

**A placebo-controlled, double-blind, randomized, trial of
Diclofenac Gel AMZ001 3.06% for the treatment of knee
osteoarthritis symptoms**

NTC03691844

Document Date: 05Aug2019

Statistical analysis plan

Final 3.0

A placebo-controlled, double-blind, randomized trial of Diclofenac Gel AMZ001 3.06 % for the treatment of knee osteoarthritis symptoms

Sponsor:	Amzell BV
Sponsor protocol number:	AMZ001-006
EudraCT number:	2018-001934-16
ClinicalTrials.gov number:	NCT03691844
Universal trial number (UTN):	Not applicable
Investigational product name:	Diclofenac Gel AMZ001 3.06%
Development phase:	2/3
Date of final SAP:	05AUG2019

Table of contents

Table of contents.....	3
List of abbreviations and definition of terms	4
1 Introduction	5
1.1 Study objectives and endpoints	5
1.1.1 Objectives.....	5
1.1.2 Endpoints	5
2 Study design.....	6
2.1 Overview of study procedures.....	7
2.2 Determination of sample size	9
2.3 Blinding.....	10
2.4 Randomization	10
3 Analysis sets	11
3.1 Intention-to-Treat Analysis Set (Full analysis set).....	11
3.2 Modified Intention-to-Treat analysis set	11
3.3 Per-protocol analysis set.....	11
3.4 Safety analysis set	12
4 Statistical analyses and presentation of data.....	12
4.1 General considerations	12
4.1.1 Data presentation.....	12
4.2 Efficacy	13
4.2.1 Primary efficacy analysis	13
4.2.2 Secondary efficacy analyses.....	14
4.2.3 Quality of life analyses.....	17
4.3 Safety.....	20
4.3.1 Adverse events	20
4.3.2 Other safety endpoints.....	21
5 Quality control.....	22
6 Layout of output.....	22
7 References	22

List of abbreviations and definition of terms

ADaM	Analysis data model
AE	Adverse event
BID	Twice daily
BMI	Body mass index
CDISC	Clinical Data Interchange Standards Consortium
CI	Confidence interval
eCRF	Electronic case report form
ECG	Electrocardiogram
EOT	End-of-text
EQ5D	EuroQol-5 Domain Questionnaire
ICOAP	Intermittent and Constant Osteoarthritis Pain
IMP	Investigational medicinal product
ITT	Intention-to-treat
LK	Likert
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed model repeated measures
mITT	Modified intention-to-treat
N/A	Not applicable
NRS	Numerical rating scale
OA	Osteoarthritis
OMERACT- OARSI	The Osteoarthritis Research Society International (OMERACT) Standing Committee for Clinical Trials Response Criteria Initiative and the Outcome Measures in Rheumatology (OARSI) committee
PGA	Patient Global Assessment
PP	Per-protocol
PT	Preferred term
Q1	First quartile
Q3	Third quartile
QD	Once daily
QoL	Quality of life
Randomized	Subject randomized to study treatment
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
Screened	Subject who enters the screening phase of the study
SD	Standard deviation
SDTM	Study data tabulation model
SE	Standard error
SOC	System organ class
TLF	Tables, listings and figures
VAS	Visual analog scale
WHO	World Health Organisation
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index
WPAI	Work and Productivity and Activity Index

1 Introduction

This document describes the planned statistical analyses and data presentations for study AMZ001-006 as outlined in protocol version 1.2 dated 22NOV2018.

1.1 Study objectives and endpoints

1.1.1 Objectives

The primary objective of this study is:

- To evaluate the change in pain intensity, in terms of the WOMAC pain score of the target knee

The secondary objectives of this study are:

- To evaluate changes in symptoms of OA
- To evaluate the changes in physical functioning
- To evaluate administration regimens of AMZ001
- To evaluate the safety and tolerability of AMZ001
- To evaluate changes in quality of life

1.1.2 Endpoints

The primary endpoint of this study (double-blind treatment groups only) is:

- The change from baseline in WOMAC pain sub-score (questions 1 to 5) in the target knee as evaluated at week 4.

The secondary endpoints of this study (double-blind treatment groups only) are:

- Changes from baseline in WOMAC total score and the WOMAC function and stiffness scores at week 4
- Changes from baseline in constant and intermittent OA pain assessed by ICOAP scores at week 4
- Changes from baseline in WOMAC pain weight-bearing score (questions 1, 2, and 5) and non-weight bearing score (questions 3 and 4) at week 4
- Changes from baseline in physical function assessed by the chair-stand test at week 4
- OMERACT-OARSI responder rate at week 4
- Total dose of rescue medication calculated as the sum of tablets used, based on pill counts
- Time between baseline and first use of rescue medication
- Changes from baseline in WOMAC pain sub-score (questions 1 to 5) between groups receiving AMZ001 QD and BID in the target knee as evaluated at week 4

- Changes from baseline in constant and intermittent OA pain assessed by ICOAP scores between groups receiving AMZ001 QD and BID at week 4
- Changes from baseline in WOMAC pain weight-bearing score (questions 1, 2, and 5) and non-weight bearing score (questions 3 and 4) between groups receiving AMZ001 QD and BID at week 4
- Changes from baseline in physical function assessed by the chair-stand test between AMZ001 QD and BID at week 4
- Changes from baseline in WOMAC total score and the WOMAC function and stiffness scores between AMZ001 QD and BID at week 4.
- Changes from baseline in the impact of OA on daily living as assessed by the PGA score at week 4
- Changes from baseline in work productivity and activity assessed by the WPAI at week 4
- Changes from baseline in quality of life assessed by the EQ5D at week 4

The safety endpoints of this study are:

- Nature, incidence and severity of AEs
- Changes in laboratory safety parameters, vital signs, 12-lead ECG parameters and weight
- Nature, incidence and severity of skin reactions on the application site

2 Study design

The trial is a multi-center, double-blind, randomized, placebo-controlled trial of AMZ001 for the treatment of knee osteoarthritis symptoms. The trial also includes a single-blind component, consisting of Voltaren[®] 1 % gel. The purpose of the trial is to evaluate the efficacy, safety and tolerability of once or twice daily application of AMZ001 during a 28-day period in subjects with radiographic and symptomatic knee osteoarthritis in either one or both knees.

A total of approximately 440 subjects are planned to be randomized. The subjects will be randomized in a ratio of 3:3:3:2 with approximately 120 subjects in each of the three double-blinded treatment groups (AMZ001 BID, AMZ001 QD and placebo) and approximately 80 subjects in the single-blinded treatment group (comparator Voltaren[®] gel 1 %).

The trial consists of three phases: screening, treatment, and follow-up, as illustrated in Table 1.

2.1 Overview of study procedures

Table 1 Overview of study procedures

Activity/Assessment	Screening 1a ¹	Screening 1b ¹	Baseline V2	On Treatment Visits				14-Day Safety Follow-up ² (± 5 days) V7
				V3	V4	V5	V6	
Study Week	-	-	0	1	2	3	4	6
Study Day	-21 to - 5	-13 to -1	1	8	15	22	29	43
Visit Window (days)	-	-	0	+/- 2	+/- 2	+/- 2	+/- 2	+/- 5
Informed Consent	X							
Medical history	X							
Inclusion/exclusion criteria	X	X	X					
Demographic data	X							
Physical examination	X						X	
Skin Tolerability Assessment			X ³	X	X	X	X	
Knee X-ray (both knees)	X							
Instructions on reporting pain ⁴	X							
WOMAC Questionnaire ^{5,6,7}		X	X ⁵	X	X	X	X	
ICOAP Questionnaire ⁶			X	X	X	X	X	
Pain Diary ⁸			X ⁸	X ⁸	X ⁸	X ⁸	X ⁸	
Patient Global Assessment			X	X	X	X	X	
WPAI			X	X	X		X	
EQ5D			X			X	X	
Satisfaction questionnaire					X		X	
Chair-stand test ⁶			X		X		X	
Selection of target knee			X					

Activity/Assessment	Screening 1a ¹	Screening 1b ¹	Baseline V2	On Treatment Visits				14-Day Safety Follow-up ² (± 5 days) V7
				V3	V4	V5	V6	
Study Week	-	-	0	1	2	3	4	6
Study Day	-21 to -5	-13 to -1	1	8	15	22	29	43
Height and weight	X						X ⁹	
Blood pressure and heart rate	X		X	X	X	X	X	
Adverse events	X ¹⁰	X	X	X	X	X	X	X
Concomitant medication	X	X	X	X	X	X	X	X
Randomization			X					
Study drug dispensation and weighing of kit(s)			X		X ¹⁵			
Dispensation of IMP "instructions for use" pamphlet			X		X			
Instruction of study drug application and frequency of use			X					
Application of morning study drug on site under staff supervision			X	X	X	X	X	
Weighing of kit(s) for compliance assessment				X	X	X	X	
Rescue medication dispense			X		(X) ¹¹	(X) ¹¹		
Rescue medication collection ¹²					(X) ¹¹	(X) ¹¹	X	
12 lead ECG	X						X	
(Rescue) Analgesic Wash-out		X	X	X	X	X	X	
Hematology	X			X	X	X	X	
Safety chemistry and urine dipstick	X			X	X	X	X	
Pregnancy test ¹³	X		X		X		X	
Blood sample for plasma level of diclofenac ¹⁴					X		X	

EQ5D: EuroQol-5 Domain. ICOAP: Intermittent and Constant Osteoarthritis Pain. WOMAC: The Western Ontario and McMaster Universities Osteoarthritis Index. WPAI: Work Productivity Activity Index.

1. If subject is washed out of analgesics at Screening Visit 1a, the examinations scheduled for screening visits 1a and 1b can be performed on the same day.
2. Phone call.
3. Skin assessment to be performed prior to application of study drug at baseline visit.
4. Subjects will be educated on appropriate expectations around their participation in a clinical study and the importance of reliably consistently and accurately reporting their pain throughout the study.
5. WOMAC pain sub-score (5 questions) to be assessed on a daily basis on paper at home during first week of treatment, i.e. the period between Visit 2 and Visit 3.
6. Assessed without analgesic medication for five half-lives of the analgesic.
7. WOMAC assessment of hips should only be performed at the screening visit.

8. To be completed daily in the evening during the first week of treatment, i.e. between Visit 2 and Visit 3). In the period between other subsequent visits, the questionnaire will be completed during the evening of the day prior to the next visit.
9. Only weight collected.
10. Conditions previously unknown to the subject at the time of the screening visit will be recorded as medical history. All subsequent events will be recorded as adverse events.
11. To be collected/dispensed if previously dispensed supply of rescue medication was depleted and empty packaging returned to site.
12. Collection of remaining rescue medication at Visit 6. Empty packaging of previously dispensed kit should be returned to the site if new rescue medication is requested.
13. Serum test at screening visit, urine dipstick at all other visits.
14. To be done after IMP application and questionnaire/test completion (approximately 1-2 hours after IMP application). Exact time of IMP application and time of blood sampling must be documented.
15. Dispensation is not mandatory for subjects allocated to Voltaren® gel

2.2 Determination of sample size

The power is determined from expected changes in the primary endpoint of the change from baseline in pain as evaluated by the WOMAC pain sub-score. The power calculation is based on data from a 4-week double-blind diclofenac sodium 2 % topical solution treatment study in knee osteoarthritis patients (Wadsworth et al. 2016). In this study using the WOMAC LK (Likert) 3.1 scale (0-4 for each question), a decrease from baseline in WOMAC pain sub-score (5 questions) was recorded in the diclofenac treatment group of 4.5 (SD 4.5) and in the placebo group 3.6 (SD 4.2).

The estimated common SD of 4.4 in the LK 3.1 scale can be converted to a SD of $4.4 * 2.5 = 11$ in the NRS scale. Calculations based on a series of assumptions are shown below in Table 2. With the assumptions of the SD of 11 (NRS scale), 120 subjects enrolled in each of the double-blinded treatment arms, 10 % dropout rate, normal distribution, 5 % level of significance two-sided, 80 % power, the study will be powered to detect a treatment difference in WOMAC pain sub-score between AMZ001 BID and placebo of 4.2 on the NRS scale.

Table 2 Power calculations for pairwise difference in WOMAC pain sub-score between AMZ001 and placebo with 108 evaluable subjects per group, 5 % level of significance, two-sided.

WOMAC pain-subscore difference	Common SD	Power
3.8	9	0.87
	11	0.72
	13	0.57
4.0	9	0.90
	11	0.76
	13	0.61
4.2	9	0.93
	11	0.80
	13	0.66

The power is not adjusted for multiple comparisons and adjustment for multiplicity will not be made as there is one test only performed for the primary endpoint.

It should be noted that the current trial is not designed or powered to evaluate superiority of AMZ001 compared to Voltaren[®] gel, but the trial will provide important information on onset and duration of action of AMZ001 to guide future, larger trials. Also, the trial is not powered to evaluate superiority of AMZ001 applied twice daily as compared to once daily, but the trial will provide information on efficacy, safety, and tolerability of twice daily administration as compared to once daily administration.

Furthermore, the trial will be used to finalize the psychometric validation of the Pain Diary and the Satisfaction Questionnaire.

2.3 Blinding

This is a double-blind study of AMZ001 (QD and BID regimens) versus placebo, which also includes a single-blind active comparator (Voltaren[®] gel 1 %).

Due to lack of availability of Voltaren[®] gel 1 % in metered dose-dispensers identical to that of the double-blinded IMP, as well as due to differences in approved daily dose application schedules, full masking of Voltaren[®] gel 1 % was not considered feasible for this trial. Instead, the Voltaren[®] gel 1 % is single-blinded to conceal the brand of the product. The single-blinded product is labelled identically to other IMPs in the trial. All double-blinded IMPs will be indistinguishable in appearance of the container, the label, as well as the gel appearance, smell and texture. The appearance of the single-blinded Voltaren[®] gel tube will be masked to fully cover any branding information.

Neither subject, investigator nor staff working on the study will be aware of treatment allocation to the double-blind treatment groups. The dispensation type (metered dose-dispenser vs. gel tube) and required frequency of application risks unblinding of the allocation of Voltaren[®] gel for the subjects allocated to the single-blinded treatment groups. Due to the difference in primary packaging material, study staff will be indirectly unblinded to allocations to the Voltaren[®] gel treatment group.

As the primary efficacy data analyses will be performed on the double-blinded treatment groups, the risk of bias associated with single-blinded treatment groups is considered irrelevant to the main data output.

All summaries produced before database lock will be presented using dummy treatments and this will be explicitly stated.

2.4 Randomization

The screening/randomization procedure will be centrally managed through an electronic Interactive Web Responding System integrated in the eCRF.

Eligible subjects will be randomized to one of four treatment groups with an allocation ratio of 3:3:3:2. Each of the three double-blinded treatment groups (AMZ001 BID, AMZ001 QD and placebo) will be randomized with approximately 120 subjects for the primary objective of assessment of an analgesic effect of topical AMZ001 gel compared with placebo. The single-blinded group of the comparator of Voltaren[®] gel 1 % will be randomized with approximately 80 subjects.

Randomization will be stratified by country to ensure balance of treatment groups with respect to intrinsic and extrinsic factors that may affect the OA outcomes.

3 Analysis sets

Every subject will be classified according to the below definitions of the Intention-to-Treat (ITT), Modified Intention-to-Treat (mITT), Per-Protocol (PP) and Safety (SAF) analysis sets at the Blind Data Review Meeting before breaking the blind. The decisions will be made by the trial team and documented in the minutes of the meeting.

Outputs on demographics and baseline characteristics, medical history, prior medication and concomitant medication will be performed using the ITT, mITT, PP and SAF analysis sets. Subject disposition will be summarized for all screened subjects and protocol deviations will be summarized for the ITT analysis set. Exposure and compliance will be performed using the ITT and SAF analysis sets.

3.1 Intention-to-Treat Analysis Set (Full analysis set)

The ITT analysis set will include all subjects randomly allocated to a treatment, based on the intention to treat “as randomized” principle (i.e. the planned regimen rather than the actual treatment given in case of any difference).

3.2 Modified Intention-to-Treat analysis set

The mITT analysis set will include all subjects from the ITT analysis set who have a baseline and at least one post-treatment WOMAC pain sub-score assessment available.

The mITT analysis set will be used to perform all efficacy analyses and summaries. Subjects will be analysed according to the randomized treatment.

3.3 Per-protocol analysis set

The PP analysis set will include all subjects from the mITT analysis set who have been treated according to the trial protocol and fulfil the following criteria:

- Absence of major clinical trial protocol deviations with respect to factors likely to affect the efficacy of the treatment, where the nature of such clinical trial deviations will be defined before breaking the blind. Major protocol violations could include although not limited to major deviations of inclusion/exclusion criteria and use of prohibited medication.
- Adequate compliance with trial medication; defined as overall compliance of the target knee

between 75% and 125%.

The PP analysis set will be used to perform sensitivity analyses for the primary endpoint.

3.4 Safety analysis set

The SAF will include all subjects who have taken at least one application of trial treatment. Subjects will be analysed according to the actual treatment they receive.

The actual treatment received by each subject will be established by the unblinded data-manager. At the Blind Data Review Meeting, the unblinded data-manager will present a **subject blinded** list of treatment misallocations for review. For subjects misallocated treatment, the actual treatment received will be decided based on below categories and documented. For subjects receiving treatment as per the randomization, actual treatment is identical to planned treatment.

Since dispensation of IMP occurs at multiple timepoints (Visit 2, Visit 4 and as needed), subjects will fall into one of the following three actual treatment categories:

1. IMP is dispensed correctly on all occasions.
2. IMP is consistently dispensed incorrectly on all occasions i.e. the subject received the same (incorrect) treatment throughout the study and the IMP received corresponds to one of the study treatment groups.
3. IMP is dispensed incorrectly on one or more occasions and the subject did not consistently receive IMP corresponding to one of the study treatment groups.

In order not to bias the safety reporting of the study treatment groups, any subjects falling into category 3 above will be included in the SAF under a separate treatment group 'AMZ – misallocations' and will therefore not be included in the safety summaries for any of the study treatment groups.

The SAF will be used for all safety and tolerability summaries.

4 Statistical analyses and presentation of data

4.1 General considerations

Descriptive summaries and analyses when applicable will be presented for the following treatment groups: AMZ001 BID, AMZ001 QD, Placebo and Voltaren 1%.

4.1.1 Data presentation

All statistical tests will be performed using a two-sided test at a 5% significance level. Results from analyses will be presented with estimates, 95% confidence intervals (CIs) and p-values.

Numerical data will be presented in summary tables by number of observations (n), the number of missing values (missing), arithmetic mean, standard deviation (SD), median, first and third quartiles (Q1, Q3), minimum (min) and maximum (max) values. The mean, median, Q1 and Q3 will be reported to one decimal place more than the raw data, the standard deviation to two decimal places more than the raw data and the minimum and maximum values to the same precision as the raw data. The summary statistics presented for categorical data will be the

number of observations (n), the number of missing values (missing) and the count and percentage of subjects in each category. Number of events will also be reported where applicable.

All data will be listed by treatment group, subject and time point (if applicable). Age and sex will be displayed in all listings.

4.2 Efficacy

4.2.1 Primary efficacy analysis

The primary efficacy analysis will include data from the double-blind treatment groups only.

4.2.1.1 Derivation of primary endpoint

The primary endpoint is the change from baseline at week 4 in the WOMAC pain sub-score in the target knee.

The WOMAC pain sub-score is derived from WOMAC Osteoarthritis Index questionnaire as the sum of questions 1 to 5. It takes values in the range 0-50, where higher values indicate greater pain. For convenience and ease of interpretation, all WOMAC scores (including the pain sub-score) will be normalised to a 0-100-point scale for data analysis.

For the primary endpoint,

$$\text{normalised WOMAC pain sub-score} = (\text{WOMAC pain sub-score}/50) * 100$$

The full WOMAC questionnaire is assessed for the target knee at baseline (Visit 2), week 1 (Visit 3) week 2 (Visit 4), week 3 (Visit 5) and week 4 (Visit 6).

4.2.1.2 Summary presentations and method of analysis

The absolute values and the absolute change from baseline in the WOMAC pain sub-score will be summarized over time by treatment group.

The primary analysis will use a restricted maximum likelihood based repeated measures mixed model (MMRM) on the dependent variable absolute change from baseline of the WOMAC pain sub-score. The analysis will include covariates of baseline value (WOMAC pain sub-score), treatment, timepoint, sex, country, the subject characteristic of unilateral/bilateral knee OA at baseline (defined by whether one or both knees meet the requirements for target knee selection) and treatment-by-timepoint interaction. An AR(1) covariance structure will be used to model the correlations between within-subject repeated measurements. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom and adjust standard errors.

Sample SAS code to be used for the MMRM analysis:

```
proc mixed data=<input data>;  
  class treatment visit sex status subject;  
  model response = treatment visit treatment*visit sex country status baseline / ddfm=kr s;  
  repeated visit / subject=subject type=ar1;  
  lsmeans visit*treatment / diffs cl obsmargins;
```

run;

where *response* is the change from baseline in WOMAC pain sub-score and *baseline* is the WOMAC pain sub-score at baseline, *status* is the categorical variable of either unilateral or bilateral knee OA at baseline and *visit* is the study timepoint.

Least square mean estimates of change from baseline in WOMAC pain sub-score along with 95% confidence intervals (CI) will be presented at each time-point for each treatment group.

Estimated treatment differences at week 4 between all treatment groups along with associated 95% confidence intervals and p-values will be presented. P-values will be interpreted according to the hierarchical step-down testing procedure specified for the primary objective and also with consideration as to whether they are secondary or exploratory objectives. Model-estimated least square means and associated 95% CI of change from baseline in WOMAC pain sub-score will be plotted over time by treatment group.

The primary efficacy analysis will be based on the mITT. The robustness of the results of the primary analysis will be investigated by a sensitivity analysis based on the PP.

Further investigation of the robustness of the primary results (in particular, the influence of time and covariates) may be performed, if deemed necessary.

4.2.2 Secondary efficacy analyses

The secondary efficacy analyses will include data from the double-blind treatment groups only.

4.2.2.1 Secondary endpoints

WOMAC Osteoarthritis Index

The WOMAC Osteoarthritis Index consists of 24 questions about aspects of the target knee over the previous 24 hours. The questions are split into three sections: *Pain*, *Stiffness* and *Difficulty Performing Daily Activities* (i.e. Physical Function). Each question is scored on a 0-10 11-point numerical scale, where higher numbers indicate greater pain, stiffness or difficulty in performing daily activities.

The following scores are derived from answers to the WOMAC questionnaire:

WOMAC scores	Range (min – max)	Derivation
Total score	0 - 240	Q1 + Q2 +..... + Q23 + Q24
Pain sub-score	0 - 50	Q1 + Q2 + Q3 + Q4 + Q5
Pain weight bearing sub-score	0 - 30	Q1 + Q2 + Q5
Pain non-weight bearing sub-score	0 - 20	Q3 + Q4
Stiffness sub-score	0 - 20	Q6 + Q7
Function sub-score	0 - 170	Q8 + Q9 +..... + Q23 + Q24

For convenience and ease of interpretation, all WOMAC derived scores will be normalised to a 0-100-point scale for data analysis where, for each score

$$\text{normalised score} = (\text{score}/\text{max possible value}) * 100$$

The full WOMAC questionnaire is assessed for the target knee at baseline (Visit 2), week 1 (Visit 3) and weeks 2, 3 and 4.

The WOMAC pain sub-scale is also completed at screening for both the left/right hips and the left/right knees. This is for the assessment of baseline status and eligibility, the determination of the target knee and whether the subject has bilateral or unilateral knee OA.

Patient Global Assessment (PGA) questionnaire

The PGA consists of a single question asking how a patient's knee OA has affected their health during the last 48 hours. It is scored on an 11-point numerical rating scale from 0 to 10, where higher scores represent a higher level of disease activity or worse health.

The PGA is assessed at baseline and weeks 1, 2, 3 and 4.

Intermittent and Constant OA Pain (ICOAP)

The ICOAP consists of 11 items (questions) assessing pain in the target knee over the previous week. Each question is scored on a 5-point scale as follows, where higher scores indicate more severe pain:

- 0 = no pain
- 1 = mildly
- 2 = moderately
- 3 = severely
- 4 = extremely

The following scores are derived from answers to the ICOAP questionnaire:

ICOAP scores	Range	Derivation
Total	0 - 44	Q1 + Q2 + + Q10 + Q11
Constant pain	0 - 20	Q1 + Q2 + Q3 + Q4 + Q5
Intermittent pain	0 - 24	Q6 + Q7 + Q8 + Q9 + Q10 + Q11

All ICOAP scores will be normalised to a score out of 100 using the formulae:

$$\text{Normalized total pain score} = (\text{total pain score}/44) * 100$$

$$\text{Normalized constant pain sub-score} = (\text{constant pain sub-score}/20) * 100$$

$$\text{Normalized intermittent pain sub-score} = (\text{intermittent pain sub-score}/24) * 100$$

The ICOAP is assessed at baseline and weeks 1, 2, 3 and 4.

Chair-stand test

The chair-stand test is the number of times a subject can stand up and then sit down again in 30 seconds without using their arms. Higher numbers correspond to greater physical function.

The chair-stand test is performed at baseline, week 2 and week 4.

Rescue Medication (paracetamol/acetaminophen)

The total dose of rescue medication (g) taken will be determined as

$$\text{total \# of tablets of rescue medication used} \times 0.5$$

where,

$$\text{total \# of tablets used} = (\text{\#of tablets used between V2 \& V4}) + (\text{\#of tablets used between V4 \& V6})$$

and

$$\text{\# of tablets used between } V_i \text{ and } V_j = (\text{\# tablets dispensed at } V_i) - (\text{\# tablets returned at } V_j) - (\text{\# tablets unaccounted for at } V_j)$$

The average dose of rescue medication per day while on study drug (g/day) will be determined as

$$\frac{\text{total dose of rescue medication used (g)}}{\text{number of days on study drug}}$$

where,

$$\text{number of days on study drug} = \begin{cases} \text{duration of exposure,} & \text{if trt end date} < \text{V6 account date} \\ \text{duration of exposure} - 1, & \text{if trt end date} = \text{V6 account date} \end{cases}$$

Note: a consistent approach to determining ‘number of days on study drug’ is taken across exposure, compliance and rescue medication calculations.

The time between baseline and first intake of rescue medication (days) will also be determined as

$$\text{Date of first intake of rescue medication} - \text{treatment start date}$$

For subjects that have first intake of rescue medication on the Visit 2 (treatment start) date, the time to first intake of rescue medication will be set to 0.5 days.

4.2.2.2 Summary presentations and methods of analysis

All summaries and statistical analysis for the secondary endpoints will be performed for the mITT set only.

For the following continuous secondary endpoints, absolute values and absolute change from baseline will be summarised over time by treatment group. Statistical analysis and hypothesis testing will be conducted as for the primary efficacy analysis using a mixed effects repeated measures model with data from double-blind treatment groups only.

- WOMAC total score
- WOMAC function score
- WOMAC stiffness score
- WOMAC pain weight bearing score
- WOMAC pain non-weight bearing score

- Patient Global Assessment (PGA)
- ICOAP total OA pain score
- ICOAP constant OA pain score
- ICOAP intermittent OA pain score
- Chair stand test, number of repetitions

The number of subjects using any rescue medication (i.e. at least one tablet), the total dose of rescue medication (g), and the average dose of rescue medication per day on study drug (g/day) will be summarized by treatment group. Time to first use of rescue medication will be summarised by unadjusted Kaplan Meier estimates of median time to first use of rescue medication with 95% CI. For subjects not taking any rescue medication, the total dose of rescue medication will be set to 0 mg and time to first use of rescue medication will be censored at the Visit 6/EOT visit date or the last date of subject participation in the study if this is not available.

The average dose of rescue medication per day on study drug will analysed using an analysis of covariance model with treatment, sex, baseline target knee WOMAC pain sub-score, country and subject characteristic of baseline unilateral/bilateral knee OA as explanatory variables. Least square mean estimates of each dependent variable along with 95% confidence intervals (CI) will be presented for each treatment group. Estimated treatment differences between all treatment groups along with associated 95% confidence intervals and p-values will be presented.

The time between baseline and first use of rescue medication will be analysed using a Cox semi-parametric proportional hazards model with treatment, sex, baseline target knee WOMAC pain sub-score, country and subject characteristic of baseline unilateral/bilateral knee OA as covariates. The relative hazards and associated confidence intervals for all pairwise comparisons will be estimated from the Cox model and displayed using Kaplan-Meier curves. Frequencies and rates will be presented by treatment group. Unadjusted Kaplan Meier estimates of median 'survival' time and 95% CI will be presented by treatment group, as stated above.

If the assumptions underlying any of the pre-specified analyses are not considered appropriate for any of the endpoints once data is available, then alternate analyses may be presented.

4.2.3 Quality of life analyses

The quality of life analyses will include data from the double-blind treatment groups only.

4.2.3.1 QoL endpoints

Work Productivity and Activity Impairment (WPAI) Questionnaire: General Health

The WPAI General Health questionnaire consists of six questions. Responses to the questions are used to derive WPAI outcomes, which are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity, i.e. worse outcomes.

The WPAI questions and derived endpoints are as follows:

WPAI Questions:

- 1 = currently employed (YES/NO)
- 2 = hours missed due to health problems
- 3 = hours missed other reasons
- 4 = hours actually worked
- 5 = degree health affected productivity while working (0-10 11-point NRS)
- 6 = degree health affected regular activities (0-10 11-point NRS)

WPAI derived endpoints	Derivation
Percent work time missed due to health	$Q2/(Q2+Q4)$
Percent impairment while working due to health	$Q5/10$
Percent overall work impairment due to health	$Q2/(Q2+Q4) + [(1-(Q2/(Q2+Q4))) \times (Q5/10)]$
Percent activity impairment due to health	$Q6/10$

Derived scores are multiplied by 100 to express them in percentages.

If any of the WPAI questions required to calculate a given WPAI outcome is missing, no data imputation will be performed and the WPAI outcome will be set to missing.

A note-to-file issued on 04JUN19 documents the strategy to be followed in cases where the reported total number of working hours (the sum of hours in WPAI questions 2, 3 and 4) are considered outside the expected normal range (0-70 hours/week) and thus are believed to be based on a misunderstanding of questions 2, 3 and 4. To exclude the invalid values, an algorithm will be applied to objectively select the participants for whom the more extreme values have been reported. The algorithm will **exclude the data from questions 2, 3 and 4 for all visits** for subjects that meet at least one of the following criteria:

- Report more than 70h work week at any time during the trial OR
- Report more than 60h work week at any time during the trial AND
 - Increase of more than 80% in total work hours per week from one or more previous visits, with a defined minimum reference value of 25 hours per week
 - Decrease of more than 50% in total work hours per week from one or more previous visits, with a defined minimum reference value of 25 hours per week.

The algorithm will be applied prior to unblinding, to ensure unbiased analysis of the valid data only. The data from the WPAI questionnaire not affected by the total number of hours reported (i.e. questions 1, 5 - if applicable - and 6) will not be excluded from analysis.

The WPAI is assessed at baseline and weeks 1, 2 and 4.

EQ5D

The EQ-5D is a standardized measure of health status that consists of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) each measured on 5 levels (no problems, slight problems, moderate problems, severe problems and extreme problems) and a

VAS of the subject's self-rated health. The VAS is scored from 0-100 where 0 corresponds to 'the worst health you can imagine' and 100 to the 'the best health you can imagine'.

The EQ5D is assessed at baseline, week 3 and week 4.

OMERACT-OARSI Responder Criteria

The OMERACT-OARSI responder criteria (Pham, et al., 2004) are based on the WOMAC pain and function sub-scores and the Patient Global Assessment (PGA) questionnaire.

A subject is defined as a responder if they demonstrate either

- a) High improvement in at least 1 of the WOMAC pain or function sub-scores
or
- b) Moderate improvement in at least 2 of the WOMAC pain sub-score, WOMAC function sub-score or PGA.

where *high improvement* is defined as

- Relative improvement of $\geq 50\%$ (i.e. relative change $\leq -50\%$) from baseline AND
- Absolute improvement of ≥ 20 (i.e. absolute change ≤ -20) from baseline

and *moderate improvement* is defined as

- Relative improvement of $\geq 20\%$ (i.e. relative change $\leq -20\%$) from baseline AND
- Absolute improvement of ≥ 10 (i.e. absolute change ≤ -10) from baseline.

and

relative change = percentage change during the study

$$= ((\text{final score} - \text{baseline score}) / \text{baseline score}) * 100$$

absolute change = absolute change during the study

$$= \text{final score (normalised)} - \text{baseline score (normalised)}$$

In the situation that at least one of the three endpoints (i.e. WOMAC pain and function sub-scores and PGA) required to assess the responder criteria is missing:

- If the available endpoints satisfy the responder criteria, then the subject will be classified as a responder. (e.g. If a subject has high improvement on the WOMAC pain sub-score but the WOMAC function sub-score and PGA are missing then the subject is a responder).
- If the available endpoints do not satisfy the responder criteria AND there is no possibility that they could have been a responder when considering possible values for the missing score(s), then the response will be set to non-response. (e.g. If a subject does not achieve moderate response on either the WOMAC pain sub-score or the PGA and the WOMAC function sub-score missing then the subject is a non-responder).
- If the available endpoints do NOT satisfy the responder criteria BUT the subject could have been a responder depending on the possible values for the missing score(s), then the

response will be set to missing (e.g. If either of the WOMAN pain or function sub-scores are missing then the response rate will be set to missing).

OMERACT-OARSI responder criteria will be evaluated at weeks 1, 2, 3 and 4.

4.2.3.2 Summary presentations and methods of analysis

Summaries and analysis will be performed for the mITT set only.

For the following QoL endpoints, absolute values and absolute change from baseline will be summarized over time by treatment group. Statistical analysis will be conducted as for the primary efficacy analysis using a mixed effects repeated measures model.

- WPAI % time missed
- WPAI % impairment while working
- WPAI % overall work impairment
- WPAI % activity impairment
- EQ5D VAS

4.3 Safety

Safety parameters will be evaluated for the safety analysis data set.

4.3.1 Adverse events

Adverse events (AEs) will be classified according to MedDRA version 21.0.

An AE overview summary table will be prepared including the number of subjects reporting an AE, the percentage of subjects (%) with an AE, and the number of events (E) reported for the following categories:

- Treatment-emergent AEs
- Deaths
- Serious adverse events (SAEs)
- AEs leading to withdrawal of study drug
- Severe and life-threatening AEs
- Adverse drug reactions

Treatment emergent AEs are defined as any AE that has occurred after the first administration of IMP administration.

ADRs are defined as all untoward and unintended responses to an IMP related to any dose administered, i.e. all AEs which are probably or possibly related to IMP.

Treatment-emergent AEs will be summarized in a table by dictionary level, i.e., system organ class (SOC) and preferred term (PT) for MedDRA. The table will display the total number of subjects reporting an AE, the percentage of subjects (%) with an AE and the number of events (E)

reported. AEs will be presented by SOC sorted alphabetically and PT sorted in decreasing frequency of occurrence.

Pre-treatment AEs will be listed only.

Summary tables will be prepared for:

- All TEAEs
- All non-serious TEAEs
- TEAEs by causality (related/unrelated)
- TEAEs by severity
- TEAEs leading to death
- TEAEs leading to withdrawal of study drug
- Treatment emergent SAEs

Data listings will be provided for:

- All AEs sorted by centre and subject number
- All AEs sorted by MedDRA SOC and PT
- SAEs
- AEs leading to death
- AEs leading to withdrawal of study drug
- Pre-treatment AEs

4.3.2 Other safety endpoints

4.3.2.1 Laboratory measurements

Absolute values and change from baseline in hematology and clinical biochemistry parameters will be summarized by visit and treatment group using descriptive statistics.

Frequency and percentage for categorical urinalysis data will be presented by visit and treatment group, when relevant.

Laboratory values will be flagged if outside the reference range.

A listing of abnormal values will be presented.

4.3.2.2 Vital signs and ECG

Descriptive statistics for vital signs parameters (systolic and diastolic blood pressure, heart rate, and body temperature), weight and 12-lead ECG parameters along with ECG interpretation will be presented in the same way as the laboratory parameters.

4.3.2.3 Skin tolerability assessment

Frequency and percentage for categorical skin tolerability assessment data will be presented by visit and treatment group.

5 Quality control

Programs used to derive the subject-level ADaM analysis dataset and the efficacy endpoints from the SDTM domains will be reviewed through parallel programming by an independent programmer. Other ADaM datasets will be reviewed by code review. Furthermore, the programs involving the statistical analyses of primary and secondary efficacy endpoints will be reviewed using parallel programming by an independent statistician. For the review of outputs, it is not necessary that the formats or layouts match.

Other output programs will be reviewed by the method of code review. All quality control activities for individual programs will be carried out in compliance with SOP 703, Programming of Single Use SAS Programs. All review findings and their follow-up will be documented in a Program Overview Form.

6 Layout of output

An ICH E3 based numbering will apply for the EOT documents. RTF files in landscape orientation will be used as the format for single output files with font type as Times New Roman of size 9. PNG files will be used for graphics, and Microsoft Office Word 2010 (or later) files (DOCX) files will be used as the format for TLF collections (EOT documents).

All output will be produced using SAS version 9.4 or a later version.

The sponsor name, protocol number, SAS program and output name and run date will appear bottom right in a footnote. The header of the EOT document will include date of collection and status (draft/final).

On an agreed timeline before the Database Release Meeting, a draft version of the EOT documents using dummy treatment groups will be produced and distributed for review. An Output Review Meeting will be held to consolidate the review findings and prepare a list of revision requests to be used for the second draft of the EOT documents. The second delivery of EOT documents (with dummy treatment groups) will have been subject to full QC.

7 References

- Measure of Intermittent and Constant Osteoarthritis Pain: ICOAP User's Guide, Version 6: July 7 2010.*
- Pham, T., Heijde, D. v., Altman, Anderson, Bellamy, Hochberg, . . . Dougados, W. a. (2004). OMERACT-OARSI Initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. *OsteoArthritis and Cartilage*, 389-399.
- Wadsworth, L., Kent, J., & Holt, R. (2016). Efficacy and safety of diclofenac sodium 2% topical solution for osteoarthritis of the knee: a randomized, double-blind, vehicle-controlled, 4 week study. *Current Medical Research and Opinion*, 32(2) 241-250.