A RANDOMIZED, CONTROLLED, DOUBLE-BLIND STUDY TO EVALUATE THE EFFICACY AND SAFETY OF AN INTRA-ARTICULAR INJECTION OF AMPION™ IN ADULTS WITH PAIN DUE TO SEVERE OSTEOARTHRITIS OF THE KNEE

STATISTICAL ANALYSIS PLAN

STUDY NUMBER: AP-013

NCT 03988023

7 JUNE 2019

STATISTICAL ANALYSIS PLAN APPROVAL SHEET

Protocol Number: AP-013

Product: Ampion™

Protocol Title: **A RANDOMIZED, CONTROLLED, DOUBLE-BLIND STUDY TO EVALUATE THE EFFICACY AND SAFETY OF AN INTRA-ARTICULAR INJECTION OF AMPION™ IN ADULTS WITH PAIN DUE TO SEVERE OSTEOARTHRITIS OF THE KNEE**

Author: Gary Stevens, PhD

Version: 1.0

Version date: 7 June 2019

Approved by:

Lorena DeRenzo 13 JUN 2019
Name Date

Holli Cherevka 13 Jun 2019
Name Date

GARY STEVENS 13 JUN 2019
Name Date

STATISTICAL ANALYSIS PLAN

A RANDOMIZED, CONTROLLED, DOUBLE-BLIND STUDY TO EVALUATE THE EFFICACY AND SAFETY OF AN INTRA-ARTICULAR INJECTION OF AMPION™ IN ADULTS WITH PAIN DUE TO SEVERE OSTEOARTHRITIS OF THE KNEE

Protocol Number AP-013

Version:	1.0
Date:	7 June 2019
Biostatistician:	Gary Stevens, PhD

Table of Contents

1. INTRODUCTION

This statistical analysis plan (SAP) outlines the proposed statistical methods to be implemented during the review of data to ensure that it confirms with categories determined by the CRF or the anticipated ranges for continuous variables and analysis of data collected within the scope of Ampion Protocol AP-013, "A Randomized, Controlled, Double-Blind Study To Evaluate The Efficacy And Safety Of An Intra-Articular Injection Of Ampion™ In Adults With Pain Due To Severe Osteoarthritis Of The Knee," dated 7 June 2019.

It is not intended that each and every table, listing, or graph will be included in the clinical study report (CSR). It is also possible that additional analyses will be conducted after review of the data. Any analyses or summaries not specified in the SAP, but performed after review of the data, will be identified in the CSR as post hoc.

2. OBJECTIVES

2.1 Primary Objectives

The co-primary trial objectives are to evaluate the greater efficacy for pain improvement and functional improvement of 4 mL Ampion versus saline IA injection when applied to patients suffering from severe OAK. Efficacy will primarily be assessed with WOMAC A pain and WOMAC C function scores.

Mean change in the WOMAC A weekly pain scores from Baseline to Week 12 will be compared between Ampion and saline control. Mean change in WOMAC C function score from Baseline to Week 12 will be compared between Ampion and saline control. This will ensure that Ampion is superior to saline in improving pain and function.

2.2 Secondary and Exploratory

The secondary trial objectives include the following evaluations of an intra-articular injection of Ampion, as compared to saline, when administered to patients suffering from severe OAK: assessment of safety; improvement in the Patient Global Assessment; improvement in pain utilizing a 100 mm VAS scale; the amount of rescue analgesia.

Primary and secondary objectives will be reported by age subsets, race, and sex.

Exploratory objectives include analyzing the effect, if any, of the efficacy of Ampion, as compared to saline, on stiffness; Outcome Measures in Rheumatology Clinical Trials and Osteoarthritis Research Society International (OMERACT-OARSI) responder status; proportion of treatment failures; and change in pain and function as a cumulative distribution function (CDF) plot.

Analysis with biomarkers, such as those for analgesic use and chemokines and cytokines associated with inflammation and cartilage regeneration, may be analyzed to compare differences between treatment groups using urine samples collected at Week 12.

Additionally, a Bayesian meta-analysis will be performed for WOMAC A pain sub score using all KL 4 patients in studies AP-003A, AP-003-B, AP-003-C, AP-004, and AP-013.

2.3 Study Design

This is a randomized, saline-controlled, double-blind study with a 28-day screening period for each patient followed by a 24-week participation period. This trial will use an adaptive design with promising zone method for sample size re-estimation at the interim analysis.

The clinical effects of treatment on severe OAK will be evaluated at baseline 2, 4, 6, 8, 10, 12, and 24 weeks, using the Western Ontario and McMaster Universities Arthritis Index (WOMAC®) osteoarthritis Index 3.1, and the Patient's Global Assessment of disease severity (PGA).

Daily assessments of WOMAC A score using a subject diary will be assessed for seven days beginning one week prior to the clinic visits at Week 0 (baseline) and Week 12 to allow for calculation of a weekly average of WOMAC A Daily Pain scores, which will be used for conducting the primary efficacy analysis.

The WOMAC® is a validated pain scoring system and sets the standard for the patient response. In order not to bias the collection of data, only questions from the validated WOMAC pain scale will be asked of patients. Clinical meaningfulness will be determined by the end results of this trial, specifically by the apparent clinical benefit versus any adverse events or any increased apparent risk. Safety will be assessed by recording adverse events (through 24 hours post-dose and at all follow up contacts) and physical examination and vitals (Baseline, Weeks 6, 12 and 24).

2.4 Sample Size

This trial will use an adaptive design with promising zone method for sample size re-estimation.

This trial is designed for 1,034 patients, 517 per study arm, randomized 1:1 for Ampion and saline IA injection. There will be an interim analysis when at 724 patients, 362 per study arm, have completed their Week 12 visit.

A promising zone analysis will be performed at the interim. Following the interim analysis, enrollment will either a) continue up to a total of 1034 patients, 517 patients per study arm, randomized 1:1 for Ampion and saline IA injection; b) allow for sample size re-estimation up to a maximum of 1.5 times the original trial size for a total of 1,552 patients, 776 per study arm, randomized 1:1 for Ampion and saline IA injection.

The Ampion effect was estimated as the average change in pain from Baseline to Week 12 for KL 4 subjects receiving Ampion from all prior single injection studies randomized 1:1 with saline control. Saline effect was estimated as the average change in pain from Baseline to Week 12 for KL 4 subjects receiving saline from all prior single injection studies randomized 1:1 for Ampion and saline. The sample size for assessing the WOMAC A Pain score alone would be 690 total patients for 90% power.

The Ampion effect was estimated as the average change in function from Baseline to Week 12 for KL 4 subjects receiving Ampion from all prior single injection studies randomized 1:1 with saline control. Saline effect was estimated as the average change in function from Baseline to Week 12 for KL 4 subjects receiving saline from all prior single injection

studies randomized 1:1 for Ampion and saline. The sample size for assessing the WOMAC C Function score alone would be 984 patients for 90% power.

The total sample size is chosen as the larger of the two values. To be conservative sample sizes were estimated with a 5% dropout so 1,034 patients will be enrolled to achieve 984 total patients. All ITT patients will be included in the statistical analysis. At the assumed effect size there is a 90% probability of success at 1,034 patients.

The study design of 1,034 total patients achieves 97% power assuming a WOMAC A pain score with mean pain decreases of -0.83 for Ampion and -0.62 for saline and a standard deviation of 0.85.

3. STUDY ENDPOINTS AND COVARIATES

3.1 Primary Endpoints

The trial will test whether Ampion can produce a superior improvement in pain vs. saline control and a superior improvement in function vs. saline. This goal is achieved using two co-primary endpoints:

Change in the WOMAC A weekly pain sub score between Baseline and Week 12
Change in WOMAC C physical function sub score between Baseline and Week 12

An average of the daily WOMAC A pain assessments will be used to calculate the weekly pain score, at baseline and at Week 12.

Mean 12-week change in the WOMAC A pain score will be compared between Ampion and saline control. Mean 12-week change in WOMAC C physical function score will be compared between Ampion and saline control.

3.2 Secondary Endpoints

- Change in PGA between Baseline and Week 12
- Change in 100mm VAS Pain scale between Baseline and Week 12
- Use of rescue analgesia (amount of acetaminophen used between Baseline and Week 12)
- Incidence and severity of treatment-emergent adverse events (TEAEs)

3.3 Exploratory Endpoints

- Change in WOMAC B stiffness sub score between Baseline and Weeks 2, 4, 6, 8, 10, 12, and 24
- Response status based on the OMERACT-OARSI criteria at Weeks 2, 4, 6, 8, 10, 12, and 24
- Proportion of treatment failures at Week 12
- Change in pain and function analysis as a cumulative distribution function (CDF) plot
- A Bayesian meta-analysis for WOMAC A pain sub score including KL 4 patients from studies AP-003-A, AP-003-B, AP-003-C, AP-004, and AP-013

- Change in WOMAC A pain sub score between Baseline and Weeks 2, 4, 6, 8, 10, and 24
- Change in WOMAC C function sub score between Baseline and Weeks 2, 4, 6, 8, 10, and 24
- Change in WOMAC A pain sub score between Baseline and Weeks 2 through 12 and Weeks 2 through 24
- Change in WOMAC C function sub score between Baseline and Weeks 2 through 12 and Weeks 2 through 24
- Change in PGA between Baseline and Weeks 2, 4, 6, 8, 10, and 24
- Change in WOMAC C physical function sub score between Baseline and Weeks 2, 4, 6, 8, 10, 12, and 24
- Difference in urine analysis between treatment groups at Week 12

3.4 Covariates

Analyses of change will include treatment assignment, Week, and the Baseline measure of the variable being analyzed as a covariate in the model.

3.5 Subset analyses

The two co-primary endpoints and three efficacy secondary endpoints will be re-analyzed separately for (a) ages <60, [60, 70), and ≥70, (b) men and women, (c) white and other, (d) Non-Hispanic and Hispanic.

3.6 Sensitivity analyses

The two co-primary endpoints and three efficacy secondary endpoints will be re-analyzed separately in the per-protocol population to exclude treatment failures.

4. HYPOTHESES

4.1 Primary Hypotheses

The trial has co-primary hypotheses – requiring success on both for a successful trial. First, pain reduction must be demonstrated for Ampion vs. saline control. Next a function improvement must be shown on Ampion vs. saline control.

Reduction in Pain, as measured by a change in the WOMAC A pain subscore as measured by the 5-point Likert scale between Baseline and Week 12, will be greater in patients treated with Ampion than with saline:

Null Hypothesis:	$\mu A = \mu S$
Alternate Hypothesis:	$\mu A < \mu S$

Where μA = mean pain reduction in Ampion arm; μS = mean pain reduction in saline arm. Lower mean change scores indicate greater pain reduction.

Improvement in function, as measured by a reduction in the WOMAC C function sub score as measured by the 5-point Likert scale between Baseline and Week 12, will be greater in patients treated with Ampion than with saline:

Null Hypothesis: $\mu A = \mu S$
Alternate Hypothesis: $\mu A < \mu S$

Where μA = mean function reduction in Ampion arm; μS = mean function reduction in saline arm. Lower mean change scores indicate greater improvement in function.

5. DEFINITIONS

5.1 Abbreviations

Abbreviation	Definition
ATC	Anatomical Therapeutic Chemical Classification System
ITT	Intent to Treat Population
MedDRA®	Medical Dictionary for Regulatory Activities
mL	Milliliter
mmHg	Millimeters of mercury
MMRM	Mixed-effects model with repeated measures
NA	Not applicable
OA	Osteoarthritis
OARSI	Osteoarthritis Research Society International
OMERACT	Outcome Measures in Rheumatology
PGA	Patient's global assessment of disease severity
PP	Per protocol population
SAE	Serious adverse event
SAS	Statistical Software from SAS Institute
SD	Standard deviation
SEM	Standard error of the mean
TEAE	Treatment-emergent adverse event
WOMAC®	Western Ontario and McMaster Universities Arthritis Index

5.2 Definitions

Adverse Event (AE)

An adverse event (AE) is defined as any undesired medical occurrence in a patient or clinical investigation patient receiving a pharmaceutical product and which does not necessarily have a causal relationship with this treatment. Therefore, an AE can be any unfavorable sign and unintended sign (including an abnormal laboratory finding),

symptom, or disease temporarily associated with the use of a study drug, whether or not related to the study drug.

AEs will be graded for severity using the following categories. Missing grade will be assigned a grade of 3 (severe) in tabulations:

Grade 1 (MILD): The symptom is barely noticeable to the study patient and does not influence performance or functioning. Concomitant medication is not ordinarily indicated for relief of mild AEs.

Grade 2 (MODERATE): The symptom is of sufficient severity to make the study patient uncomfortable and to influence performance of daily activities. Concomitant medication may be indicated for relief of moderate AEs.

Grade 3 (SEVERE): The symptom causes severe discomfort, sometimes of such severity that the study patient cannot continue in the study. Daily activities are significantly impaired or prevented by the symptom. Concomitant medication may be indicated for relief of severe AEs.

Relationship to study drug will be coded using the following categories. Missing relatedness will be assigned to 'Related' in tabulations:

Unrelated: The adverse event is unlikely to have been caused by study drug.

Possibly related: It is unclear whether the adverse event may have been caused by study drug.

Related: The adverse event is likely to have been caused by study drug.

Relationship to procedure will be coded using the following categories. Missing relatedness will be assigned to 'Related' in tabulations:

Unrelated: The adverse event is unlikely to have been caused by procedure.

Possibly related: It is unclear whether the adverse event may have been caused by procedure.

Related: The adverse event is likely to have been caused by procedure.

Treatment-Emergent AE:

A treatment-emergent AE (TEAE) is any AE that begins or increases in severity after the initial dose of study drug.

Serious Adverse Event:

A serious adverse event (SAE) isany untoward medical occurrence that occurs at any dose that:

- Results in death

- Is life-threatening (patient is at immediate risk of death from the event as it occurred)
- Requires in-patient hospitalization (formal admission to a hospital for medical reasons) or prolongation of existing hospitalization
- Results in persistent or significant disability/incapacity
- Results in a congenital anomaly/birth defect

Hospitalizations for elective surgery or other medical procedures that are not related to a treatment-emergent AE are not considered SAEs.

Age

Patient's age is defined as its integer value in years at enrollment.

Baseline

For any variable, unless otherwise defined, Baseline at Day 0, Visit 1. If data are not available from Visit 1, then the last measurement recorded prior to Day 0 will be selected from the screening period.

Change from Baseline:
The arithmetic difference between a post-Baseline value and the Baseline value:
Change from Baseline = (Post-Baseline Value – Baseline Value)
Percentage Change from Baseline = [(Post-Baseline Value – Baseline Value) / Baseline Value] x 100

End of Study

End of study is at Visit 10 (Week 24), unless terminated early.

Enrollment date

Enrollment date is the same as the randomization date and is designated Day 0.

Study drug

Study drug in this study is 4 mL Ampion injection or 4mL saline.

Randomization date

Randomization date is the day the patient is assigned a randomization number on study Day 0.

Study Day 0

Day 0 is defined as the first day that study drug is administered to the patient.

Study Day

Day of treatment: study day = (visit date - date of Study Day 0).

Treatment failure

Treatment failures are defined as patients who receive additional medications or surgical procedures in order to treat worsened symptoms in the treated knee during the study; specifically, treatment failures are any patients who:

- Increase their corticosteroid dose of ≤ 10 mg above their starting dose of corticosteroid during the study to treat worsened symptoms in the treated knee;
- Increase use of NSAID pain medications by a clinically significant amount during the study to treat worsened symptoms in the treated knee;
- Begin opioid treatment during the study to treat worsened symptoms in the treated knee;
- Receive intra-articular injection of pain medicine in the treated knee during the study to treat worsened symptoms in the treated knee;
- Undergo total knee replacement or other surgical procedure in the treated knee prior to assessment of the primary efficacy endpoint.

6. ANALYSIS SUBSETS

6.1 Data Subsets

6.1.1 Safety Analysis Set

The safety analysis population is defined as all patients who are randomized and receive study medication (Ampion or saline). Patients will be analyzed as treated. Summaries of data will include all collected data.

6.1.2 Intent to Treat (ITT) Analysis Set

The intent-to-treat (ITT) analysis population is defined as all patients who are randomized. All efficacy analyses will be performed in the ITT population. Patients will be analyzed as randomized.

The ITT analysis set is the primary effectiveness analysis population. For the primary effectiveness analysis at Week 12, multiple imputation will be used to assign a value to those cases with missing data. The full conditional specification method with predictive means matching as described in Berglund & Heeringa (2014) will be used. This method uses all of an individual's known primary outcome measures at Baseline, 2, 4, 6, 8, 10, and/or 12 weeks to impute any missing values.

6.1.3 Per Protocol (PP) Analysis Set

The per protocol (PP) analysis population is defined as all patients who met all entry criteria and had no major protocol violations. All efficacy analyses will be repeated in the PP population. These analyses will be supportive of the ITT analysis. Patients will be analyzed as treated. No imputations will be performed.

7. INTERIM ANALYSIS AND EARLY STOPPING GUIDELINES

There will be one interim analysis once 724 subjects have reached final assessment at Week 12.

At the interim analysis, the conditional probability based upon the observed treatment effect on the WOMAC C function score will be calculated. The minimal conditional probability that allows a sample size increase without inflating Type I error under the promising zone for a sample size increase up to 150% is 0.391. Any conditional power < 0.391 results in going to 1,034 patients, randomized 1:1 for Ampion and saline IA injection. Sample size re-estimation according to the promising zone is done for conditional powers between 0.391 and 0.90 with a sample size increase capped to 1,552 total patients, randomized 1:1 for Ampion and saline IA injection.

The final analysis, defined by a sample size established using the promising zone, will use the critical value with one-sided $\alpha = 0.025$ for WOMAC A (Ampion vs. saline control) and WOMAC C (Ampion vs. saline control).

7.1 Data Handling and Electronic Transfer of Data

See Data Management Plan (DMP).

7.2 Handling of Missing and Incomplete Data

All data collected under this study protocol will be included in the assessment of patient safety. Missing or incomplete AE data will assume greatest relationship to study drug and/or severity. Incomplete adverse event and concomitant medication dates will be imputed as described in Section 12. If imputed dates are used, then they will be identified as such in the final study report. If an AE start date is missing or partially missing and no additional information is available from the site in order to establish whether the event started before or after the dose of study drug, the event will be considered to have started after dose. Partially missing dates where the month and year is prior to Day 0 will be classified as pre-dose.

For the co-primary effectiveness analysis of WOMAC pain change and WOMAC C function change, missing change scores will be imputed when the ITT analysis population is used. Missing data will be imputed by the fully conditional specification predictive mean matching multiple imputation approach of Berglund and Heeringa (2014). A sensitivity analysis will be conducted in which the missing 12-week endpoint is replaced by the Baseline WOMAC pain score.

For the secondary endpoints of PGA, 100mm VAS, and number of acetaminophen pills used, missing values will be imputed when the ITT analysis population is used. Imputation methods employed will be multiple imputation (SAS PROC MI and SAS PROC MIANALYZE) and single imputation methods (i.e., regression-based estimates, Baseline imputation, LOCF, and a worst/best case imputation).

Per the WOMAC User Guide, the patient's response will be regarded as invalid, and the subscale(s) deficient, in the event that two pain, both stiffness, or more than four physical function items are omitted. Further, if one pain, one stiffness, or 1-3 physical function units

are missing, the average value for the subscale will be substituted in lieu of the missing item value(s).

7.3 Detection of Bias

Any breaking of the blind for individual patients prior to formal unblinding will be documented in the clinical study report. Data collected after unblinding will be noted.

7.4 Outliers

No formal outlier tests are planned. If necessary, values outside the pre-defined range, such as a WOMAC score not between zero and four, or a PGA score not between one and five, would be queried and excluded prior to database lock.

7.5 Testing/Validation Plan

All statistical analyses will be programmed using SAS® software version 9.3, or later. Graphic displays may be produced using R, version 3.0.0, or later. Standard macros will be used in programming when possible. Testing and validation plans for all programs will be developed in accordance with contract research organization guidelines and will include independent programming of tables and analyses.

8. STATISTICAL METHODS OF ANALYSIS

8.1 General Principles

Data will be summarized by saline and Ampion arms. Descriptive statistics on continuous variables will include mean, standard deviation, median, 25^{th} and 75^{th} percentiles, and range. Change from Baseline will include a 95% confidence interval. Categorical variables will be summarized using frequency counts and percentages. Data listings of individual patient's data will be provided.

Statistical testing for the primary endpoints of pain and function will be performed using a one-sided p-value compared to α=0.025.

Other efficacy analyses will be presented with one-sided p-values. Safety analyses will be presented as two-sided.

8.2 Patient Accountability

The number of patients who are randomized, receive study drug, and complete the study will be summarized. The number of patients included in the safety, ITT, and PP analysis sets will be included in the table. Attendance at each visit, including missed visits, discontinuations, lost to follow-up, and percentage accountability will be summarized. A list of patients who withdraw early will be provided. It will include the reason and timing of the withdrawal. Similarly, the reason any patient is excluded from an analysis set will also be provided. In addition, significant known protocol deviations will be noted for individual patients; a summary table may also be provided.

8.3 Demographic and Baseline Characteristics

Age, race, ethnicity, sex, height, weight, and Baseline measures of WOMAC and PGA will be summarized by treatment arm for all patients randomized, using descriptive statistics.

8.4 Safety Analyses

The safety profile will be based on adverse events, concomitant medications, vital signs, and results of physical examination. All treated patients will be included in the safety analyses.

8.5 Adverse Events

Adverse events will be grouped by system organ class and by preferred term within system organ class according to the latest version of the MedDRA coding dictionary. The number of patients reporting at least one adverse event and each adverse event will be summarized treatment group. Tables and/or narratives of any on-study death, serious or significant adverse events, including early withdrawals because of adverse events, will be provided should they occur.

8.5.1 Concomitant Medications

The number and percent of patients receiving concomitant medications or treatments prior to and during the study and at the final visit will be tabulated and presented overall and by treatment group for the Safety analysis dataset. Concomitant medications, using the generic medicine name, and treatments will be summarized using descriptive statistics and will be presented by type of drug (WHO drug classification preferred term and Anatomical Therapeutic Chemical Classification [ATC] Level 1) by treatment group.

8.5.2 Vital Signs

Vital signs will be listed for each patient. These will include temperature, respiration, pulse, and blood pressure. Summaries over time and changes from Baseline will be provided.

8.5.3 Physical Examinations

Any new or abnormal findings will be recorded as adverse events. Status of each body system will be summarized at each visit.

8.6 Efficacy Analyses

All efficacy variables will be assessed at Baseline (Day 0), Week 2, Week 4, Week 6, Week 8, Week 10, Week 12, and Week 24. Differences between treatments (Ampion and saline arms) will be presented using mean differences and 95% confidence intervals. Endpoints will be compared between Ampion and saline.

For summaries based on the PP, missing data will not be estimated or carried forward in any of the descriptive analyses. For summaries based on the ITT population, estimates from the MMRM will be shown incorporating imputed missing data via the full conditional specification method with predictive mean matching.

8.6.1 Primary Effectiveness Endpoints

Change in WOMAC A pain score as measured by a 5-point Likert scale, from Baseline to Week 12, will be analyzed by including change from Baseline at 2, 4, 6, 8, 10, and 12 weeks in a mixed-effects model with repeated measures (MMRM). Treatment assignment Week, the treatment*week interaction and Baseline WOMAC pain score will be included as fixed effects. Patient ID will be included to denote repeated measures and a compound symmetric covariance matrix will be used for residuals. Least squares means and treatment difference with 95% confidence intervals will be presented.

SAS code used will follow the following format:

```
Proc mixed;
Class PtID treatment week;
Model delta_pain = treatment baseline_pain week treatment*week;
Repeated week / subject=newID type = CS;
Lsmeans active*week / SLiCE=week diff;
run;
```

The primary analysis will include the repeated measures of WOMAC C function sub score from Baseline to Weeks 2, 4, 6, 8, 10, and 12 and will use a mixed-effects model with repeated measures MMRM model, with Week as a fixed effect, and including patient ID as a repeated measure with a compound symmetric covariance matrix for residuals.

The primary effectiveness endpoints will be analyzed using the ITT analysis population. In order to preserve the randomization, any missing values at 12 weeks will be imputed. Week 12 visit data collected more than 14 days after the target day of 84 will be used and not imputed. The primary method of imputation will the full conditional specification method with predictive means matching (Berglund & Heeringa 2014). This imputation method uses all observed values of the primary outcome at Baseline, 2, 4, 6, 8, 10, and 12 weeks to impute an individual's missing values. To determine the possible effect of missing data on the primary analysis, the following sensitivity analyses will be performed: 1 – multiple imputation (SAS PROC MI, using the default method MCMC along with PROC MIANALYZE); 2 – last observation carried forward (LOCF) 3 – a regression-based prediction using individual's prior data and time of prior data will be used to impute Week 12 data; and 4 – a conservative imputation approach will be performed where the greatest pain reduction observed in the study will be imputed for the saline arm and the worst change in pain score observed in the study will be used to impute missing 12-week values in the Ampion arm. Any imputed value that results in more than one decimal place will be rounded to one decimal place.

An analysis of site effect will be performed. If the main effect has a p-value of 0.15 or smaller in a secondary analysis, site will be included in the model and summaries of the primary effectiveness endpoint will be produced by site as well as pooled over sites.

WOMAC A Pain scores and the WOMAC C functions scores and their respective change and percent change from Baseline will be summarized by visit. The Change from Baseline scores will include a 95% CI. For the WOMAC A Pain subscore and WOMAC C Function subscore, a test for the main effect for Ampion vs. saline will be included on the summary tables at weeks indicated as primary or secondary endpoints.

8.6.2 Secondary Effectiveness Endpoints

Secondary effectiveness endpoints will be analyzed using the ITT and the PP analysis populations. All secondary efficacy tests will be performed and reported with one-sided p-values for superiority of Ampion.

Change from Baseline to 2, 4, 6, 8, 10, 12, and 24 weeks for the PGA will be analyzed using a MMRM model. Treatment assignment, Week, and Baseline PGA will be included as fixed effects and a compound symmetric covariance matrix for residuals to reflect the repeated measures.

Change from Baseline to 12 weeks for the 100mm VAS Pain Score will be analyzed using a MMRM model. Treatment assignment, and Baseline PGA will be included as fixed effects and a compound symmetric covariance matrix for residuals to reflect the repeated measures.

Use of rescue analgesia (amount of acetaminophen used) will be compared by treatment arm. A Wilcoxon rank-sum test will be used to compare overall pill count use over the study between saline and Ampion.

8.6.3 Subset analyses

The two co-primary endpoints and three secondary effectiveness endpoints will be analyzed by subsets for:

(a) ages <60, [60, 70), and ≥70,
(b) men and women,
(c) white and other,
(d) Non-Hispanic and Hispanic

8.6.4 Sensitivity analyses

The two co-primary endpoints and three efficacy secondary endpoints will be re-analyzed separately in the per-protocol population to exclude treatment failures.

8.6.5 Exploratory Endpoints

A Bayesian meta-analysis for Likert WOMAC A pain subscore including KL4 patients from studies AP-003-A, AP-003-B, AP-003-C, AP-004, and AP-013. The model is constructed for WOMAC A change from baseline for patient *i* in study *j*

$$Y_{i,j} \sim N(\mu_{i,j}, \sigma^2)$$

$$\mu_{i,j} = \alpha_j + \delta_j T_i + \beta BL_i$$

$$\alpha_j \sim N(\alpha, \sigma_\alpha^2)$$

$$\delta_j \sim N(\delta, \sigma_\delta^2)$$

$$\alpha \sim N(0,100)$$

$$\delta \sim N(0,100)$$

$$\sigma_\alpha \sim U(0,10)$$

$$\sigma_\delta \sim U(0,10)$$

$$\sigma \sim U(0,10)$$

Where T_i is the treatment assignment (0 saline or control; 1=Ampion) and BL_i is the baseline WOMAC A score. α_j's represent the control means across studies. δ_j's represent the treatment effects across studies. σ_α and σ_α represent the heterogeneity of control rates and treatment effects, respectively. Sensitivity analyses will be reported that increase the priors on the variance parameters from U(0,10) to U(0,500). The posterior means and 95% credible intervals will be reported for each treatment by study and treatment effect by study.

Change in WOMAC A pain sub score between Baseline and Weeks 2, 4, 6, 8, 10, and 24 will be analyzed using the MMRM model described for the primary analysis.

Change in WOMAC B stiffness sub score between Baseline and Weeks 2, 4, 6, 8, 10, 12 and 24 will be analyzed using the MMRM model described for the primary analysis.

Change in WOMAC C physical function sub score between Baseline and Weeks 2, 4, 6, 8, 10, and 24 will be analyzed using the MMRM model described for the primary analysis.

Change in PGA between Baseline and Weeks 2, 4, 6, 8, 10, and 24 will be analyzed using the MMRM model described for the primary analysis.

OMERACT-OARSI responder status, defined using Scenario D criteria, will be analyzed using separate logistic regression models for each visit.

Treatment failures will be analyzed with a Fisher's exact test.

Cumulative distribution function (CDF) plots will be used to compare WOMAC A pain and WOMAC C function endpoints for Ampion vs. saline arms.

If collected, the differences in urine analysis will be evaluated between treatment groups at Week 12.

Continuous secondary endpoints and their change from Baseline will be summarized by visit and treatment arm. Change from Baseline will include a 95% CI.

8.6.6 Post Hoc/Ad Hoc Analyses

Any analysis not described in this plan will be considered post-hoc and identified as such in the CSR.

8.6.7 Multiplicity Adjustment

Hierarchical testing for the ITT analysis population will be applied to the analysis of the primary endpoints. It will also be utilized for the secondary endpoints, assuming significance of the co-primary effectiveness endpoints are established. Given significance

for the primary effectiveness tests (WOMAC A and WOMAC C (as specified)), then and only then would PGA score be tested at the one-sided alpha of 0.025; if the PGA score is significant then and only then, will the 100 mm VAS hypotheses be tested at a one-sided alpha of 0.025; if the 100 mm VAS hypothesis is significant then, and only then, will the amount of rescue analgesia through Week 12 be tested at a one-sided alpha 0.025. Multiplicity adjustment will not be applied to exploratory endpoints.

9. CHANGES FROM PROTOCOL-SPECIFIED ANALYSES

Modifications to the planned statistical analyses should be minimized. Nonetheless, the data obtained from the study may indicate that the planned analyses are inappropriate, that additional analyses need to be performed, or that the design of the study needs to be modified, due to factors such as the distribution of the data or imbalance in important covariates. The study report will provide a detailed explanation for any deviations from the planned analyses.

10. LIST OF PLANNED TABLES, FIGURES, AND LISTINGS

Tables are categorized and numbered in accordance with ICH E3 guidelines. Each table, figure and listing is presented by treatment arm.

Accountability tables listed below will also include an overall column. Listings will be sorted by treatment, patient ID, and by visit, if multiple visits exist.

10.1 Tables

14.1 Accountability

14.1.1	Accountability (Analysis population: All Enrolled)
14.1.2	Enrollment by Site (Analysis population: All Enrolled)
14.1.3	Analysis Populations (Analysis population: All Enrolled)
14.1.4	Patient Disposition (All Screened)
14.1.5	Major Protocol Deviations (Analysis population: All Enrolled)
14.1.6.1	Demographics and Baseline Characteristics (Analysis population: ITT)
14.1.6.2	Demographics and Baseline Characteristics (Analysis population: PP)

The following efficacy tables will be provided both for ITT and PP. ITT tables will have a .1 suffix and PP tables will have a .2 suffix added to the table numbers (ex. 14.2.1.1 = ITT and 14.2.1.2 = PP)

14.2 Efficacy

14.2.1	Summary of WOMAC A Pain Subscale
14.2.2	Summary of WOMAC B Stiffness Subscale
14.2.3	Summary of WOMAC C Physical Function Subscale
14.2.4	Summary of 100mm VAS Pain Score
14.2.5	Summary of PGA
14.2.6	Summary of Rescue Analgesia Use
14.2.7	Summary of OMERACT-OARSI responder rates
14.2.8	Summary of WOMAC A Pain Subscale by Age

14.2.9	Summary of WOMAC A Pain Subscale by Sex
14.2.10	Summary of WOMAC A Pain Subscale by Race
14.2.11	Summary of WOMAC A Pain Subscale by Ethnicity
14.2.12	Summary of WOMAC C Pain Subscale by Age
14.2.13	Summary of WOMAC C Pain Subscale by Sex
14.2.14	Summary of WOMAC C Pain Subscale by Race
14.2.15	Summary of WOMAC C Pain Subscale by Ethnicity
14.2.16	Summary of 100mm VAS Pain Score by Age
14.2.17	Summary of 100mm VAS Pain Score by Sex
14.2.18	Summary of 100mm VAS Pain Score by Race
14.2.19	Summary of 100mm VAS Pain Score by Ethnicity
14.2.20	Summary of PGA by Age
14.2.21	Summary of PGA by Sex
14.2.22	Summary of PGA by Race
14.2.23	Summary of PGA by Ethnicity
14.2.24	Summary of Rescue Analgesia Use by Age
14.2.25	Summary of Rescue Analgesia Use by Sex
14.2.26	Summary of Rescue Analgesia Use by Race
14.2.27	Summary of Rescue Analgesia Use by Ethnicity
14.2.28	Summary of treatment failure rates (%)

14.3 Safety

14.3.1	Overall Summary of Treatment-Emergent Adverse Events (Analysis population: Safety)
14.3.2	Incidence of Treatment-Emergent Adverse Events by System Organ Class and Preferred Term (Analysis population: Safety)
14.3.3	Incidence of Treatment-Emergent Adverse Events by System Organ Class and Preferred Term by Maximum Severity (Analysis population: Safety)
14.3.4	Incidence of Treatment-Emergent Related Adverse Events by System Organ Class and Preferred Term (Analysis population: Safety)
14.3.5	Incidence of Treatment-Emergent Related Adverse Events by System Organ Class and Preferred Term by Maximum Severity (Analysis population: Safety)
14.3.6	Incidence of Treatment-Emergent Serious Adverse Events by System Organ Class and Preferred Term (Analysis population: Safety)
14.3.7	Incidence of Treatment-Emergent Adverse Events by Preferred Term in Descending Order of Frequency (Analysis population: Safety)
14.3.8	Incidence of Treatment-Emergent Related Adverse Events by Preferred Term in Descending Order of Frequency (Analysis population: Safety)
14.3.9.1	Summary of Vital Signs and Change from Baseline over Time for Pulse (bpm) (Analysis population: Safety)
14.3.9.2	Summary of Vital Signs and Change from Baseline over Time for Temperature (F) (Analysis population: Safety)
14.3.9.3	Summary of Vital Signs and Change from Baseline over Time for Systolic BP (mmHg) (Analysis population: Safety)

14.3.9.4	Summary of Vital Signs and Change from Baseline over Time for Diastolic BP (mmHg) (Analysis population: Safety)
14.3.9.5	Concomitant Medication Use by ATC Level 1 and WHO Preferred Term (Analysis population: Safety)
14.3.9.6	Medication Started on Study by ATC Level 1 and WHO Preferred Term (Analysis population: Safety)
14.3.9.7	Concomitant Medications by WHO Preferred Term in Descending Order of Use (Analysis population: Safety)
14.3.9.8	Medication Started on Study by WHO Preferred Term in Descending Order of Use (Analysis population: Safety)

10.2 Listings

Subject Accountability

1. Randomization List (including: patient ID, randomization number, randomized treatment and treatment administered, and date of treatment)
2. Inclusion and Exclusion Criteria
3. Protocol Deviations
4. Patients Withdrawing from the Study Prematurely (date and reason)
5. Analysis Populations with Reason for Exclusion (if populations differ)

Demographics and Baseline Characteristics

6. Demographics and Baseline Characteristics (age, sex, race, ethnicity, weight, height, Baseline WOMAC and PGA scores)
7. Medical History
8. Baseline Medication Use

Efficacy

9. WOMAC A, B, and C Scores (includes all subscales)
10. PGA Scores
11. 100mm VAS Pain Scores
12. Rescue Analgesia
13. OMERACT-OARSI responder rates
14. Treatment failures

Safety

15. All Adverse Events [with indication of TEAE]
16. Physical Exams

17. Vital Signs Data [systolic and diastolic blood pressure, pulse rate, respiration, and temperature]

10.3 Figures

Note that all figures, unless otherwise stated, will be line plots showing mean ± SEM at each visit for each treatment arm.

1. WOMAC A Pain Subscale at Baseline and Week 12 [Boxplot]
2. Summary of WOMAC A Pain Subscale over Time
3. Summary of WOMAC B Stiffness Subscale over Time
4. Summary of WOMAC C Physical Function Subscale over Time
5. Summary of PGA over Time
6. Summary of Rescue Analgesia Used [stacked histogram]
7. Summary of OMERACT-OARSI responder rates over time [Bar chart]
8. Change in WOMAC A Pain Subscale from Baseline to Week 12 for subgroups [Forest plot]
9. Change in WOMAC C Function Subscale from Baseline to Week 12 for subgroups [Forest plot]

11. LITERATURE CITATIONS / REFERENCES

SAS Institute Inc. SAS Language: version 8 first edition. SAS Institute, Inc, Cary, NC, USA, 1990.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org/.

12. HANDLING OF MISSING OR INCOMPLETE DATES FOR ADVERSE EVENTS AND CONCOMITANT MEDICATIONS

12.1 Imputation Rules for Partial or Missing Stop Dates

If the month and year are present, impute the last day of the month. If only the year is present, impute December 31 of that year. If the stop date is entirely missing, assume the event or medication is ongoing. If a partial or complete stop date is present and the 'ongoing' or 'continuing' box is checked, then it will be assumed that the AE or concomitant medication stopped and the stop date will be imputed, if partial.

		Stop Date						
		Complete: yyyymmdd		Partial: yyyymm		Partial: yyyy		Missing
Start Date		**<1st Dose**	**≥1st Dose**	**<1st Dose yyyymm**	**≥1st Dose yyyymm**	**<1st Dose yyyy**	**≥1st Dose yyyy**	
Partial: yyyymm	**=1st Dose yyyymm**	2	1	2	1	N/A	1	1
	≠ 1st Dose yyyymm		2		2	2	2	2
Partial: yyyy	**=1st Dose yyyy**	3	1	3	1	N/A	1	1
	≠ 1st Dose yyyy		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1 = Impute the date of first dose
2 = Impute the first of the month
3 = Impute January 1 of the year
4 = Impute January 1 of the stop year

Note: For patients who were never treated (first dose date is missing), partial start dates will be set to the first day of the partial month.

Note: If the start date imputation leads to a start date that is after the stop date, then do not impute the start date.

13. SCHEDULE OF ASSESSMENTS AND PROCEDURES

	Screening	Pre-Baseline (phone)	Baseline	24-hours (phone)	Week 2 (phone)	Week 4 (phone)	Week 6	Week 8 (phone)	Week 10 (phone)	Pre-Week 12 (phone)	Week 12	Week 24	Early Term.
Assessments & Procedures	**Visit -1 Day-28 to 0**	**Visit 0 Day -6 to 0**	**Visit 1 Day 0**	**Visit 2 Day 1**	**Visit 3 Day 14 ± 7**	**Visit 4 Day 28 ± 7**	**Visit 5 Day 42 ± 7**	**Visit 6 Day 56 ± 7**	**Visit 7 Day 70 ±7**	**Visit 8 Day 77 ±7**	**Visit 9 Day 84 ± 7**	**Visit 10 Day 174 ± 7**	
Informed Consent	X												
Inclusion/ exclusion criteria	X		X										
Medical history/prior medications	X		X										
Concomitant medications	X		X	X	X	X	X	X	X		X	X	X
Urine sample			X**								X		
Physical examination	X		X				X				X	X	X
Vital Signs	X		X				X				X	X	X
Randomization			X										
Review Diary Completion Instructions	X								X				
Phone Contact Daily Collection Reminder		X								X			
WOMAC A (Likert)	X	X*	X***		X	X	X	X	X	X*	X***	X	X
WOMAC A (VAS)	X	X*	X***							X*	X***		X
WOMAC B and C (Likert)	X		X		X	X	X	X	X		X	X	X
PGA	X		X		X	X	X	X	X		X	X	X
X-ray[1]	X												
Study treatment			X										
Dispense Rescue medication			X										
Review Rescue medication				X	X	X	X	X	X		X	X	X
Adverse Events			X	X	X	X	X	X	X		X	X	X

Visits are in clinic except for Day 1 and Weeks 2, 4, 8 and 10 when patients will be contacted by telephone

[1]X-ray may be acquired at Screening to satisfy inclusion criteria, "Index knee must be symptomatic for greater than six months with a clinical diagnosis of OA and supported by radiological evidence (KL 4) that is not older than six months prior to the date of screening".

*Patients will complete 7 daily pain score assessments up to and including the Baseline and Week 12 visits to allow calculation of WOMAC A weekly pain scores. The Screening and first Pre-Baseline daily pain score collection may occur on the same day, in office or at home, as appropriate.

**Urine will only be collected at Baseline if necessary to conduct a pregnancy test; collection to occur as applicable

***The WOMAC may occur in office or at home, and may serve as the 7th WOMAC Daily Pain collection