CRF page when gel was applied or the maximum date recorded at the "Treatment Period Dosing Log (Extended)" CRF page, if any.

Study days will be defined relative to the start of the treatment period. For dates after the start of the treatment period:

Study Day = Date of Assessment – Start of the treatment period + 1

For dates prior to the start of the treatment period:

Study Day = Date of Assessment – Start of the treatment period

All dates will be displayed in DDMMMYYYY format.

7.1.2 Adjustment for multiplicity

The study will test three hypotheses:

1. Equivalence of the Test and Reference treatments
2. Superiority of Test to Placebo treatment
3. Superiority of Reference to Placebo treatment

Possible inflation of Type I error rate will be controlled by declaring the study a success only if all 3 hypotheses are simultaneously true. Therefore there will be no need to apply any multiplicity adjustment methods.

7.2 Subject Disposition

The number of subjects screened for the study, failing screening (with breakdown by reason), number of subjects entering the run-in period, number of subjects discontinued during the run-in period (with break-down by reason for discontinuation) will be presented.

The number of subjects randomized in the double-blind period of the study, included in the safety, intent to treat, modified intent to treat, per protocol populations, prematurely discontinued from the study (along with the reasons for discontinuation) will be presented by treatment. Number of subjects enrolled by center will be tabulated for all randomized subjects, safety, mITT and PP populations.

7.3 Inclusion/Exclusion criteria

Number and percentage of subjects violating each inclusion/exclusion criterion will be presented by criterion. This analysis will be based on all screened subjects.

Violation of inclusion/exclusion criteria will be listed.

7.4 Diagnosis of Osteoarthritis according to ACR Criteria

Number and percentage of subjects answering Yes to the following CRF questions will be presented by treatment for the safety population:

- Is the subject having knee pain?
- Did the Subject meet at least 3 of the following 6 criteria for the target knee?
    - Age > 50 years
    - Stiffness < 30 minutes
    - Presence of crepitus
    - Bony tenderness
    - Bony enlargement
    - No Palpable warmth

Kellgren-Lawrence grade will also be summarized descriptively, presenting the number and percentage of subjects with each grade (0 to 4).

All details of the osteoarthritis diagnosis collected on the CRF will be listed.

7.5 Joint examination

Results of joint examination will be summarized descriptively for each question asked at the Joint Examination CRF page. Results also will be listed.

7.6 Demographic and Baseline Characteristics

Demographic characteristics will include:

- age
- gender
- race
- ethnicity.

Baseline characteristics include:

- height
- weight
- body mass index (BMI)
- Baseline POM VAS target knee pain score

- Baseline WOMAC Pain Subscale score.

Descriptive statistics will be presented for age (years), BMI, VAS and WOMAC scores. Frequency counts and percentage will be presented for sex, race, ethnicity. BMI will be reported in kg/m^2.

Age will be derived from Informed Consent Signed Date (INFCSD) and Date of Birth (DOB) as the number of whole years between those two dates.

Demographic and baseline characteristics will be evaluated for comparability across treatment groups in the following manner. Continuous variables (age, BMI, VAS and WOMAC scores) will be analyzed with an analysis of variance with factors of treatment and investigational center. Overall p-value for the global null hypothesis of all groups being equal will be displayed. Categorical variables (gender, ethnicity, race) will be analyzed with a Cochran-Mantel-Haenszel general association test, stratified by investigational center.

These analyses will be performed for the ITT, Safety, mITT and PP populations.

Demographic and baseline characteristics will be presented in the by-subject listings.

7.7 Medical History

Medical history will be recorded at Visit 1 Screening. Reported terms will be coded using MedDRA dictionary version 20.0.

Medical history will be summarized by MedDRA system organ class and preferred term for the safety population.

7.8 Prior and Concomitant Medication

Prior and concomitant medications will be assessed at screening and at each subsequent study visit. Medications will be coded using WHO Drug Dictionary version Sep-2016 B2.

Medication will be classified as prior, if the end date is known and is prior to the first use of the study medication (any study medication, including run-in). Medications that are ongoing or ended after the first use of the study medication will be classified as concomitant. If the end date of the medication is unknown, it will also be considered concomitant. If the end date is partially known, the medication will be considered concomitant, unless the known part of the end date

rules out the possibility that the medication could have ended after the start of the study drug. Please see appendix 11.3 for more details on handling of partial dates.

Prior and concomitant medications will be separately summarized by ATC class (the highest available level) and preferred name for the safety population.

The prior and concomitant medications will be listed in a by-subject listing.

7.9 Protocol Deviations

Protocol deviations will be derived algorithmically and additionally reported by study centers according to the MYL-1601N-3002 Protocol Deviations Guidance Document. Eventually all approved protocol deviations will be entered in the Protocol Deviations eCRF page.

Subjects with protocol deviations deemed to have an impact on treatment evaluation will be excluded from the Per Protocol population. The list of protocol deviations which lead to exclusion of subjects from the PP population will be finalized at the Blind Data Review (BDR) meeting to be held prior to database lock.

Major deviations that have an impact on treatment evaluation may include:

- Violation of inclusion/exclusion criteria
- Use of prohibited medications
- Missing WOMAC Pain Subscale assessments
- Final visit out of window
- Missing visits
- Violation of rescue medication use rules
- IP compliance <75% or >125%
- Administration of incorrect or expired IP

Protocol deviations will be summarized by deviation category and treatment group. This analysis will be performed for all randomized subjects.

7.10 Handling of Missing Data

For demographic and baseline characteristics, each variable will be analyzed using all available data. Subjects with missing data will be excluded only from analyses for which data are not

available. For efficacy analysis, patients who discontinued the study early will be included in the ITT and mITT analysis with LOCF approach; subjects who discontinue the study due to lack of treatment effect will also be included in the PP analysis with LOCF approach.

For WOMAC pain total score, the following imputation method will be used for handling missing values. If one item is missing, the missing value will be imputed by averaging remaining 4 items and total score will be calculated using sum of 5 items. If more than one items are missing, then the total pain score will be assigned to missing value at that visit. The LOCF method will be used for endpoint analysis if Visit 5 value is missing.

7.11 Efficacy Analyses

All efficacy analyses will be based on the randomized treatment group as determined by the randomization schedule.

7.11.1 Center pooling

To eliminate potential effect of random fluctuations at small center on the efficacy endpoints small centers will be pooled. A study center will be pooled if it doesn't meet both of the following conditions:

- It has at least 10 patients in the Per Protocol population;
- It has at least one patient in each treatment group in the Per Protocol population.

The smallest center that does not meet the above requirements will be pooled with the next smallest center. The procedure will be repeated until all pooled centers meet the above two requirements.

Pooled center will be used in all efficacy analyses.

7.11.2 Analyses of Primary Endpoint

At all post-screening visits the subjects will complete the WOMAC Osteoarthritis Pain Subscale Ratings for the target (most severe) knee. The Pain Subscale score will be reported on the eCRF.

Missing Week 4 assessments will be imputed as follows. In the ITT and mITT analysis, if a subject discontinued from the study prior to Week 4 or misses the Week 4 assessment for any

other reason, the LOCF rule will be used to impute the score (the change in the score from the baseline visit to the last completed visit prior to discontinuation will be carried forward). In the Per Protocol analysis, if a subject discontinued from the study prior to Week 4 due to treatment failure, the LOCF rule will be used to impute the score (the change in the score from the baseline visit to the last completed visit prior to discontinuation due to treatment failure will be carried forward). If a subject is missing the Week 4 assessment for any other reason, the subject will be excluded from Per Protocol population.

The baseline score will be defined as the results from the latest exam prior to the start of study drug applications. For each post-baseline visit the change from baseline in the WOMAC Pain Subscale score will be calculated. The baseline value, time point value and change from baseline to the timepoint will be analyzed using descriptive statistics and will be tabulated by center, visit and treatment group. This summary will also be prepared for the subset of subjects with baseline WOMAC Pain Subscale score ≤ 10.

Mean WOMAC Pain Subscale score will also be plotted graphically by visit with +/- SD bars. A separate line will be plotted for each treatment group.

7.11.2.1 Analysis of clinical equivalence of test and reference treatments

To show the clinical equivalence, estimates of LS mean change from baseline to Week 4 in the WOMAC Pain Subscale score will be calculated for the test and reference treatment, and then the 90% CI for the LS mean ratio will be constructed using Fieller's method. Bioequivalence will be established if the 90% confidence interval for the ratio of test/reference LS means is contained within the interval [0.80, 1.25].

To this end, first an ANCOVA model will be fit with change from baseline to Week 4 in the WOMAC Pain Subscale score as outcome, baseline WOMAC Pain Subscale score as covariate and treatment, center and treatment-by-center interaction as factors on the data from test and reference treatments only (excluding Placebo subjects):

$$Y_{ijk} = b + a \cdot B_{ijk} + T_i + C_j + TCI_{ij} + \varepsilon_{ijk}$$

Where

- Y_{ijk} is the change from baseline to Week 4 in WOMAC Pain Subscale score for the kth subject in the jth center with ith treatment
- b is model intercept
- B_{ijk} is the baseline WOMAC Pain Subscale score for the kth subject in the jth center with ith treatment
- a is the model parameter associated with baseline WOMAC Pain Subscale score
- T_i is the effect of the ith treatment
- C_j is the effect of the jth center
- TCI_{ij} is the interaction effect of the ith treatment and the jth center
- ε_{ijk} is the random error (normally distributed with zero expected value)

If the treatment-by-center interaction factor is not significant at the 0.05 level, the model will be rerun without the interaction term. Treatment LS means and standard errors will be estimated from this model. Then Fieller's formula will be applied; covariance between treatment means will be assumed to be 0. See appendix 11.2 for complete description of Fieller's formula.

Equivalence will be illustrated graphically by a plot including a vertical line corresponding to the 90% CI around the Test/Reference ratio with the ratio itself indicated by a dot with reference lines for 80% and 125%.

Analysis of bioequivalence will be performed primarily on the Per Protocol population, and additionally on the ITT and mITT populations as a sensitivity analysis.

7.11.2.2 Analysis of superiority to placebo control

The analysis of superiority will be performed for the primary endpoint separately for the Test treatment versus the Placebo treatment and for the Reference treatment versus the Placebo treatment. Each of these analyses will be performed using an ANCOVA model with change from baseline to Week 4 in the WOMAC Pain Subscale score as outcome, baseline WOMAC Pain Subscale score as covariate and treatment, center and treatment-by-center interaction as factors (mathematical statement of this model is the same as for the equivalence model). If the treatment-by-center interaction factor is not significant at the 0.05 level, the model will be rerun without the interaction term. The model will be fit on data from the Test and Placebo treatment

for analysis of superiority of the Test treatment and separately on data from the Reference and Placebo treatment for analysis of superiority of the Reference treatment. From this model, the least square (LS) mean estimate for each treatment group with the 95% CI will be calculated; further, an estimate of the LS mean difference between the active treatment (Test or Reference) and Placebo with 95% CI and the p-value for test of no difference will be calculated.

Superiority will be established if the mean change from baseline to Week 4 for each active treatment is estimated to be less than, and statistically significantly different from ($p<0.05$ for test of no difference) that for the Placebo.

Superiority will be illustrated graphically by a plot including two vertical lines (for Test drug vs. Placebo and Reference Drug vs. Placebo) corresponding to the 95% CI around the Active – Placebo difference with the difference itself indicated by a dot with reference line for 0 difference.

Analysis of superiority will be performed primarily using the mITT population, and additionally on the ITT and PP populations as a sensitivity analysis.

7.11.3 Analyses of Secondary Endpoint Outcomes

The following secondary endpoints will be evaluated in the same way as the primary endpoint, but will be considered as supportive information.

- The change from baseline to Week 2 in total WOMAC Pain Subscale score for the Target Knee.
- The change from baseline to Week 4 in total WOMAC Pain Subscale score for the Target Knee for the subgroup of subjects whose baseline WOMAC pain score is less than or equal to 10.
- The change from baseline to Week 4 in POM VAS score for the Target Knee.

Missing Week 2 assessments will be imputed as follows. In the ITT and mITT analysis, the LOCF rule will be used to impute the score, i.e. the change in the score from the baseline visit to the Early Termination visit, provided it occurs after baseline and within 18 days, (2 weeks + 4 days allowed window) will be carried forward. In the Per Protocol analysis, if a subject

discontinued from the study prior to Week 2 due to treatment failure, the LOCF rule will be used to impute the score as for ITT and mITT analyses. If a subject is missing the Week 2 assessment for any other reason, the subject will be excluded from Per Protocol analysis.

In the analysis of secondary endpoints a 95% CI for the Test/Reference ratio will also be presented in addition to 90% CI.

A descriptive summary by visit will also be created for POM VAS score.

Mean POM VAS score will also be plotted graphically by visit with Mean +/- SD error bars. A separate line will be plotted for each treatment group.

The following secondary endpoints will be summarized descriptively by treatment group:

- Number of days of acetaminophen (paracetamol) consumption during the trial.
- Average daily dose of acetaminophen (paracetamol) during the trial
- Total dose (mg) of acetaminophen (paracetamol) during the trial
- Percentage of patients that used acetaminophen (paracetamol) at least once during the trial

When calculating these endpoints, only days from the day of the first application of the randomized study drug to the day of the last application of the randomized study drug will be counted.

Number of days with any acetaminophen consumption will be counted from the subject's diary.

Average daily dose of acetaminophen will be calculated as follows:

(Sum over all days from the first application of the randomized study drug to the last application of the randomized study drug of [Strength (mg) * Number of Tablets]) / ([Date of the last application of the randomized study drug] - [Date of the first application of the randomized study drug] + 1)

Total dose (mg) of acetaminophen will be calculated as follows:

Sum over all days from the first application of the randomized study drug to the last application of the randomized study drug of [Strength (mg) * Number of Tablets])

For these calculations the days without any consumption of acetaminophen will count as zero tablets.

These analyses will be performed on mITT, ITT and PP populations.

7.11.4 Subgroup analyses

The following subgroups will be considered:

1. Subjects with baseline Kellgren-Lawrence grade 1 or 2;
2. Subjects with baseline Kellgren-Lawrence grade 3.

The analyses of the primary endpoint described above will be repeated by subgroup. This will be an exploratory analysis and inference about equivalence and superiority of the study drug on the subgroups will not be drawn.

7.11.5 Forest Plot

A forest plot representing Test/Reference ration with 90% CI will be presented for all endpoints and subgroups for the PP, mITT and ITT popuplations:

1. Change from Baseline to Week 4 in WOMAC Pain Subscale Score:
    a. All subjects
    b. Subjects with Baseline WOMAC Pain Subscale Score <= 10
    c. Subjects with Baseline Kellgren-Lawrence Grade 1-2
    d. Subjects with Baseline Kellgren-Lawrence Grade 3
2. Change from Baseline to Week 2 in WOMAC Pain Subscale Score: for all subjects
3. Change from Baseline to Week 4 in POM VAS Score: for all subjects

Each CI will be represented by a horizontal line with a dot marking the ratio itself. Reference lines will be drawn to represent ratios of 100%, 80% and 125%

7.12 Safety Analyses

7.12.1 Adverse Events

Adverse Events will be coded using the Medical Dictionary of Regulatory Activities (MedDRA version 20.0) AE coding system for purposes of summarization.

An AE will be considered as treatment-emergent if the time of onset is after the time of the first randomized study drug administration or if it increased in severity during the study period. AEs with unknown start dates will be counted as treatment-emergent unless the AE resolution date is prior to the study drug start date. If the start date is partially missing, AE will be considered treatment-emergent unless the known part of the start date or the end date rules out that it occurred after the start of the study drug. See appendix 11.3 for more details on handing of partial dates.

A TEAE is defined as treatment-related if it is recorded as definitely, probably or possibly related to the study medication at the eCRF. In case the relatedness was not assessed, the most conservative result – related will be chosen for the analysis.

An overall summary will include, by treatment group and overall, the number and percentage of subjects reporting at least one TEAE and subjects discontinuing the study drug due to TEAE, as well as counts of AEs by severity, relationship, count of serious AEs and deaths.

A summary of the frequencies (number and percentage) of subjects with TEAEs and counts of TEAEs will be presented by system organ class and preferred term. Similar summaries will be created for serious TEAEs, TEAEs leading to treatment discontinuation and for TEAEs leading to death (if any).

Similar summaries of non treatment-emergent AEs and serious non treatment-emergent AEs by system organ class and preferred term will be created for the run-in population.

Additionally number and percentage of subjects with TEAE will be presented by preferred term only in the descending order of frequency for TEAEs experienced by >1% of subjects.

TEAEs will also be summarized by system organ class, preferred term and maximum severity as well as by system organ class, preferred term and closest relationship to the study drug (not related or related).

A subject experiencing the same AE multiple times will only be counted once for that preferred term. Similarly, if a subject experiences multiple AEs within the same system organ class that subject will be counted only once in that system organ class. In summaries by severity or by relationship, if a subject has the same AE on multiple occasions, the highest severity, and the closest relationship to study drug, respectively, will be used.

All information pertaining to adverse events noted during the study will be listed by subject, detailing verbatim, preferred term, system organ class, start date, stop date, severity, outcome, action taken and causal relationship to the study drug.

7.12.1.1 Application Site Reactions

Application site reactions are a subset of treatment-emergent adverse events. They will be identified by specific preferred terms. The list of preferred terms for application site reactions will be finalized prior to the database lock.

Number and percent of subjects and number of events will be summarized for application site reactions by preferred term and treatment. Application site reactions also will be listed.

7.12.2 Exposure to Product

The subjects will be instructed to use the diary to document all doses taken by checking the yes or no box for the appropriate date.

Study drug compliance will be defined as

[Actual number of applications] / [Planned number of applications] * 100%

where:

- Actual number of applications will be calculated from the subject diary as the number of gel application in the treatment period dosing log.
- Planned number of applications is 112 (28 days, 4 times daily).

Subjects will be considered treatment compliant if they apply at least 75% and not more than 125% of doses.

Duration of exposure will be defined as [Date of last study drug application] – [Date of first study drug application] + 1;

Compliance, actual number of applications and duration of exposure will be summarized descriptively by treatment group for the safety population.

Duration of exposure to the placebo treatment in the run-in period will also be summarized descriptively for the run-in population.

7.12.3 Vital Signs

Vital signs will be assessed at each visit. Systolic and diastolic blood pressure, respiratory rate, heart rate and body temperature will be measured.

Vital signs and their changes from baseline will be summarized descriptively at baseline and at each post-baseline visit (Visit 4 and Visit 5). Results also will be listed.

7.12.4 Physical Examination

Physical examination will be performed at Visits 1, 3, 4 and 5. The following body systems will be assessed: general appearance; skin; head, eyes, ears, nose, throat; lymph nodes; cardiovascular; abdominal; musculoskeletal; nervous; respiratory and other if applicable. Each body system will be classified as either normal or abnormal.

A summary of these findings will be presented by visit, body system and treatment group. Results also will be listed.

7.12.5 Laboratory Tests

Blood samples for safety laboratory tests (clinical chemistry and haematology) will be collected at Visit1 (Screening) and Visit 5. If the gap between the screening lab assessments and baseline visit (Visit 3) is more than 14 days, the lab assessments should be repeated at baseline visit.

The following tests will be performed:

- Hematology: Hemoglobin, Total RBC count, Total WBC count, Platelet count, Differential Leukocyte counts: Neutrophils, Lymphocytes, Eosinophils, Monocytes, Basophils.
- Blood Chemistry: Alkaline Phosphatase, Serum Creatinine, Random Glucose, SGPT (ALT) & SGOT (AST), Serum Bilirubin.

Values of laboratory tests and their changes from baseline will be summarized descriptively by visit, test and treatment group.

Shifts from baseline to Visit 5 among the categories Low (below the low limit of normal), Normal (within the normal limits) and High (above the upper limit of normal) will be presented. Only subjects with both baseline and Visit 5 assessments will be included

Results also will be listed.

7.12.6 ECG

ECG will be performed Visit1 (Screening) and Visit 5. Overall interpretation will be recorded as Normal, Abnormal not clinically significant or Abnormal clinically significant.

Number and percentage of subjects with each level of interpretation will be presented by treatment and visit. In addition, a summary of shifts from Visit 1 (Screening) to Visit 5 will be presented.

All ECG results will be listed.

8. CHANGES FROM PROTOCOL-SPECIFIED ANALYSES

There are no changes from the protocol-specified analyses.

9. LIST OF PLANNED TABLES, FIGURES, AND LISTINGS

See separate document with the table, figure and listing shells.

10. LITERATURE CITATIONS / REFERENCES

1. Study Protocol: "A Randomized, Double blind, Three-arm, Parallel, Placebo-controlled, Clinical Study to Evaluate the Bioequivalence using Clinical Endpoint of Diclofenac Sodium Gel, 1% (Mylan Inc.) to Voltaren® Gel (Diclofenac Sodium Topical Gel) 1% (Novartis Consumer Health, Inc.) in Patients with Osteoarthritis (OA) of the Knee " version 1.1 dated May 3, 2017.
2. Draft Guidance on Diclofenac Sodium Gel 1%. Office of Generic Drugs, FDA. Recommended March 2011.

11. APPENDICES

11.1 Study visit Schedule

Visit	Screening*	Start of Placebo Run-in Period*	Baseline Visit (Day 0)	Telephonic Contact[k]	Interim Visit (Day 14±4)	Telephonic Contact[k]	End of Study (Day 28±4)/Early Termination
Days	28 days prior to randomi-zation	At least 7 days (+4 days window)	Day 0	Midway between Visits 3 and 4	Day 14±4	Midway between Visits 4 and 5	Day 28±4
Visit No.	Visit 1	Visit 2	Visit 3		Visit 4		Visit 5
Informed Consent Process	X						
Demographics (e.g., height, weight)	X						
Inclusion and Exclusion Criteria	X	X	X				
Medical History	X						
Physical Examination[a]	X		X		X		X
12-Lead ECG[b]	X						X
Osteoarthritis of the knee according to ACR criteria	X						
X-Ray of Target Knee[c]	X						
Discontinuation of current osteoarthritis therapy							
Vital signs (HR, RR, BP, temperature)[d]		X					
Pregnancy test[e]	X	X[f]	X[f]		X[f]		X
Lab Safety Tests (Hematology & Clinical chemistry) [g]	X[h]						X
Pain on Movement (POM) assessment in both knees VAS		X	X		X		X
WOMAC® (version 3.1) pain score for the target knee[i]		X	X		X		X
Gel application and diary completion training		X	X				

Visit	Screening*	Start of Placebo Run-in Period*	Baseline Visit (Day 0)	Telephonic Contact[k]	Interim Visit (Day 14±4)	Telephonic Contact[k]	End of Study (Day 28±4)/Early Termination
Days	**28 days prior to randomi-zation**	**At least 7 days (+4 days window)**	**Day 0**	**Midway between Visits 3 and 4**	**Day 14±4**	**Midway between Visits 4 and 5**	**Day 28±4**
Visit No.	**Visit 1**	**Visit 2**	**Visit 3**		**Visit 4**		**Visit 5**
Subject diary dispensing		X (for run-in period)	X (for treatment period)				
Review of the subject diary			X		X		X
Collection of the subject diary			X				X
Study medication dispensing		X	X		X		
Collection of unused study medication			X		X		X
Rescue medication dispensing			X		X (if required)		
Collection of unused rescue medication							X
Assessment of treatment compliance			X	X	X	X	X
Application site evaluation			X		X		X
Adverse events[j]	X	X	X	X	X	X	X
Concomitant medication assessment		X	X	X	X	X	X

a A physical examination includes the evaluation of general appearance, eyes, head and neck, abdomen, lymph nodes, skin, cardiovascular system, respiratory system, and musculoskeletal system.
b A standard 12-lead ECG will be recorded after 5 minutes of rest in the supine position.
c Evidence of osteoarthritis with Kellgren-Lawrence grade 1-3 disease. X-ray of target knee will be performed on screening if not performed for more than 1 year prior to baseline.
d Vital signs will be measured by the investigator or designee after 5 minutes of rest (sitting).
e Pregnancy test is only required for females of childbearing potential. Serum pregnancy to be performed at screening and end of treatment visit.
f The urine pregnancy test can be performed if clinically indicated. If urine pregnancy is found to be positive, serum pregnancy will be done for confirmation.
g All hematology and clinical chemistry laboratory assessments are to be conducted preferably in a fasted state.
h The hematology and biochemistry (lab assessments) at screening to be performed within 14 days of baseline visit. If the gap between lab assessments and baseline visit is more than 14 days, the lab assessments should be repeated at baseline visit.
i WOMAC pain score (pain score = 0 to 20), will be determined by the subject's responses to five questions (S1–S5) using a 5-point Likert scale (i.e.,

‘none’=0; ‘mild’=1, ‘moderate’=2; ‘severe’=3; ‘extreme’=4). The questions pertain to the amount of pain the subject is currently experiencing in the target knee [i.e., ‘How much pain do you have’ when ‘Walking on a flat surface’ (S1), ‘Going up or down stairs’ (S2), ‘At night while in bed’ (S3), ‘Sitting or lying’ (S4), ‘Standing upright’ (S5)]

jAEs: The active reporting period to Mylan or its designated Representative begins from the time that the patient provides informed consent, which is obtained prior to the patient’s participation in the study, i.e., prior to undergoing any study-related procedure and/or receiving investigational product, through and including 28 days after the last application of the investigational product. All AEs should be followed until resolution or deemed stable or until the event is found to be due to another known cause (concurrent condition or medication) and clinical judgment indicates that further evaluation is not warranted, with Mylan concurring with that assessment. This may imply that observations will continue beyond the last planned visit per protocol, and that additional investigations may be requested by the Monitoring Team up to and as required by the sponsor. However, SAEs should be reported any time after the active reporting period, when the Investigator becomes aware and the SAE is considered to be reasonably related to the study drug. All SAEs should be immediately (24 hours) reported

kTelephone contact: Principal investigator or designee will contact the subjects one time between the schedule visit for assessment of the well-being, adverse event, concomitant medication, and treatment compliance. Additional telephonic contact will be done to remind the subject regarding stoppage of rescue medication usage prior to scheduled visit and next visit date. The same will be documented in the source notes.

* Based on the requirement for wash out period, screening and start of run-in can occur on the same day. The wash out period of 7 days or ≥ 5 half-life of previous osteoarthritis drug therapy, whichever is longer, will be maintained before the baseline visit.