
| | |
|---|---|
| DOCUMENT: | STATISTICAL ANALYSIS PLAN |
| PROTOCOL | |
| Number: | CLIN2002-PO |
| Title: | A Randomized, Double-Blind, Placebo-Controlled, Titration-to-Effect Study of Orally Administered CR845 in Patients with Osteoarthritis of the Hip or Knee |
| Date: | 04NOV2016, Version 3.0 |
| STUDY PHASE | Phase 2 |
| SAP STATUS | SAP Version 1.0 |
| SAP VERSION DATE | 18JAN2017 |
| SPONSOR: | CARA Therapeutics, LP<br>Four Stamford Plaza<br>Stamford CT 06902<br>(203) 406 3700 |

## SIGNATURE(S)/APPROVAL PAGE

**AUTHOR(S)**

[redacted]

20 Jan 2017
Date

CARA Therapeutics LP

[redacted]

20 JAN 2017
Date

**APPROVERS**

[redacted]

1/20/17
Date

CARA Therapeutics LP

## TABLE OF CONTENTS

# 1. LIST ◎F ABBREVIATIONS

| | |
|---|---|
| AE | Adverse Event |
| ALT | Alanine Aminotransferase |
| AST | Aspartate Aminotransferase |
| ATC | Anatomical-Therapeutic-Chemical |
| BMI | Body Mass Index |
| BUN | Blood Urea Nitrogen |
| CI | Confidence Interval |
| CRF | Case Report Form |
| CSR | Clinical Study Report |
| DB | Double-blind |
| ECG | Electrocardiogram |
| FAP | Full Analysis Population |
| FDA | US Food and Drug Administration |
| ICH | International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use |
| IV | Intravenous |
| IWRS | Interactive Web Response System |
| LNH | Low/Normal/High |
| LOCF | Last Observation Carried Forward |
| MedDRA | Medical Dictionary for Regulatory Activities |

| | |
|---|---|
| MMRM | Mixed Models with Repeated Measures |
| NRS | Numerical Rating Scale |
| OA | Osteoarthritis |
| PP | Per Protocol |
| PR | ECG PR Interval |
| QRS | ECG QRS Duration |
| QT | ECG QT Interval |
| QTcF | ECG QT Interval, Fridericia's Correction |
| RR | ECG RR Duration |
| SAP | Statistical Analysis Plan |
| SAE | Serious Adverse Event |
| SAF | Safety Population |
| SD | Standard Deviation |
| SE | Standard Error |
| SI | Système International |
| SOC | MedDRA System Organ Class |
| TEAE | Treatment-emergent Adverse Event |
| WBC | White Blood Cell |
| WOMAC | Western Ontario and McMaster Osteoarthritis Index |
| WHO DD | World Health Organization Drug Dictionary |

# 2. INTRODUCTION

This statistical analysis plan (SAP) summarizes the planned presentation and analyses of the clinical data from CARA Therapeutics Protocol CLIN2002-PO (version 3.0 dated 04 November 2016), This is a multicenter, randomized, double-blind, placebo-controlled, titration-to-effect study of orally administered CR845 in patients with osteoarthritis (OA) of the hip or knee.

CR845 is a selective kappa-opioid receptor agonist with a peripheral mechanism of action as a novel therapeutic agent for the treatment of acute and chronic pain. CR845 is a potent and selective kappa receptor agonist with more than 30,000-fold selectivity over mu and delta opioid receptors and with no known activity at other non-opioid receptors, ion channels, or transporters. Its unique peptidic structure significantly differs from that of small molecule kappa opioid agonists developed to date, which, for the most part, are active within the Central Nervous System. Intravenous (IV) and oral formulations of CR845 are being developed as novel treatments for post-operative and chronic pain, respectively, with the potential for reduced side effects and improved tolerability compared with currently available opioid medications.

The planned analyses identified in this SAP may be included in regulatory submissions and/or future manuscripts. Additional exploratory analyses, not identified in this SAP, may be performed to support the clinical development program. Unplanned, analyses that are performed but not identified in this SAP will be clearly identified in the clinical study report (CSR). The structure and content of this SAP provides sufficient detail to meet the requirements identified by the US Food and Drug Administration (FDA) and International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH): Guidance on Statistical Principles in Clinical Trials.

# 3. STUDY ◎BJECTIVES

The primary objective of this study is to characterize the analgesic efficacy of orally administered CR845 in patients with OA of the hip or knee.

There are 4 secondary objectives:

- To compare the treatment effect on patient function, pain and stiffness between CR845- and placebo-treated patients as measured by the Western Ontario and McMaster Osteoarthritis Index (WOMAC).
- To compare the treatment effect on Patient Global Impression of Change between CR845- and placebo-treated patients.
- To compare the use of analgesic rescue medications between CR845- and placebo-treated patients.

- To characterize the safety and tolerability of titration-to-effect administration of oral CR845.

# 4. STUDY DESIGN

## 4.1. General Study Design and Plan

This is a multicenter, randomized, double-blind, placebo-controlled, titration-to-effect study of orally administered CR845 in patients with OA of the hip or knee.

CR845 is a highly charged peptide with limited bioavailability when administered orally. Based upon previous studies with oral CR845, a significant inter-patient variability in plasma concentrations of CR845 has been observed that is likely due to differences in the absorption of CR845 from the gastrointestinal tract. It is presumed that patients who absorb CR845 well will not need higher doses to achieve an adequate analgesic response and patients who absorb the drug poorly will need to be titrated up to a higher dose to achieve a similar effect. The purpose of this Titration-to-Effect Period is to identify the optimal dose level of CR845 for each patient, based upon tolerability and efficacy, prior to the beginning of the 4-week Maintenance Treatment Period. Once patients enter into the Maintenance Treatment Period, the dose level can no longer be changed.

The study schedule is comprised of the following:

- Screening period (Visit 1/Day -14 to Day -1)
    - Patients will keep a diary of their pain scores for the 7 days prior to baseline (Day 1). Only patients who report a daily pain intensity in the index joint ≥ 5 during 4 or more of the last 7 days prior to randomization, with 2 consecutive days ≥ 5 occurring *just prior to randomization*, and continue to satisfy all other inclusion/exclusion criteria will be entered into the Titration-to-Effect period and will be randomized to receive either CR845 or placebo in a 2:1 ratio stratified by primary OA joint (knee versus hip).
- Blinded 4-week Titration-to-Effect Period (Days 1-28 / Weeks 1-4):
    - Patients will receive either CR845 1mg or matching placebo twice daily (BID) for 1 week (7 days). Placebo patients will continue to take matching placebo throughout the study, regardless of the decisions described below which will be made for all patients.
    - At the end of each week, there will be a clinic visit to review the patient diary, rescue medication use, Adverse Events, any other safety related events and

serum sodium levels. The clinical team will decide on the next dose level to give the patient for the following week.

- At the end of each week for the first 3 weeks, the clinical team will decide whether or not the patient should stay at their current dose level BID, move to the next higher dose level BID (if not currently taking 5mg), move to the next lower dose level BID (if not currently taking 1mg) or discontinue from the study.

- At the end of the 4th week, the clinical team will decide whether or not the patient should stay at their current dose level BID or move to the next lower dose level BID (if not currently taking 1mg) for the maintenance treatment period or whether they should be discontinued from the study.

- Blinded 4-week Maintenance Treatment Period (Days 29-57/Weeks 5-8):

  - 4 weeks of treatment at the optimal dose of CR845 (or matching placebo) following the Titration-to-Effect period.

  - Patients will return to the clinic at the end of Week 6 (Day 43) and Week 8 (Day 57) to complete safety and efficacy assessments.

- Follow-Up Period (1-week post treatment)

  - 7-10 days after the last dose of study medication, patients will return to the study site for a follow-up safety evaluation

If at any time during the Titration-to-Effect or Maintenance Treatment Period, the patient's serum sodium level is > 150 mmol/L, dosing should be paused for patient assessment (i.e., retesting of sodium approximately 1 hour after the patient is given a glass of water, and clinical evaluation). Based upon the outcome of the clinical evaluation by the clinical staff, the dosing of study medication will either be continued or terminated.

Study medication will be taken BID with a full glass of water (250 mL). It is recommended that the medication be taken in the morning (at least 2 hours before or 4 hours after breakfast) and on a night (at least 4 hours after the last meal).

Acetaminophen is the only allowable rescue medication for pain beginning from Day -5 until the end of the Maintenance Treatment Period and will be provided as 325-mg tablets and its use (number of tablets taken in the previous 24 hours) will be reported each evening in the patient diary. Acetaminophen should be taken as directed for pain, with a maximum allowable dose of 8 tablets per day (given 1 to 2 tablets every 4 to 6 hours as needed). However, the use of acetaminophen is not allowed for the 12 hours prior to a scheduled office visit until after the efficacy assessments have been completed, in order to minimize the confounding effects of rescue medication on these measures.

Patients will record their pain intensity score (all days), the number of rescue medication tablets taken for target or other pain (Day -5 onwards) and AM/PM dosing of study drug (Day 1 onwards) in a patient diary.

## 4.2. STUDY POPULATION

The study population will consist of male and female patients aged ≥ 25 years with moderate to severe pain associated with OA of the hip or knee. All inclusion/exclusion criteria below must be met before the patient is enrolled.

### 4.2.1. Inclusion criteria

1. Voluntarily provides written informed consent to participate in the study prior to any study procedures.

2. Is able to speak, read, and communicate clearly in English or Spanish; is able to understand the study procedures.

3. Male or female ≥ 25 years of age.

4. Body mass index (BMI) ≤ 40 $kg/m^2$

5. Has OA of the hip or knee according to American College of Rheumatology (ACR) criteria.

6. Reports an average pain intensity level ≥ 5 in the index joint at Screening on a 0-10 NRS scale.

7. Is either opioid-naïve (defined as taking < 10 mg a day of morphine equivalent 14 days prior to screening) or opioid-experienced. If receiving opioid analgesic medication for OA, patients must be on a stable dose ≤ 40 mg of morphine equivalents for 14 days prior to screening.

8. Willing to discontinue all currently used pain medication beginning 5 days prior to the Baseline Visit and throughout the study. Acetaminophen use is allowed.

9. If female:

a. Of childbearing potential – the patient must be willing to practice an acceptable form of birth control (defined as the use of an intrauterine device, a barrier method with spermicide, condoms, any form of hormonal contraceptives, or abstinence from sexual intercourse) for the duration of treatment and for at least 3 days following the last dose of study drug.

b. Of non-childbearing potential – the patient must be surgically or biologically sterile (hysterectomy, bilateral oophorectomy, bilateral tubal ligation, or postmenopausal for at least 1 year).

10. If male, the patient must be surgically or biologically sterile. If not sterile, the patient must agree to use an acceptable form of birth control with a heterosexual partner (as described in inclusion criterion #9) or abstain from sexual relations during the treatment period and for 3 days following the last dose of study drug.

11. Is free of other physical, mental, or medical conditions that, in the opinion of the Investigator, would make study participation inadvisable.

12. Reports a daily pain intensity score in the index joint ≥ 5 (on a 0-10 NRS scale) during 4 or more of the last 7 days prior to randomization, with 2 consecutive days ≥ 5 occurring just prior to randomization.

### 4.2.2. Exclusion criteria

1. Has had a joint replacement in the index joint.

2. Has received an intra-articular injection of corticosteroids or hyaluronic acid in the index joint(s) within 3 months prior to the Screening Visit.

3. Has started a new medication for chronic illness within 30 days prior to the Screening Visit.

4. Is receiving opioid analgesic treatment for OA of the hip or knee at a dose > 40 mg of morphine equivalent.

5. Uses antipsychotics, antiepileptics, sedatives, hypnotics, or antianxiety agents, selective serotonin reuptake inhibitors (SSRIs), tricyclic antidepressants with a dose change <30 days prior to day 1 of the study.

6. Has a history or current diagnosis of substance dependence (except caffeine or nicotine) or alcohol abuse, according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5).

7. Has a positive urine drug screen for drugs of abuse at Screening.

8. Has been diagnosed with a condition of hyperhidrosis (excessive sweating) or primary hypodipsia (a reduced sense of thirst).

9. Has a history (within 6 months) of clinically meaningful orthostatic changes in vital signs OR, at Screening, has a decrease in systolic blood pressure by > 20 mm Hg or a decrease in diastolic blood pressure by 10 mm Hg together with an increase in heart rate of > 30 beats per minute when transitioning from supine to standing measurements.

10. Has a medical condition (e.g., a cardiovascular, pulmonary, hepatic, renal, hematologic, gastrointestinal, endocrine [adrenal hyperplasia], immunologic, dermatologic, neurologic, oncologic, or psychiatric) or a significant laboratory abnormality that, in the Investigator's opinion, would jeopardize the safety of the patient or is likely to confound the study measurements.

11. Has had any gastric bypass surgery, sleeve, or lap-band (for weight loss).

12. Has a corrected QT interval > 450 msec in males, or > 470 msec in females or clinically significant abnormality on screening ECG.

13. Has a serum sodium level > 143 mmol/L at Screening or Baseline (i.e., Visit 2 prior to initiating study drug).

14. Has impaired renal function indicated by serum creatinine > 2 × the reference upper limit of normal (ULN).

15. Has a serum alanine aminotransferase (ALT) or aspartate aminotransferase (AST) > 2.5 × the reference ULN, or total bilirubin > 2 × the ULN at Screening.

16. Has, in the opinion of the Investigator, any clinical signs of dehydration or hypovolemia (e.g., symptomatic hypotension) or associated laboratory abnormalities (e.g., elevated hematocrit or elevated blood urea nitrogen [BUN] > 1.5 × the reference ULN) at Screening.

17. Has taken opioid or non-opioid pain medication (e.g., Nonsteroidal anti-inflammatory drug such as naproxen or cyclo-oxygenase-2 inhibitors) within 5 days prior to study drug administration. Acetaminophen use is allowed.

18. Has received another investigational drug within 30 days prior to Baseline or has planned to participate in another clinical trial while enrolled in this study.

## 4.3. RANDOMIZATION AND BLINDING

Randomization

Subjects will be assigned unique subject numbers upon enrollment into the screening period. Subject numbers will be recorded on all clinical investigation documentation (ie, CRFs, clinical drug supply labels, laboratory kits, ECGs, etc.).

Eligible patients will be randomized to receive either CR845 or placebo in a 2:1 ratio, and will be stratified according to their primary joint affected by OA (knee vs. hip). Unique randomization numbers, generated according to the randomization specifications approved by CARA Therapeutics and INC Research, LLC, will be assigned through Medidata Balance and will determine the drug supply given to each subject.

Blinding

CR845 will be provided as enteric-coated tablets at doses of 1 mg, 2.5 mg and 5 mg with matching placebo tablets. All tablets are white in color with no markings and are identical in appearance, regardless of dose.

Because of the known titration scheme during the Titration-to-Effect Period, the clinical staff may be aware of the dose being administered, but all attempts should be made to keep the patients blinded as to the study dose being administered during both titration and treatment.

Emergency unblinding (via IWRS) of treatment assignment for a patient may be necessary due to a medical emergency, such as an Serious Adverse Event (SAE) that is unexpected and for which a causal relationship to study drug cannot be ruled out or any other significant medical event (e.g. pregnancy).

If emergency unblinding is required for a medical emergency:

- Only the Investigator will make the decision to unblind the treatment assignment.
- Only the patient with the medical emergency will be unblinded.

The Investigator should notify the INC Medical Director and/or designee prior to breaking the blind, whenever possible. In all cases, the INC Medical Director and/or designee must be notified within 24 hours of the blind being broken.

## 4.4. STUDY ASSESSMENTS

The schedule of study visits and measurements is shown in Table 1. A visit window of +/- 1 day will be allowed.

**Table 1 Schedule of Visits and Procedures**

| | Visit 1 | Visit 2 | Visit 3 | Visit 4 | Visit 5 | Visit 6 | Visit 7 | Visit 8 | Visit 9 |
|---|---|---|---|---|---|---|---|---|---|
| | Day -14 to Day -1 | Day 1 | Day 8 | Day 15 | Day 22 | Day 29 | Day 43 | Day 57 | Follow-Up Visit |
| **Study Procedure** | *Screening* | *Baseline* | *Titration to Effect Period* | | | | *Maintenance Treatment Period* | | *Follow-Up Period* |
| Informed consent | X | | | | | | | | |
| Inclusion/exclusion criteria | X | X | | | | | | | |
| Medical history, demographics, and identify index joint | X | | | | | | | | |
| Record prior medications[1] | X | X | | | | | | | |
| Height and weight | X | | | | | | | | |
| Drug screening test | X | | | | | | | | |
| Training on PRO completion [2] | X | | | | | | | | |
| Training on Patient Diary Completion [3] | X | | | | | | | | |
| Provide acetaminophen [4] | X | X | X | X | X | X | X | | |
| Discontinue current pain[5] medications | X[5] | | | | | | | | |
| | *Screening* | *Baseline* | *Titration to Effect Period* | | | | *Treatment Period* | | *Follow-Up* |
| Assign to treatment [6] | | X | | | | | | | |
| Review PRO and Patient Diary completion [2,3] | | X | X | X | X | X | X | X | |
| Drug accountability | | X | X | X | X | X | X | X | X |
| Instruct on fluid replacement | | X | X | X | X | X | X | | |
| Dispense study drug [7] | | X | X | X | X | X | X | | |
| Record concomitant medications[8] | X | X | X | X | X | X | X | X | |
| Review of information and titration determination and enter into IWRS | | | X | X | X | X | | | |

1. *Prior Medications:* Prior medications are defined as medications taken at any time during the 30 days prior to Screening until the time of dosing on Day 1, whether continuing or not. Any changes to prior medications (dose, regimen, etc.) during the course of the study must be considered as a new concomitant medication.
2. *Patient-reported Outcome Training:* A member of the clinical staff will review the patient-reported outcome assessments for this study with each patient during the Screening and Baseline Visits. Patient-reported outcome assessments for this study are the Numeric Rating Scale, the Western Ontario and McMaster Osteoarthritis Index, and the Patient Global Impression of Change.
3. *Patient Diary:* Patients will be shown how to capture their daily pain scores and the number of rescue medication tablets taken each day (the preceding 24 hours) in their diaries beginning 7 days before their scheduled Baseline Visit and continuing for the duration of the Titration-to-Effect and Maintenance Treatment Periods.
4. *Rescue Medication:* A bottle of analgesic rescue medication (acetaminophen, 325 mg) will be provided at the Screening Visit to be used (if needed) throughout the study. Patients will be instructed to take acetaminophen as needed for pain (up to 2 tablets of 325mg per administration for a maximum of 8 tablets per day). Patients will also be instructed to bring the bottle of medication back with them for each clinic visit for drug accountability and dispensation of another bottle of rescue medication if needed. Patients should also be instructed not to take any analgesic rescue medication within 12 hours of a scheduled clinic visit to prevent confounding of the pain assessments during the visit.
5. *Discontinuation of Pain Medication:* Patients will be instructed to discontinue their current nonsteroidal anti-inflammatory drugs and opioid analgesic medications starting on Day -5 and continuing for the duration of the study (through Day 57).
6. *Treatment Assignment:* Patients will be enrolled and assigned to treatment following successful completion of the Baseline assessments.
7. *Dispensing of Study Medication:* Study medication will be dispensed to patients in bottles of 15 tablets at the Baseline Visit. The patients will take their first dose of study medication at the clinic.
8. *Concomitant Medications:* Concomitant medications are defined as medications taken at any time from post-dose on Day 1 until the end of the study (i.e., at the Follow-up Visit). Any new concomitant medication added during the Titration-to-Effect and Maintenance Treatment Periods must also have the reason for taking the new concomitant medication recorded in the case report form. Any concomitant medication administered due to an adverse event must also be recorded on the adverse event case report form page.

## Table 1 Schedule of Visits and Procedures (continued)

| | Visit 1 | Visit 2 | Visit 3 | Visit 4 | Visit 5 | Visit 6 | Visit 7 | Visit 8 | Visit 9 |
|---|---|---|---|---|---|---|---|---|---|
| | Day -14 to Day -1 | Day 1 | Day 8 | Day 15 | Day 22 | Day 29 | Day 43 | Day 57 | Follow-Up Visit |
| ***Safety*** | ***Screening*** | ***Baseline*** | ***Titration to Effect Period*** | | | | ***Treatment Period*** | | ***Follow-Up*** |
| Physical examination [9] | X | | | | | | | X | |
| Vital signs [10] | X | X | X | X | X | X | X | X | X |
| 12-lead ECG | X | | | | | | | X | |
| Pregnancy test [11] | X | X | | | | | | | |
| Clinical laboratory evaluations [12] | X | X | | | | | | X | |
| Serum sodium measurement [13] | X | X | X | X | X | X | X | X | |
| Adverse event assessment | | X | X | X | X | X | X | X | X |
| | | | | | | | | | |
| ***Effectiveness*** | ***Screening*** | ***Baseline*** | ***Titration to Effect Period*** | | | | ***Treatment Period*** | | ***Follow-Up*** |
| Daily joint pain (NRS) [14] | X | X | X | X | X | X | X | X | |
| WOMAC Index [15] | | X | X | X | X | X | X | X | |
| Patient Global Assessment of OA[16] | | | | | | | | X | |

*9. Physical Exam:* A physical examination will be performed at Screening and on Day 57 (or early termination). On Day 57, any changes from the previous physical exam conducted at Screening will be noted.
10. *Vital Signs:* Vital signs will be measured at each clinic visit. Vital signs (blood pressure and heart rate) at the Screening visit will be conducted in both the supine and standing positions. For the purpose of these assessments, patients should be assessed in the supine position first and then, after sitting up for at least 1 minute, should be standing for at least 3 minutes prior to measuring blood pressure or heart rate. For the remainder of the study, vital signs will be assessed in the sitting position. Body temperature will be measured at screening and at Day 57 (or early termination).
11. *Pregnancy Testing:* For female patients of childbearing potential, a serum pregnancy test will be conducted at Screening and a urine pregnancy test will be conducted at the Baseline Visit prior to the administration of study drug.
12. *Laboratory Evaluations:* Clinical laboratory evaluations will consist of hematology and blood chemistry and will be conducted at Screening, at Baseline (prior to drug administration), and on Day 57 (or early termination).
13. *Serum Sodium Measurement:* Blood samples to measure serum sodium levels will be taken at all visits (or early termination). If at any time during the study, the patient's serum sodium level is > 150 mmol/L, dosing should be paused for patient assessment (i.e., retesting of sodium in approximately 1 hour after the patient is given a glass of water and clinical evaluation). Based upon the outcome of the clinical evaluation by the clinical staff, the dosing of study medication will either be: a) continued or b) terminated.
14. *Daily Joint Pain:* The average daily index joint pain intensity will be assessed using the numeric rating scale. The numeric rating scale score will be recorded into the electronic case report form by the site at the Screening Visit and by the patient each day (at approximately 6:00 PM) in their patient diary. The patient should make all attempts to assess their joint pain at the same time each day. During the Screening Period, assessments will be conducted for 5 days prior to the Baseline Visit.
15. The Western Ontario will be completed at each clinic visit from Baseline to Day 57 (or early termination).
16. The Patient Global Impression of Change will be completed on Day 57 (or early termination).

A visit window of +/- 1 day will be allowed

# 5. STUDY VARIABLES AND DEFINITIONS

## 5.1. EFFICACY VARIABLES

### 5.1.1. Primary Efficacy Variable

The primary efficacy variable is the weekly mean pain intensity score calculated using the 'average pain over the last 24 hours' scores collected in the daily diaries using an 11-point scale where 0 = no pain, and 10= pain as bad as you can imagine. The weekly mean pain intensity score will be defined as the sum of non-missing daily "average pain over the last 24 hours" scores reported during that week (Days -7 to -1 [Baseline], Days 1-7 [Week 1], Days 8-14 [Week 2], Days 15-21 [Week 3], … , Days 50-56 [Week 8]) divided by the number of days with non-missing scores for that week. If a subject reports less than 3 days of pain scores during a week, the weekly mean pain intensity score will be set to missing.

The primary efficacy endpoint is defined as the change from baseline at Week 8 with respect to the weekly mean of the daily 24-hour pain intensity for the index joint. This will be compared between the CR845 and placebo group.

### 5.1.2. Secondary Efficacy Variables

The following will be secondary efficacy variables/endpoints and will be compared between the CR845 and placebo group:

- Change from baseline in the Western Ontario & McMaster Osteoarthritis (WOMAC) Index Total Score at each week
- Change from baseline in the WOMAC Pain Intensity, Stiffness and Function Sub-scale Scores at each week
- Treatment Pain Response: Proportion of patients with at least 30% and 50% improvement from baseline in the weekly mean pain intensity at week 8
- Patient Global Impression of Change (PGIC) at Week 8 or End of Treatment visit
- Supplemental Analgesic Medication Used: Average daily number of acetaminophen tablets used over the study and during the Maintenance Treatment Period
- Proportion of patients withdrawing from treatment due to lack of analgesic efficacy

**Western Ontario & McMaster Osteoarthritis (WOMAC) Index**

The WOMAC Index is a self-administered assessment used to measure pain, stiffness, and physical function in patients with osteoarthritis. The WOMAC Index contains 24 questions across 3 different sub-scales (pain, stiffness, and function) all with scoring 0-10 (0 = No Pain/Stiffness/Difficulty, 10 = Extreme Pain/Stiffness/Difficulty). The WOMAC Index Total Score is calculated as the sum of all 24 scores. Each WOMAC Index Subscale score (Pain/Stiffness/Function) is calculated as the sum of all scores within that subscale. Patients will complete the WOMAC at each clinic visit starting with the Baseline Visit.

**Treatment Pain Response**

A patient's pain response to treatment is defined as the percent improvement from baseline with respect to the weekly mean of "average pain in the last 24 hours" pain score during the last week of the Maintenance Treatment Period (Week 8) . If a patient's mean weekly pain score during the last week of the Maintenance Treatment Period is greater than the baseline score (i.e., the patient has an increase in pain compared to baseline), his/her response to treatment will be assigned a value of 0 (i.e. the patient will be considered a non-responder). Patients who discontinue study drug early will be considered non-responders to treatment and will be assigned a pain response of 0.

The percentage of subjects achieving levels of treatment response 10% to 100% by 10% increments as defined above will be calculated with 30% of key interest as it has been shown to represent a clinically important improvement in pain.

**Patient Global Impression of Change (PGIC)**

The PGIC score collected at Week 8, or at the End of Treatment visit (for patients who discontinue treatment early) ranges from 1 (Very Much Improved) to 7 (Very Much Worse).

**Supplemental Analgesic Medication Used**

The average daily number of acetaminophen tablets used during the study will be calculated using data recorded in the patient diary as the sum of the total number of tablets used divided by the length of exposure (in days). A similar algorithm will be used to derive the average daily number of acetaminophen tablets used during each week (baseline will be study day -5 to study day -1), the Titration-to-Effect Period, the Maintenance Treatment Period, Study Day 1 to 56 Period and the End of Treatment week

Due to the nature of the data, the average daily number of analgesic rescue use (in this case acetaminophen) tend to have a large portion of zeroes (contributed by subjects not taking any supplemental pain medication) followed by continuous distribution of non-zero values. Therefore, the supplemental pain medication for OA will be grouped categorically into the following classes: (1) no supplemental acetaminophen tablets, (2) 0 to ≤ 0.5 tablets, (3) > 0.5 to ≤ 1.0 tablets, (4) > 1.0 to ≤ 2.0 tablets, and (5) > 2.0 tablets.

**Withdrawal Due to Lack of Analgesic Efficacy**

This will be obtained from the treatment completion form of the eCRF.

## 5.2. SAFETY VARIABLES

The safety variables are:

- Adverse events
- Clinical laboratory evaluations (including serum sodium levels)
- Vital signs : systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate and body temperature
- ECGs: heart rate, PR interval, QRS duration, QT interval, RR duration, QTcF, and clinical interpretation of ECG test results

### 5.2.1. Adverse Events

Adverse events that occur after the signing of the informed consent up to study completion, including the Follow-up visit 7-10 days after the last dose of study medication, will be recorded on CRFs for all subjects who enter the study. An additional follow-up telephone call may be performed up to 30 days after the last dose as a final assessment of treatment-emergent adverse events unresolved at the final clinic visit attended. Information gathered by telephone related to status of adverse events will be documented in the patient case report form.

### 5.2.2. Clinical Laboratory Evaluations

The following clinical laboratory tests will be performed at Screening, Day 1 (Baseline prior to drug administration) and at Day 57 (or early termination), with additional blood sampling for the valuation of serum sodium as follows:

**Table 2 – Clinical Laboratory Tests**

| | |
|---|---|
| **Hematology:** | Hemoglobin, hematocrit, red blood cell (RBC) count, RBC ndices, mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), platelet count (or estimate), white blood cell (WBC) count including differential. |
| **Serum Chemistry:** | Total bilirubin, ALT, AST, glucose (fasting/non-fasting), serum creatinine, Blood Urea Nitrogen (BUN), electrolytes (sodium, potassium, chloride, and calcium), alkaline phosphatase, gamma glutamyl transferase, and creatinine. |
| **Serum Sodium** | Serum sodium levels will be measured at each clinic visit. f at any time during the study, the patient's serum sodium level s > 150 mmol/L, dosing should be paused and the patient should be evaluated. Based upon the outcome of the clinical evaluation of the patient by the clinical staff, the dosing of study medication will either be: a) restarted or b) terminated. |

The following laboratory tests will also be performed:

- For female patients of childbearing potential, a serum pregnancy test will be conducted at Screening and a urine pregnancy test will be conducted at the Day 1 (Baseline) visit prior to the administration of study drug.
- Urine drug screen at Screening for amphetamines, barbiturates, benzodiazepines, cocaine, opioids, phencyclidine (PCP), and tetrahydrocannabinol (THC).

Blood samples for hematology and serum chemistry will be sent to a central laboratory for analysis. Urine drug screens, pregnancy tests, and sodium levels will be conducted at the study sites.

### 5.2.3. Vital Signs

Vital sign measurements of systolic blood pressure, diastolic blood pressure, heart rate and respiratory rate will be obtained at each clinic visit. Body temperature (T) will only be obtained at Screening and Day 57 (or early termination).

At screening, blood pressure and heart rate will be assessed both in the supine and standing position. For the purpose of these evaluations, patients should be assessed in the supine position first and then, after sitting up for at least one-minute, should be standing for at least 3 minutes prior to measuring blood pressure or heart rate. Otherwise, all vital signs will be collected while the patient is at rest in a seated position.

### 5.2.4. ECGs

An ECG will be performed at Screening and Day 57 (or early termination) and will consist of heart rate, PR duration, QRS duration, QT interval, RR interval, uncorrected QT interval and corrected Fridericia's QT interval. The investigator will interpret the ECG as Normal, Abnormal – Not Clinically Significant or Abnormal – Clinically Significant.

# 6. SAMPLE SIZE DETERMINATION

Approximately 405 patients with OA of the hip and knee will be randomized in a 2:1 ratio (270 to CR845 and 135 to placebo) at approximately 40 sites in the United States.

[redacted]

[redacted]

A sample size re-estimation will be made after approximately 50% of the planned 405 patients have been randomized and have either completed the 8-week treatment period or have discontinued from treatment early (see Section 8.3).

The sample size will be calculated to ensure 90% conditional power such that:

$$t_2 \geq \frac{1.96 + \sqrt{0.5}\, t_1}{\sqrt{0.5}}$$

where $t_1$ and $t_2$ are the t-statistics from separately performing the primary analysis for stage 1 (interim) and stage 2 (post interim) patients respectively (see Section 8). Maintaining a 2:1 ratio, the required total sample size ($n_2$) to achieve the above criteria will be calculated using the difference between least square means and the corresponding standard error divided by sqrt($1/n_{11} + 1/n_{12}$) from stage 1 analysis (where $n_{11}$ and $n_{12}$ are the sample size in each treatment group from stage 1).

A maximum total sample size of 480 randomized patients will be allowed for the study regardless of the sample size re-estimation at the interim analysis.

# 7. ANALYSIS POPULATIONS

Several analysis populations are described below. The analysis of the Full Analysis and Per Protocol populations will be based on planned (randomized) treatment, the analysis of the Safety population will be based on actual treatment received.

## 7.1. ENROLLED POPULATION

The **enrolled population** consists of all individuals who signed the informed consent form.

## 7.2. EFFICACY POPULATIONS

### 7.2.1. Full Analysis Population (FAP)

The Full Analysis population (FAP) consists of all randomized patients who received at least 1 dose of double-blind study drug will be included in the efficacy evaluation.

### 7.2.2. Per Protocol (PP) Population

The Per Protocol population (PP) consists of all randomized patients in the FAP without any major protocol violations that could affect the efficacy analyses. The criteria for determining what constitutes a major protocol violation will be defined prior to database lock and the breaking of the blinded treatment assignment code.

The requirement for a PP population and exclusion of subjects from it will be decided at a blinded data review meeting (BDRM) and decisions will be documented in a separate BDRM document approved and signed by all authors of this SAP (or equivalents) prior to database lock and unblinding of the data.

## 7.3. SAFETY POPULATION

The Safety population (SAF) is the group of subjects who were randomized and received at least one dose of double-blind study drug. The FAP and the Safety populations are identical.

## 7.4. PROTOCOL DEVIATIONS

Protocol deviations will be identified in several ways: through programmatic checks, through medical reviews, and by clinical research associates during site monitoring. The protocol deviations include but are not limited to:

1.) Exclusionary medical history at screening

2.) Violation of sodium criteria at screening or at admission to surgery

3.) Prior, concomitant, or supplemental medication deviations

4.) Violation of inclusion/exclusion criteria related to the pain intensity scores

5.) Other known violation of inclusion/exclusion criteria

6.) Deviations with respect to concomitant analgesic rescue medication

At the BDRM, protocol deviations will be classified as minor or major and whether they would affect efficacy sufficiently to exclude the patient from the PP population.

# 8. GENERAL STATISTICAL CONSIDERATIONS

This SAP contains a detailed description of the analytical variables and the statistical methods/procedures that will be presented in support of this clinical study (CLIN2002-PO). Any modification to the proposed analyses will be documented in the integrated statistical/clinical report. Additionally, any exploratory analyses not described in the Statistical Analysis Plan will be documented in the integrated statistical/clinical report.

The overall objectives of this study will be achieved through descriptive and/or inferential analyses involving the efficacy and safety variables collected in this study. All variables specified in the protocol will be reported in listing format. All such variables (other than those used purely for screening purposes) and resulting change from baseline values will be summarized using descriptive statistics for each treatment group. Selected change from baseline values will be compared between the treatment groups statistically. Within this general framework, attention is called to variables that will be used to evaluate the primary hypothesis, pre-specified secondary hypotheses, and additional hypotheses of special interest.

The primary population for efficacy analysis will be the FAP population. Sensitivity analyses based on the PP population will be conducted for the primary efficacy variable if greater than or equal to 20% of FAP patients are excluded from the PP population. The primary population for safety analysis will be the safety population (see Section 7.0 for definitions)

All hypothesis tests comparing CR845 and placebo will be two-sided and conducted at the 5% significance level, unless otherwise specified.

All summary tables will be presented by treatment group (if applicable) and over all patients in the analysis population being summarized. Continuous variables will be summarized using the following descriptive statistics: n, mean, SD, median, minimum, and maximum. Categorical variables will be summarized using counts and percentages. Unless specified otherwise, percentages presented will be calculated using all patients in the analysis population being summarized (with randomized or received treatment as/if appropriate) as the denominator.

SAS® Version 9.2 or later will be used to produce all statistical tables, listings and figures to be contained in Section 14 of the integrated statistical/clinical report.

## 8.1. ADJUSTMENTS FOR COVARIATES

All statistical analyses of efficacy variables comparing CR845 versus placebo will be adjusted by the location of the primary OA joint (hip vs knee) unless otherwise specified.

Statistical models without repeated measures which compare the change from baseline endpoints will additionally be adjusted for the value observed at baseline prior to first dose of treatment and effect of the study center.

Mixed Models with Repeated Measures (MMRM) comparing results which are included across all time-points will additionally be adjusted for the baseline value, week, treatment by week interaction and the effect of the patient.

## 8.2. HANDLING OF INCOMPLETE OR MISSING DATA

No missing or incomplete safety data will be imputed.

Weekly Mean Pain Intensity Scores
If a patient reports at least 3 days with non-missing pain intensity scores within a particular week, the weekly mean pain intensity score will be derived using the non-missing scores only. This is identical to imputing the missing scores as the mean of the non-missing scores. If the patient reports less than 3 days with non-missing pain intensity scores within a particular week, the weekly mean pain intensity score will be treated as missing data.

Missing weekly mean pain intensity scores will be imputed using multiple missing data and pattern mixture methodology as detailed in Section 10.1. Sensitivity analyses will be conducted exploring different patterns for missing data. An MMRM will be performed using only observed data without multiple imputation and an analysis using the change from baseline to the last non-missing weekly mean pain intensity score will also be provided.

WOMAC Index Total Score and WOMAC Pain Intensity, Stiffness and Function Sub-Scale Scores
If at least 50% of items within a subscale are non-missing, the missing items of the WOMAC score at a weekly visit will be imputed using the mean of non-missing items within the same subscale. Otherwise, the item, the subscale score and the total score will all remain missing. The total and subscale scores will be derived as the sum of all respective observed and imputed scores. If all values of a particular subscale are missing at a particular week then the total and corresponding subscale score will be treated as missing.

An MMRM will be performed using only observed data for WOMAC related endpoints.

Treatment Pain Response
Patients with a missing weekly mean pain intensity score for Week 8 will be assumed to be a non-responder and assigned a pain response of 0%.

Rescue Medication
If the patient recorded that they did not confirm rescue medication was taken on a particular day, yet the number of tablets was recorded, it will be assumed rescue medication was taken.

All other efficacy endpoints
Missing data will not be imputed.

[redacted]

[redacted]

[redacted]

## 8.4. Multi-center Studies

There is no plan to include study center in the primary analysis because the MMRM will already include a random effect for patient. A model with random patient effects nested within study center would be unwieldy due to expected small numbers of patients per study center and large number of centers.

However, to satisfy ICH guidance (E9 Section 3.2), the difference between treatment means in the change from baseline with respect to the weekly mean of the daily 24-hour pain intensity for the index joint for each center will be plotted by week (observed data only). Only centers with at least 2 patients per treatment group will be included in this plot.

## 8.5. Multiple Comparisons / Multiplicity

One primary efficacy endpoint has been specified in the protocol and the study has been powered sufficiently to show statistical significance in this endpoint. However, as this study is a Phase 2 study, there will be no adjustment for multiplicity for secondary endpoints. The results for the primary efficacy analysis will be the key result to aid planning of future Phase 3 studies. Results from other endpoints may be used to aid the generation of future hypothesis to be explored and/or warrant inclusion of such endpoints in future studies.

## 8.6. Examination of Subgroups

Summaries relating to the Pain Intensity and WOMAC endpoints will be presented by location of the primary OA joint.

# 9. STUDY POPULATION CHARACTERISTICS

Unless specified otherwise, all summaries of study population characteristics will be summarized for the FAP. Since the SAF and FAP are equivalent populations, an additional summary table for patients in the SAP will only be produced if any patients are mis-randomized.

## 9.1. SUBJECT DISPOSITION

A complete accounting of patient participation in the study will display the number of patients that were enrolled (i.e. signed informed consent) and the number and percentage of patients that were:

- Screen failures (Enrolled population)
- Randomized (Enrolled population)
- Randomized following re-screening (Enrolled population)
- Dosed with study drug (Enrolled population)
- Included in each population (FAP, SAF, PP)
- Completed the 8-week treatment with study drug
- Discontinued the study drug early with reasons
- Completed the follow-up visit

In addition, the number and percentage of patients completing each week of the Titration to Effect Period and completing the Maintenance Treatment periods, as well as the assigned doses across each week/period will be summarized.

Listings including the patient disposition, status with respect to inclusion and exclusion criteria, and population eligibility will be presented.

## 9.2. PROTOCOL DEVIATIONS

Major protocol deviations and deviations that may affect the analysis of efficacy will be summarized for all patients. All protocol deviations will be listed, and sorted by patient number and date of deviation (if available). This by-patient listing will identify whether a specific protocol deviation is major or minor, and whether it is considered to potentially affect the analysis of efficacy.

## 9.3. DEMOGRAPHICS AND BASELINE CHARACTERISTICS

Demographic and baseline characteristics will be summarized based on the information provided during the screening visit. The demographic characteristics to be summarized are

listed below with details of whether they will be summarized as a categorical (with categories if not direct from the eCRF) or continuous variable:

- Age at study entry – continuous
- Age at study entry – categorical
  - 25 to 44; 45 to 64; 65 to 74; 75+
- Gender – categorical
- Race – categorical
- Ethnicity - categorical
- Weight (kg) – continuous
- Height (cm) – continuous
- Body Mass Index (BMI) ($kg/m^2$) – continuous
- Duration of time with OA – continuous
- Location of primary OA - categorical

Age at study entry will be calculated as (date the informed consent was signed – date of birth) / 365.25 rounded down to the nearest integer.

The supportive data for the summary of demographic and baseline characterisitcs will be listed, and sorted by patient number.

## 9.4. MEDICAL HISTORY

The medical history reported on the eCRF will be coded to Medical Dictionary for Regulatory Activities Terminology (MedDRA) terminology. Coded medical history terms will be summarized by SOC and Preferred Term. Medical history data will be listed including the verbatim investigator description of the relevant history, the coded terms, the start/stop dates, and whether the history is ongoing, if not stop date is provided.

A separate listing will be created with all the distinct levels of SOC, Preferred Terms and the verbatim investigator description reported in the study. Sorting will be alphatically by SOC, Preferred Term and then verbatim description.

## 9.5. DOSING AND EXTENT OF EXPOSURE

Treatment compliance (%) will be calculated as follows:

$$100 * \frac{\text{Number of doses actually taken}}{\text{Expected number of doses to have been taken}}$$

During the Titration to Effect and Maintenance Treatment periods as well as over the study, it is most likely the first dose will be a PM dose (following a clinic visit) and the last dose will be an AM dose (prior to a clinic visit). Therefore, the expected number of doses over a particular period will be calculated as [end date– start date] * 2.

The number of doses actually taken will be compared to the expected number of doses to derive the number of missed doses. There will be monitoring by the clinic staff to ensure that the diary dosing coincides with the drug accountability at each visit.

The number of missed doses (0, 1, ≥ 2) and treatment compliance (continuous and as categories <80%, 80% to 110%, >110%) will be summarized for the Titration to Effect, Maintenance Treatment and overall study period. Additionally, the length of exposure to study drug (continuous and as overlapping categories: any exposure, ≥ 1 week, ≥ 2 weeks, ≥ 4 weeks, ≥ 8 weeks), maximum titration level (1mg, 2,5mg and 5mg) and average dose taken over the study will be summarized.

## 9.6. PRIOR AND CONCOMITANT MEDICATION AND THERAPIES

Prior medications are defined as medications taken at any time during the 30 days prior to Screening until the time of dosing on Day 1, whether continuing or not. Any changes to prior medications (dose, regimen, etc.) during the study will be considered a new *Concomitant Medication*.

*Concomitant medications* are defined as any medications taken following the administration of the first dose of study drug on Day 1 through the end of the study (i.e., at the Follow-up Visit). These will be categorized into two groups, 1) During treatment with study drug and 2) Following end of treatment with study drug.

Whether a medication is prior, concomitant or both will be assessed by comparing the start date of the medication to the date of the first dose. If the medication start and/or end date is fully missing or sufficiently partially missing such that it cannot be determined whether it was taken prior to the first dose of study drug, then the medication will be considered concomitant. Similarly, if the medication start and/or end date is fully missing or sufficiently partially missing such that it cannot be determined whether it was taken following the last dose of study drug then it will be classed as being concomitant during treatment with study drug.

Prior and Concomitant medications (during and after treatment) will be coded using the Anatomical-Therapeutic-Chemical (ATC) classification text from the World Health Organization Drug Dictionary (WHO DD) and summarized seperately. The number and percentage of patients taking a medication will be displayed by ATC Class 1 and Preferred Term. Patients will only be counted one time in each unique ATC Class 1 and Preferred Term.

All prior and concomitant medications will be listed and will include both the medication name as reported by the investigator as well as the coded information.

A separate listing will be created with all the distinct levels of ATC Class 1, Preferred Terms and the verbatim investigator description reported in the study. Sorting will be by ATC Class 1, Preferred Term and then verbatim description.

# 10. EFFICACY ANALYSES

All efficacy analyses will be conducted on the FAP and PP population if neccesary (see Section 8).

In addition to the statistical analyses described in the sections below, descriptive summaries (as per Section 8) and data listings sorted by unique patient number will be presented for all efficacy variables.

## 10.1. PRIMARY EFFICACY VARIABLE

To account for the interim unblinded sample size re-assessment, the analyses relating to the primary efficacy variable will employ weighted test statistics as proposed by Cui, Hung and Wang in order to maintain the 5% two-sided type 1 error rate. As there is no potential for early stopping of the study, no adjustments need to be made to the significance levels for the combined final analysis. The primary analysis and sensitivity analyses described in this section will be performed on each stage (pre and post interim patients) seperately and test statistics ($t_1$ and $t_2$) from each stage will be combined to obtain the CHW Z statistic as follows:

$Z_{CHW} = \sqrt{0.5}\, t_1 + \sqrt{0.5}\, t_2$ from which the two-sided p-value will be obtained using the normal distribution.

Point estimates and 95% CI will be obtained using methodology suggested by Hung and Lawrence[2]:

Point estimate:

$$\hat{\delta} = \frac{t_1 \hat{\delta}^{(1)} + \sqrt{1 - t_1}\,\sqrt{t^* - t_1}\,\hat{\delta}^{(2)}}{t_1 + \sqrt{1 - t_1}\,\sqrt{t^* - t_1}}$$

where
$t_1 = 0.5$
$t^*$ = (Re-estimated total sample size)/405
$\hat{\delta}^{(1)}$ = Least square mean difference from stage 1 analysis
$\hat{\delta}^{(2)}$ = Least square mean difference from stage 2 analysis

$$\textbf{95\% CI} = \hat{\delta} \pm \left( \frac{\hat{\delta}}{Z_{CHW}} * 1.96 \right)$$

The primary analysis and each sensitivity analysis will also be performed using the full combined dataset with no CHW adjustment. Presentation of results will start with these analyses (additionally including the CHW results) and then follow with the results of the analysis from each individual stage.

The primary efficacy variable will be analyzed using a mixed effect model with repeated measures (MMRM). The model will contain treatment (placebo vs CR845), treatment-by-week interaction as fixed effects, baseline value and OA joint (hip vs knee) as covariates, and patient as a random effect.

The treatment difference between CR845 (across all doses) and placebo at Week 8 (the primary efficacy endpoint) will be estimated as the simple contrast in the treatment effect. The 2-sided 95% confidence interval (CI), based on the difference in the least squares (LS) means between the 2 treatment groups (CR845 vs placebo) will be presented. The LS means (and 95% CI) will be additionally be presented for each treatment for each week.

The MMRM model will be implemented using SAS PROC MIXED. A restricted maximum likelihood (REML) will be used. The Kenward-Roger[3] approximation will be used to estimate the denominator degrees of freedom and adjust standard errors. An appropriate covariance matrix will be used to model the within patient errors until there are no convergence issues in the following order: 1) unstructured, 2) Toeplitz, 3) autoregressive [1], 4) compound symmetry, 5) variance components. For analyses of multiple imputed datasets, this selection will continue until analyses for all imputations converge.

The treatment of missing data will be addressed using pattern mixture model methodology. Within each treatment, 4 data patterns will be defined as follows:

- Pattern 1: will include patients who completed treatment on study drug
- Pattern 2: will include patients who discontinue treatment early due to AEs
- Pattern 3: will include patients who discontinue treatment early due to lack of therapeutic efficacy (LOTE)
- Pattern 4: will include patients who discontinue treatment early due to reasons other than AE event or LOTE.

The pattern mixture model methodology will be implemented using SAS procedures MI and MIANALYZE. Specifically, the following steps will be followed:

- Intermittent missing data will first be imputed using the Markov Chain Monte Carlo (MCMC) method implemented with the SAS MI procedure, which is appropriate for non-monotonic missing data.
- Data for patients who discontinues early will be multiply imputed as follows:
    - If the patient discontinues due to AE (pattern 2) or LOTE (pattern 3), then missing data will be assumed to follow a distribution similar to the baseline values observed in the patient's randomized treatment.
    - If the patient discontinues due to reasons other than AE or LOTE (pattern 4), at each time point, missing data will be assumed to follow a distribution similar to scores for patients that are still in the study and randomized to the same treatment group.
- Results of the MMRM on the multiply imputed data will be summarized using the SAS MIANALYZE procedure using Rubin's rule.

A detailed plan of the above procedure is included as an attachment in Section 16.1.

Three sensitivity analyses (described below) will be conducted to assess the impact of the treatment of missing data on the study results. These sensitivity analyses may not necessarily reach statistical significance at the 5% level. However, a consistency of treatment effect (similar estimate of CR845 mean relative to placebo) is expected across the primary and all sensitivity analyses.

The MMRM analysis will be applied using observed data only to provide a reference in addition to the following sensitivity analyses:

**Sensitivity analysis 1 (Multiple Imputation Missing at Random[4])**
This sensitivity analysis assumes that patients with missing data follow the same trajectory as other patients in their respective treatment arm that have complete data.

- Intermittent missing data will first be imputed using the MCMC method implemented with the SAS MI procedure, which is appropriate for non-monotonic missing data.
- Data missing after patients discontinue treatment early will then be multiply imputed with the SAS MI procedure using a method appropriate for monotone missingness (e.g., regression statement). At each time point, missing data will be assumed to follow a distribution similar to scores for patients who are still in the study and randomized to the same treatment group.
- Results of the MMRM on the multiply imputed data sets will be summarized by the SAS MIANALYZE procedure.

**Sensitivity analysis 2 (Multiple Imputation, Missing not at random, placebo-based imputation[4])**
This sensitivity analysis assumes that patients in the active treatment group who discontinue early will have the same evolution, after treatment is stopped, as patients in the placebo group (who are not exposed to treatment by definition). Patients who discontinue treatment early in the placebo group are assumed to behave as the placebo patients that remain in the study.

- Intermittent missing data will first be imputed using the MCMC method implemented with the SAS MI procedure, which is appropriate for non-monotonic missing data.
- Data missing after patients discontinue treatment early will be multiply imputed using multiple calls of the SAS MI procedure. At each time point, missing data will be imputed using data from patients in the placebo group that have complete data at that time.
- Results of the MMRM on the multiply imputed data sets will be summarized by the SAS MIANALYZE procedure.

**Sensitivity analysis 3 (Single Imputation, MNAR, Hybrid BOCF /LOCF):**
This sensitivity analysis makes different assumptions about the pain trajectory of patients who discontinue study drug early depending on the reason for discontinuation. For patients who

discontinue due to AE, Baseline Observation Carried Forward (BOCF) will be used. For patients who discontinue due to reasons other than AE, Last Observation Carried Forward (LOCF) will be used. It is plausible that patients who discontinue due to LOTE experience worsening pain scores after randomization; in this case, their last observed pain scores would be higher (worse) than their baseline pain scores. This sensitivity analysis assumes that missing data for these patients is similar to observed pain scores for patients who discontinue due to LOTE but are still in the study.

## 10.2. SECONDARY EFFICACY VARIABLE(S)

**Change from baseline in the Western Ontario & McMaster Osteoarthritis (WOMAC) Index Total Score at each week**

The change from baseline in the WOMAC Index total score will be analyzed using an MMRM with treatment (placebo vs CR845) and treatment-by-week interaction as fixed effects, baseline value and OA joint (hip vs knee) as covariates, and patient as a random effect. The primary comparison between groups will be based on estimates and contrasts for the weeks 4, 6, and 8 means. The comparison at week 8 will be viewed as secondary. A 95% CI will be constructed for the weeks 4, 6 and 8 mean differences between CR845 and placebo based on the MMRM.

The MMRM model (based on observed data) will be implemented using SAS PROC MIXED. REML will be used. An appropriate covariance matrix will be used to model the within patient errors. An appropriate covariance matrix will be used to model the within patient errors until there are no convergence issues in the following order: 1) unstructured, 2) Toeplitz, 3) autoregressive [1], 4) compound symmetry, 5) variance components. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom and adjust standard errors.

**Change from baseline in the WOMAC Pain Intensity, Stiffness and Function Sub-scale Scores at each week**

The change from baseline in the WOMAC sub-scales for pain intensity, stiffness and function will be analyzed using an MMRM model similar to the one presented above for the WOMAC index total score.

**Treatment Pain Response: Proportion of patients with a X0% improvement from baseline in the weekly mean pain intensity at week 8**

The proportion of patients with a response to treatment that is ≥ 30% will be analyzed using logistic regression, including treatment, baseline weekly pain score, and OA joint as independent variables. Site will be included as a random effect. Additionally, the proportions will be compared using a Cochran-Mantel-Haenzel test stratified by OA joint (hip or knee) and

dichotomized baseline weekly pain score (<6.7, ≥6.7). These analyses will be repeated for patients who have a response to treatment ≥ 50%.

The cumulative proportion of patients that achieve at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100% response to treatment will be summarized and plotted for each treatment.

**Patient Global Impression of Change (PGIC) at Week 8 or End of Treatment visit**

The PGIC will be summarized as both a categorical and continuous variable (including 95% CI for the mean). The proportion of patients with PGIC scores of "Very Much Improved" and "Much Improved" will also be summarized and compared between treatments using a logistic regression with treatment as a main effect and OA joint as a covariate. Additionally, this proportion will be compared using a Cochran-Mantel-Haenzel test stratified by OA joint (hip or knee) and dichotomized baseline pain score (<6.7, ≥6.7).

**Supplemental Analgesic Medication Used: Average daily number of acetaminophen tablets used over the study and during the Maintenance Treatment Period**

The average daily number of acetaminophen tablets used during the study and during the Maintenance Treatment Period will be compared between treatments using a Wilcoxon Rank Sum Test stratified by Primary OA Joint (Hip or Knee) and dichotomized baseline pain score (<6.7, ≥6.7) (van Elteren test[4]). The categorized groupings will also be summarized.

**Proportion of patients withdrawing from treatment due to lack of analgesic efficacy**

The proportion of patients withdrawing from treatment due to lack of analgesic efficacy will be analyzed using logistic regression, including treatment, baseline pain score, and OA joint as independent variables. Additionally, the proportions will be compared using a Cochran-Mantel-Haenzel test stratified by OA joint (hip or knee) and dichotomized baseline pain score (<6.7, ≥6.7).

# 11. SAFETY ANALYSES

Evaluation of safety will be performed for all patients in the SAF population.

## 11.1. ADVERSE EVENTS

Adverse events will be coded using MedDRA. Verbatim description of the adverse events and the MedDRA System Organ Class (SOC) and Preferred Term (PT) for all AEs will be contained in the patient data listings. A separate listing sorted by MedDRA SOC and PT will include all verbatim descriptions associated with each SOC/PT category.

Treatment emergent AEs (TEAEs) will be defined as AEs where onset occurs from the day of first dose of study medication up to and including the 2nd day following the day of last dose of study medication.

All reported AEs (whether treatment-emergent or not) will be included in by-subject AE listings. A separate listing will be created with all the distinct levels of SOC, Preferred Terms and the verbatim investigator description reported in the study. Sorting will be alphabetically by SOC, Preferred Term and then verbatim description.

If the adverse event start date is completely/partially missing, the AE will be considered a TEAE unless the following can be determined:

- The partial AE start date is sufficient to show that the AE started prior to the day of first dose of study drug.
- The AE stop date shows that the AE ended prior to the day of first dose of study drug.
- The partial AE start date is sufficient to show that the AE started after the 2 days following the day of last dose of study drug.

The incidence of TEAEs will be presented using counts and percentages of patients with AEs and tabulated by SOC and PT. SOC will be sorted alphabetically and PT within SOC will be sorted by descending frequency based on the incidence across subjects overall. If a patient has multiple occurrences (start and stop) of an event associated with a specific SOC or PT within a SOC, a patient will only be counted once in the incidence count for the SOC or PT within SOC respectively.

The following specific summary tables will be generated:

- Overall incidence of adverse events
- Incidence of TEAEs by SOC overall and by PT within each SOC
- Incidence of TEAEs related to Treatment by SOC overall and by PT within each SOC
    - If a patient reports two or more adverse events that code to the same PT, the event with the maximum relationship will be included in the table. TEAEs with a missing relationship will be assumed to be treatment related.
- Incidence of TEAEs by Maximum Severity, SOC overall and by PT within each SOC
    - If a patient reports two or more adverse events that code to the same PT, the event with the maximum severity will be included in the table. TEAEs with a missing severity will be assumed to be severe.

- Incidence of TEAEs occurring in ≥ 5% of patients in at least one treatment group by SOC overall and by PT within each SOC
- Incidence of treatment-related TEAEs occurring in ≥ 5% of patients in at least one treatment group by SOC overall and by PT within each SOC
- Incidence of Serious TEAEs by SOC overall and by PT within each SOC
- Incidence of TEAEs leading to treatment discontinuation by SOC overall and by PT within each SOC
- Incidence of non TEAEs following last dose of study drug for patients taking opioids by SOC overall and by PT within each SOC

Deaths, other SAEs, and AEs leading to treatment discontinuation, dose interruption, or dose reduction will be listed including the treatment group and dosing information (if applicable), start and stop dates/times of the AE, and days on study relative to the day of first dose of study drug.

The following listings will be provided:

- Listing of subjects who died
- Listing of subjects with Serious TEAEs
- Listing of subjects with non treatment-emergent SAEs
- Listing of subjects with TEAEs leading to study drug discontinuation, dose reduction, or dose interruption

## 11.2. Clinical Laboratory Evaluations

Chemistry and Hematology Laboratory Data

Laboratory tests from unscheduled study assessments outside planned protocol windows will not be included in the summary tables, except in the derivation of the laboratory test value at Baseline. If multiple assessments are recorded within a planned protocol window, the closest assessment to the planned time point will be used. If there are two assessments within equal proximity either side of the planned time point the earlier assessment will be used.

All laboratory tests results, whether scheduled or unscheduled, will be included in the by-subject listings. Laboratory values will be reported in standard Système International (SI) units

For all laboratory listings and summaries described in this section, the following definitions will be applied to each laboratory test:

- Screening value = the last scheduled or unscheduled (repeat) test result obtained prior to, but not including Day 1
- Baseline value = the last scheduled or unscheduled (repeat) test result obtained prior to the first dose of study drug
- End of treatment = the last scheduled test result (Day 57 or early termination) obtained within 48 hours from the last dose of study drug.

Chemistry and Hematology laboratory results at Day 57 (or early termination), and the corresponding change from baseline will be summarized for the safety population by presenting summary statistics (n, mean, SD, median, and range).

Laboratory test results will be assigned an L/N/H classification according to whether the value was below (L), within (N), or above (H) the laboratory parameter's reference range. Within each treatment arm, a 3 by 3 tables shift tables will compare the Baseline LNH classification of each laboratory test to the LNH classification at Day 57 (or early termination).

Selected laboratory test results will be flagged as markedly abnormal based on the ranges defined in Table 4 below. A summary table of the incidence of markedly abnormally low and markedly abnormally high laboratory tests results will be presented by treatment arm for all patients in the safety population. For these calculations, each subject may be counted once in the laboratory parameter value high and in the laboratory parameter low categories as applicable.

## Table 4. Laboratory Ranges Used to Identify Markedly Abnormal Laboratory Values

| | Units | |
|---|---|---|
| **Laboratory Parameter** | **Lower Limit[a]** | **Upper Limit[a]** |
| **Hematology** | | |
| Hemoglobin | <10 g/dL or <100 g/L | —— |
| Platelets | $<75.0 \times 10^9$/L or $<75000/mm^3$ | —— |
| Leukocytes | $<3.0 \times 10^9$/L or $<3000/mm^3$ | —— |
| Leukocytes - infants, young children | <75% LLN [b] | —— |
| Lymphocytes | <LLN | —— |
| Lymphocytes - infants, young children | <75% LLN | —— |
| Neutrophils | $<1500 \times 10^6$/L or $<1500/mm^3$ | —— |
| **Clinical Chemistry** | | |
| Electrolytes | | |
| Potassium | < LLN | >5.5 mmol/L |
| Bicarbonate ($CO_2$) | ≤15 mEq/dL or ≤15 mmol/L | —— |
| Liver Function Tests | | |
| Alkaline phosphatase | —— | >3 × ULN [b,c] |
| AST | —— | >3 × ULN [c] |
| ALT | —— | >3 × ULN [c] |
| Total bilirubin | —— | >1.5 × ULN |
| Renal Function Tests | | |
| Creatinine | —— | >1.5 × ULN |
| Other Chemistry | | |
| Calcium | <8 mg/dL or <2.0 mmol/L | >11.5 mg/dL or >2.9 mmol/L |
| Phosphorous | <2.5 mg/dL or <0.8 mmol/L | — |
| Glucose | <55 mg/dL or <3.0 mmol/L | >160 mg/dL or >8.9 mmol/L |
| Uric acid | —— | > ULN |
| Cholesterol | —— | >300 mg/dL or >7.75 mmol/L |
| Triglycerides | —— | >2.5 × ULN |
| Albumin | <3 g/dL or <30 g/L | —— |

[a] National Cancer Institute Common Toxicity Criteria, Version 2, Grade 2 limits (Grade 1 limit if Grade 2 limit missing).
[b] LLN = lower limit of the laboratory reference (normal) range; ULN = upper limit of the laboratory reference (normal) range.
[c] Modified Version 2 of the National Cancer Institute Common Toxicity Criteria.

A patient listing, sorted by patient identifier, will present all laboratory results (scheduled or unscheduled) and will include flags to identify whether a result is outside of the normal range and whether it is markedly abnormal.

A listing of all subjects with at least one markedly abnormal laboratory value will be prepared; for a given patient, this listing will not only include the laboratory results flagged as markedly abnormal, but will also present the results from all other scheduled and unscheduled evaluations of the same laboratory parameter, whether or not they are markedly abnormal.

A separate listing of patients that have liver function test results for AST or ALT > 3 ULN or total bilirubin ≥ 1.5 ULN will be presented.

Additional analyses of liver function tests (AST, ALT, and total bilirubin) will be conducted by generating shifts tables from baseline to Day 57 (or early termination) using the categories presented in Table 5 below.

**Table 5. Categories for Shift Analysis of AST, ALT, and Total Bilirubin**

| AST/ALT | Total Bilirubin |
|---|---|
| ≤ 1 x ULN | 1 x ULN |
| > 1 - ≤ 3 x ULN | > 1 – ≤ 1.5 x ULN |
| >3 - ≤ 5 x ULN | > 1.5 – ≤ 2 x ULN |
| >5 - ≤ 10 x ULN | > 2 x ULN |
| >10 - ≤ 20 x ULN | |
| > 20 x ULN | |

Sodium Laboratory Tests

Sodium values (absolute and change from baseline as appropriate) will be summarized at each timepoint (Screening, Baseline [Day 1], Week 1 [Day 8], Week 2 [Day 15], Week 3 [Day 22], Week 4 [Day 29], Week 6 [Day 43], Week 8 [Day 57] and at Day 57 [or early termination]).

The number of patients who meet the important sodium levels (> 146 mmol/L or > 150 mmol/L) anytime between first dose and within 72 hours of final dose of study drug) will be summarized. A patient listing including all sodium test results will be programmed, flagging values that are > 146 mmol/L and > 150 mmol/L. In addition, a listing of patients with at least one sodium result > 150 mmol/L (whether during or after treatment with study drug) will be prepared; this listing will include all sodium test results for the selected patients.

## 11.3. VITAL SIGNS

Vital sign measurements of Systolic Blood Pressure, Diastolic Blood Pressure, Heart Rate and Respiratory Rate (absolute and change from baseline as appropriate) will be summarized at

each timepoint (Screening [supine], Baseline [Day 1], Week 1 [Day 8], Week 2 [Day 15], Week 3 [Day 22], Week 4 [Day 29], Week 6 [Day 43], Week 8 [Day 57] and at Day 57 [or early termination]). Temperature will be summarized in similar fashion at Screening and Day 57 (or early termination).

For vital signs, abnormal values will be identified as those outside (above or below) the reference ranges that are defined in Table 6.

**Table 6. Reference Range for Vital Sign Values**

| Vital Sign Parameter | Range |
|---|---|
| Systolic blood pressure | 100–140 mmHg |
| Diastolic blood pressure | 60–90 mmHg |
| Respiratory rate | 12–20 breaths per minute |
| Heart rate | 60-100 beats per minute |
| Temperature | 36.8°C–38.3°C |

All vital sign values collected after the Baseline dose of study drug will be evaluated for clinically notable results using the criteria presented in Table 7. A listing of all subjects with at least one clinically notable vital sign value will be prepared; for a given subject, this listing will not only include flagged vital sign parameters, but will present the results of all scheduled and unscheduled vital sign evaluations for the patient. A summary table of the incidence of clinically notable vital signs may be presented.

**Table 7. Criteria Used to Identify Clinically Notable Vital Sign Abnormalities**

| Vital Sign Parameter | Value | Change From Baseline[a] |
|---|---|---|
| Systolic blood pressure | ≥180 mm Hg | Increase of ≥20 mm Hg |
| | ≤90 mm Hg | Decrease of ≥20 mm Hg |
| Diastolic blood pressure | ≥105 mm Hg | Increase of ≥15 mm Hg |
| | ≤50 mm Hg | Decrease of ≥15 mm Hg |
| Heart rate | ≥120 bpm | Increase of ≥15 bpm |
| | ≤50 bpm | Decrease of ≥15 bpm |
| Respiratory rate | <12 breaths per minute | — |
| | > 20 breaths per minute | — |

[a] Both value and change from baseline criteria must be met to qualify as a clinically notable abnormality.

## 11.4. ECGs

Continuous ECG parameters (absolute and change from baseline as appropriate) will be summarized at screening and Day 57 (or early termination). A by-patient listing will also be

prepared, with a separate listing for those patients with a clinically significant abnormal investigator interpretation of ECG results.

## 12. SUMMARY OF CHANGES FROM PROTOCOL-SPECIFIED ANALYSES

At the time of finalizing this SAP, only the following update has been made in this SAP compared to the Study Protocol V3.0 4th November 2016:

**Update 1**

Protocol Section 11.7.1 Text:
The baseline score will be calculated as the average of the 4 pain scores collected during the Screening period that qualified the patient for randomization

SAP Section 5.1.1 Text:
The weekly mean pain intensity score will be defined as the sum of non-missing daily "average pain over the last 24 hours" scores reported during that week (Days -7 to -1 [Baseline]…. divided by the number of days with non-missing scores for that week.

Reason for update
To ensure the derived baseline derivation of the weekly pain score was equivalent in terms of number of days assessed (i.e. 7 day period using any day with a recorded pain score).

## 13. REPORTING CONVENTIONS

P-values will be reported to four decimal places; p-values less than 0.0001 will be reported as p<0.0001. The mean and median will be displayed to one decimal place greater than the original value and the measure of variability will be displayed to two decimal places greater than the original value. All statistical programming and analyses will be performed using SAS® (SAS institute, Cary, NC 27513).

The following standards will be used in the data presentation:

- In text tables should be in portrait format, font type would be “Arial” and font size 9 and RTF format. Section 14 tables and figures, and Appendix 16.2 listings should be in landscape format. Output should adhere to US / ICH margins and should not require changes for European page size. For section 14 tables, a blank row will separate the header from the content of the table listing. For tables that have “n (%) only”, the placement should be centered below “N=xx” in the column header. Frequency tables will be center justified, and descriptive statistics decimal justified. ICH margins are Top=0.75”, Bottom=0.75”, Left=1”, Right=0.75”, Gutter=0”, Header=0.6”, Footer=0.6”, Width=8.27” and Height=11”.
- Percentages presented in the in-text tables should be rounded to whole numbers using the SAS rounding function. Percentages presented in Section 14 tables should be rounded to a

single decimal point (xx.x%) using the SAS rounding function. Percentages for values in tables that are <1 should be represented as "<1". If "%" is part of the column heading, do not repeat the "%" sign in the body of the table. Unless specified otherwise, "%" should reflect the total population of the treatment group(s). Any deviation from that should be part of the footnote. For 0 counts, leave the corresponding percentage blank.

- The format for minimum and maximum should be "Min, Max". Standard deviation should be the default for representing scale for demographic and baseline characteristics. For tables presenting statistical analyses, the SD will be used unless the standard error has been specified. Standard deviation should be abbreviated as "SD", and presented within parenthesis next to the mean value, without any +/- sign. The standard deviation should have one additional decimal place beyond that of the mean (eg, mean has one decimal place, SD should have two).
- "N" will represent the entire treatment group, while "n" will represent a subset of the treatment group. For tables with population designated as a row heading, "N" should be used (ie, tables where all participant data is not available for every variable within a treatment group). As a guideline, if the number is used in a denominator it should be presented as "N". If the number is used in the numerator, it should be presented as an "n".
- The heading should consist of at least 3 lines. Line 1: Table/Appendix number. Line 2: Protocol identifier / Table Title – Adjunct Description of Table Title (optional). Line 3: Population. The title for in-text tables should begin with the Table/Appendix number.
- A solid line should appear both above and below the column headings of a table. A solid line should appear at the end of the table or at the bottom of each page if the table extends to more than one page. Footnotes should start after the bottom solid line.
- Footnotes should be left-justified at the bottom of the page. Footnotes must be present on the page where they are first referenced. If more than four footnote lines are planned then a cover page may be used to display footnotes.
- The maximum length of a title is 255 characters (including spaces).
- Legends will be used for all figures with more than one variable or tiem displayed.
- Figures should be generated in black and white. Lines should be wide enough to see the line after the figure has been fotocopied.
- All tables, figures, and data listings will have the name of the program and date/time stamp on the bottom of each output.

All statistics will be validated with two independent programming codes. All tables, figures and listings will be validated by a two-person review of the output.

# 14. REFERENCES

1. Cui L, Hung HM, Wang SJ. (1999), “Modification of sample size in group sequential clinical trials”. *Biometrics*. 55:853-857.
2. Lawrence J, Hung H.M. (2003), “Estimation and Confidence after Adjusting the Maximum Information”. Biometrical Journal 45, 143–152
3. Kenward, M. G. and Roger, J. H. (1997), “Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood,” *Biometrics*, 53, 983–997.
4. Ratitch B., and O’Kelly M., “Implementation of pattern-mixture models using standard SAS/STAT procedures”, SUGI 2011.

# 15. TOCS FOR TABLES, FIGURES, LISTINGS AND OUTPUT

## 15.1. TABLES

Mock tables are presented in a separate document. The current version based on this SAP are detailed in CLIN2002-PO SAP Table Shells V1_0 18JAN17 and detailed below:

Table 14.1.1.1 Patient Disposition – Screening and Randomization Enrolled Population

Table 14.1.1.2 Patient Disposition Full Analysis Population (FAP)

Table 14.1.1.3 Patient Disposition – Weekly Dose Full Analysis Population (FAP)

Table 14.1.1.4 Patient Disposition – Analysis Populations Full Analysis Population (FAP)

Table 14.1.1.5 Major Protocol Deviations Full Analysis Population (FAP)

Table 14.1.2 Demographic and Baseline Characteristics Full Analysis Population (FAP)

Table 14.1.3 Medical History Full Analysis Population (FAP)

Table 14.1.4 Extent of Exposure to Study Drug Full Analysis Population (FAP)

Table 14.1.5 Treatment Compliance Full Analysis Population (FAP)

Table 14.1.6.1 Prior Medication Full Analysis Population (FAP)

Table 14.1.6.2 Concomitant Medication During Treatment with Study Drug Full Analysis Population (FAP)

Table 14.1.6.3 Concomitant Medication Following End of Treatment with Study Drug Full Analysis Population (FAP)

Table 14.2.1.1.1.1 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) Scores over time
– Descriptive Statistics Over All Patients Full Analysis Population (FAP)

Table 14.2.1.1.1.2 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) Scores over time
– Descriptive Statistics for Stage 1 Full Analysis Population (FAP)

Table 14.2.1.1.1.3 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) Scores over time
– Descriptive Statistics for Stage 2 Full Analysis Population (FAP)

Table 14.2.1.1.2.1 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Primary Efficacy Analysis (MMRM with MNAR Pattern Mixture Modeling) Over All Patients Full Analysis Population (FAP)

Table 14.2.1.1.2.2 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Primary Efficacy Analysis (MMRM with MNAR Pattern Mixture Modeling) for Stage 1 Full Analysis Population (FAP)

Table 14.2.1.1.2.3 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Primary Efficacy Analysis (MMRM with MNAR Pattern Mixture Modeling) for Stage 2 Full Analysis Population (FAP)

Table 14.2.1.1.3.1 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Primary Efficacy Analysis (MMRM with MNAR Pattern Mixture Modeling) Over All Patients Per-Protocol Population (PP)

Table 14.2.1.1.3.2 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Primary Efficacy Analysis (MMRM with MNAR Pattern Mixture Modeling) for Stage 1 Per-Protocol Population (PP)

Table 14.2.1.1.3.3 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Primary Efficacy Analysis (MMRM with MNAR Pattern Mixture Modeling) for Stage 2 Per-Protocol Population (PP)

Table 14.2.1.1.4.1 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 – MMRM Analysis on Observed Data Over All Patients Full Analysis Population (FAP)

Table 14.2.1.1.4.2 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 – MMRM Analysis on Observed Data For Stage 1 Full Analysis Population (FAP)

Table 14.2.1.1.4.3 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 – MMRM Analysis on Observed Data For Stage 2 Full Analysis Population (FAP)

Table 14.2.1.1.5.1 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Sensitivity Analysis 1: MMRM with MAR Pattern Mixture Modeling Over All Patients Full Analysis Population (FAP)

Table 14.2.1.1.5.2 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Sensitivity Analysis 1: MMRM with MAR Pattern Mixture Modeling for Stage 1 Full Analysis Population (FAP)

Table 14.2.1.1.5.3 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Sensitivity Analysis 1: MMRM with MAR Pattern Mixture Modeling for Stage 2 Full Analysis Population (FAP)

Table 14.2.1.1.6.1 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Sensitivity Analysis 2: MMRM with MNAR Placebo Based Pattern Mixture Modeling Over All Patients Full Analysis Population (FAP)

Table 14.2.1.1.6.2 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Sensitivity Analysis 2: MMRM with MNAR Placebo Based Pattern Mixture Modeling For Stage 1 Full Analysis Population (FAP)

Table 14.2.1.1.6.3 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Sensitivity Analysis 2: MMRM with MNAR Placebo Based Pattern Mixture Modeling For Stage 2 Full Analysis Population (FAP)

Table 14.2.1.1.7.1 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Sensitivity Analysis 3: Single Imputation BOCF/LOCF Over All Patients Full Analysis Population (FAP)

Table 14.2.1.1.7.2 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Sensitivity Analysis 3: Single Imputation BOCF/LOCF For Stage 1 Full Analysis Population (FAP)

Table 14.2.1.1.7.3 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) at Week 8 - Sensitivity Analysis 3: Single Imputation BOCF/LOCF For Stage 2 Full Analysis Population (FAP)

Table 14.2.1.1.8 Change from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) – Descriptive Statistics by Primary OA Joint Full Analysis Population (FAP)
Table 14.2.1.2.1 Responder Analysis of Percentage Improvement from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) Scores at Week 8 Full Analysis Population (FAP)
Table 14.2.1.2.2 Responder Analysis of Percentage Improvement from Baseline in the Weekly Mean of the Daily 24 Hour Pain Intensity (NRS) Scores at Week 8 by Primary OA Joint Full Analysis Population (FAP)
Table 14.2.2.1.1 Change from Baseline in the WOMAC Index Score Over Time – Descriptive Statistics Full Analysis Population (FAP)
Table 14.2.2.1.2 Change from Baseline in the WOMAC Index Total Score – MMRM Analysis Full Analysis Population (FAP)
Table 14.2.2.1.3 Change from Baseline in the WOMAC Index Total Score Over Time by Primary OA Joint Full Analysis Population (FAP)
Table 14.2.2.2.1 Change from Baseline in the WOMAC Index Pain Intensity Subscale Score Over Time – Descriptive Statistics Full Analysis Population (FAP)
Table 14.2.2.2.2 Change from Baseline in the WOMAC Index Pain Intensity Subscale Score – MMRM Analysis Full Analysis Population (FAP)
Table 14.2.2.2.3 Change from Baseline in the WOMAC Index Total Score Over Time by Primary OA Joint Full Analysis Population (FAP)
Table 14.2.2.3.1 Change from Baseline in the WOMAC Index Stiffness Intensity Subscale Score Over Time – Descriptive Statistics Full Analysis Population (FAP)
Table 14.2.2.3.2 Change from Baseline in the WOMAC Index Stiffness Intensity Subscale Score – MMRM Analysis Full Analysis Population (FAP)
Table 14.2.2.3.3 Change from Baseline in the WOMAC Index Stiffness Intensity Subscale Score Over Time by Primary OA Joint Full Analysis Population (FAP)
Table 14.2.2.4.1 Change from Baseline in the WOMAC Index Function Intensity Subscale Score Over Time – Descriptive Statistics Full Analysis Population (FAP)
Table 14.2.2.4.2 Change from Baseline in the WOMAC Index Function Intensity Subscale Score – MMRM Analysis Full Analysis Population (FAP)
Table 14.2.2.4.3 Change from Baseline in the WOMAC Index Function Intensity Subscale Score Over Time by Primary OA Joint Full Analysis Population (FAP)
Table 14.2.3 Patient Global Impression of Change (PGIC) at Day 57/ET Full Analysis Population (FAP)
Table 14.2.4 Average Daily Use of Rescue Medication for Target Joint Pain Full Analysis Population (FAP)
Table 14.2.5 Proportion of patients withdrawing from treatment due to lack of analgesic efficacy Full Analysis Population (FAP)
Table 14.3.1.1 Overall Summary of Adverse events
Safety Population (SAF)
Table 14.3.1.2 Summary of Treatment Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term Safety Population (SAF)

Table 14.3.1.3 Summary of Treatment-Related Treatment Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term Safety Population (SAF)
Table 14.3.1.4 Summary of Treatment Emergent Adverse Events (TEAEs) by System Organ Class, Preferred Term and Maximum Severity Safety Population (SAF)
Table 14.3.1.5 Summary of Treatment Emergent Adverse Events (TEAEs) occurring in >= 5% of patients in at least one treatment group by System Organ Class and Preferred Term Safety Population (SAF)
Table 14.3.1.6 Summary of Treatment-Related Treatment Emergent Adverse Events (TEAEs) occurring in >= 5% of patients in at least one treatment group by System Organ Class and Preferred Term Safety Population (SAF)
Table 14.3.1.7 Summary of Serious Treatment Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term Safety Population (SAF)
Table 14.3.1.8 Summary of Treatment Emergent Adverse Events (TEAEs) leading to treatment discontinuation by System Organ Class and Preferred Term Safety Population (SAF)
Table 14.3.1.9 Summary of Non-Treatment Emergent Adverse Events (NTEAEs) Following Last Dose of Study Drug by System Organ Class and Preferred Term Safety Population (SAF)
Table 14.3.2.1 Listing of Deaths Safety Population (SAF)
Table 14.3.2.2 Listing of Treatment-Emergent Serious Adverse Events (SAEs) Safety Population (SAF)
Table 14.3.2.3 Listing of Non Treatment-Emergent Serious Adverse Events (SAEs) Safety Population (SAF)
Table 14.3.2.4 Listing of Treatment-Emergent Adverse Events (TEAEs) leading to study drug discontinuation, dose reduction or dose interruption Safety Population (SAF)
Table 14.3.4.1.1 Summary of Hematology Quantitative Laboratory Parameters Safety Population (SAF)
Table 14.3.4.1.2 Shift Table of L/N/H Classification for Hematology Laboratory Parameters Safety Population (SAF)
Table 14.3.4.1.3 Shift Table of Incidence of Markedly Abnormal Results for Hematology Laboratory Parameters Safety Population (SAF)
Table 14.3.4.1.4 Listing of Patients with at Least One Markedly Abnormal Hematology Laboratory Parameter Safety Population (SAF)
Table 14.3.4.2.1 Summary of Sodium Laboratory Tests Safety Population (SAF)
Table 14.3.4.2.2 Categorical Summary of Sodium Laboratory Tests Safety Population (SAF)
Table 14.3.4.2.3 Listing of Patients with at Least One Sodium Result > 146 mmol/L or > 150 mmol/L Safety Population (SAF)
Table 14.3.4.3.1 Summary of Liver Function Tests Safety Population (SAF)
Table 14.3.4.3.2 Shift Tables of Categorical Classification for Liver Function Tests – AST Safety Population (SAF)
Table 14.3.4.3.3 Shift Tables of Categorical Classification for Liver Function Tests – ALT Safety Population (SAF)
Table 14.3.4.3.4 Shift Tables of Categorical Classification for Liver Function Tests – Total Bilirubin Safety Population (SAF)

Table 14.3.4.3.5 Listing of Patients with at AST > 3 ULN, ALT > 3 ULN or Total Bilirubin > 1.5 ULN Safety Population (SAF)
Table 14.3.4.4.1 Summary of Other Quantitative Serum Chemistry Laboratory Parameters Safety Population (SAF)
Table 14.3.4.4.2 Shift Table of L/N/H Classification for Other Serum Chemistry Laboratory Parameters Safety Population (SAF)
Table 14.3.4.4.3 Shift Table of Incidence of Markedly Abnormal Results for Other Serum Chemistry Laboratory Parameters Safety Population (SAF)
Table 14.3.4.4.4 Listing of Patients with at Least One Markedly Abnormal Other Serum Chemistry Parameter Safety Population (SAF)
Table 14.3.4.5.1 Summary of Systolic Blood Pressure During the Study Safety Population (SAF)
Table 14.3.4.5.2 Shift Table of Maximum L/N/H Classification for Systolic Blood Pressure During the Study Safety Population (SAF)
Table 14.3.4.5.3 Shift Table of Minimum L/N/H Classification for Systolic Blood Pressure During the Study Safety Population (SAF)
Table 14.3.4.6.1 Summary of Diastolic Blood Pressure During the Study Safety Population (SAF)
Table 14.3.4.6.2 Shift Table of Maximum L/N/H Classification for Diastolic Blood Pressure During the Study Safety Population (SAF)
Table 14.3.4.6.3 Shift Table of Minimum L/N/H Classification for Diastolic Blood Pressure During the Study Safety Population (SAF)
Table 14.3.4.7.1 Summary of Heart Rate During the Study Safety Population (SAF)
Table 14.3.4.7.2 Shift Table of Maximum L/N/H Classification for Heart Rate During the Study Safety Population (SAF)
Table 14.3.4.7.3 Shift Table of Minimum L/N/H Classification for Heart Rate During the Study Safety Population (SAF)
Table 14.3.4.8.1 Summary of Temperature During the Study Safety Population (SAF)
Table 14.3.4.8.2 Shift Table of Maximum L/N/H Classification for Temperature During the Study Safety Population (SAF)
Table 14.3.4.8.3 Shift Table of Minimum L/N/H Classification for Temperature During the Study Safety Population (SAF)
Table 14.3.4.9.1 Incidence of Clinically Notable Vital Sign Abnormalites for SBP, DBP, Respiratory Rate and Heart Rate Safety Population (SAF)
Table 14.3.4.9.2 Listing of Patients with Clincally Notable Vital Sign Abnormalities for SBP, DBP, Respiratory Rate or Heart Rate Safety Population (SAF)
Table 14.3.4.10.1 ECG Results Safety Population (SAF)
Table 14.3.4.10.2 Shift Table of Investigator Interpretation of ECG Results over the study Safety Population (SAF)
Table 14.3.4.10.3 Listing of Clinically Significant Abnormal ECG Results over the study Safety Population (SAF)

## 15.2. DATA LISTINGS FOR APPENDIX 16.2

Mock listings are presented in a separate document. The current version based on this SAP are detailed in CLIN2002-PO SAP Listing Shells V1_0 18JAN17 and detailed below:

Listing 16.2.1.1 Patient Disposition – Screen Failures

Listing 16.2.1.2 Patient Disposition over the study All Randomized Patients

Listing 16.2.2 Protocol Deviations All Randomized Patients

Listing 16.2.3 Treatment and NRS Completion and Analysis Populations All Randomized Patients

Listing 16.2.4.1 Demographics and Informed Consent All Randomized Patients

Listing 16.2.4.2 Other Baseline Characteristics All Randomized Patients

Listing 16.2.4.3.1 Medical History All Randomized Patients

Listing 16.2.4.3.2 MedDRA Classification of Medical History All Randomized Patients

Listing 16.2.4.4.1 Prior and Concomitant Medication All Randomized Patients

Listing 16.2.4.4.2 ATC Classification of Medications All Randomized Patients

Listing 16.2.5.1 Randomization and treatment administration All Randomized Patients

Listing 16.2.5.2 Dosing, Extent of Exposure and Accountability of Study Drug All Randomized Patients

Listing 16.2.5.3 Treatment Compliance All Randomized Patients

Listing 16.2.5.4 Exposure to and Compliance With Rescue Medication All Randomized Patients

Listing 16.2.6.1 Pain NRS All Randomized Patients

Listing 16.2.6.2 WOMAC Index All Randomized Patients

Listing 16.2.6.3 Patient Global Impression of Change All Randomized Patients

Listing 16.2.6.4 Usage of Acetaminophen Rescue Medication for Target Joint Pain All Randomized Patients

Listing 16.2.6.5 Scheduled Efficacy Assessments Not Conducted/Recorded SAF Population

Listing 16.2.7.1 Adverse Events (AEs) SAF Population

Listing 16.2.7.2 MedDRA Classification of Adverse Events All Randomized Patients

Listing 16.2.8.1 Hematology Laboratory Parameters SAF Population

Listing 16.2.8.2 Serum Sodium Results SAF Population

Listing 16.2.8.3 AST, ALT and Total Bilirubin Results SAF Population

Listing 16.2.8.4 Other Serum Chemistry Laboratory Parameters SAF Population

Listing 16.2.8.5 Urine Drug Tests at Screening SAF Population

Listing 16.2.8.6 Pregnancy Tests SAF Population

Listing 16.2.8.7 Vital Signs SAF Population

Listing 16.2.8.8 ECG All Randomized Patients

Listing 16.2.8.9 Scheduled Safety Assessments Not Conducted SAF Population

# 16. ATTACHMENTS

## 16.1. PRIMARY ANALYSIS PROGRAMMING PLAN

The primary efficacy endpoint is defined as the change from baseline at Week 8 with respect to the weekly mean of the daily 24-hour pain intensity in at the index joint.

**Step 1 – Change data structure from ADaM dataset**
Manipulate data into dataset (DATAIN) with multivariate structure with one record per subject and separate variables for treatment (TRTP), OA joint (STRATA) and each weekly NRS mean daily score (NRSWEEK1 to NRSWEEK8) including baseline week (BASE). Keep the reason for treatment discontinuation variable (DCTREAS). Set any imputed data (where DTYPE not missing) to missing.

**Step 2 – Impute intermittent missing data**
Impute intermittent missing data using the Markov Chain Monte Carlo (MCMC) method implemented with the SAS MI procedure with treatment, baseline value and OA joint (hip vs knee) included as follows (need to create numeric 0/1 variables TRTPN and STRATAN based on TRTP [CR845=1, Placebo=0] and STRATA [HIP=1, KNEE=0]):

```
proc mi data=DATAIN nimpute=100 seed=543571857 out=MONOTONIC;
var TRTPN STRATAN BASE NRSWEEK1-NRSWEEK8;
mcmc chain=multiple impute=monotone;
run;
```

This will create 100 records per subject with only monotonic missing data in the dataset MONOTONIC.

Set any imputed values <0 or >10 to 0 and 10 respectively.

**Step 3 – Impute missing data due to AE or Lack of Therapeutic Effect (LOTE)**
If the patient discontinues treatment due to AE or LOTE then missing data will be assumed to follow a distribution similar to the baseline values observed in the patient's randomized treatment.

- To impute missing data, the following will be performed:
  - Obtain the mean (1dp) and standard deviation (2dp) across all the patients in the treatment group at baseline (use DATAIN dataset to do this).

- Split the MONOTONIC data into 3 datasets
    1) Non-missing or missing and discontinued treatment not due to AE/LOTE (Both groups) – discard data
    2) Missing and discontinued treatment due to AE/LOTE (Placebo)
    3) Missing and discontinued treatment due to AE/LOTE (CR845)
- Sort the missing datasets (2 and 3) by imputation and then subject number
- Within each of the missing datasets (2 and 3), create a variable using the RANNOR function in SAS, seed=543571857, with mean and standard deviation as found above
- Impute any missing values of NRSWEEKx with the value of the created variable.
- Set any imputed values <0 or >10 to 0 and 10 respectively.
- Combine datasets from 2 and 3 with all data imputed and keep for Step 5.

**Step 4 - Impute missing data due to reasons other than AE or LOTE**

If the patient discontinues due to reasons other than AE or Lack of Efficacy, at each time point, missing data will be assumed to follow a distribution similar to scores for patients who are still in the study and randomized to the same treatment group.

- To impute missing data at the first week x with any missing data not due to AE/LOTE, the following will be performed to the MONTONIC dataset:
    - Split the MONOTONIC data into 2 datasets
        1) Non-missing at Week x or missing at Week x and discontinued treatment not due to AE/LOTE
        2) Missing at Week x and discontinued treatment due to AE/LOTE – discard data
    - Sort dataset (1) by imputation and subject number
    - Impute the missing data at week x in dataset (1) as follows:

```
proc mi data=DATA1 out=WEEKx seed=543571857 nimpute=1;
by  Imputation ;
var TRTPN STRATAN BASE NRSWEEK1-NRSWEEKx;
monotone regression;
run;
```

    - Set any imputed values <0 or >10 to 0 and 10 respectively.

- To impute missing data at the next week y with any missing data not due to AE/LOTE, the following will be performed to the WEEKx dataset:
    - Split the WEEKx data into 2 datasets

3) Non-missing at Week y or missing at Week y and discontinued treatment not due to AE/LOTE
4) Missing at Week y and discontinued treatment due to AE/LOTE – discard data

- Sort dataset (3) by imputation and subject number
- Impute the missing data at week y in dataset (3) as follows:

```
proc mi data=DATA3 out=WEEKy seed=543571857 nimpute=1;
by _Imputation_;
var TRTPN STRATAN BASE NRSWEEK1-NRSWEEKx;
monotone regression;
run;
```

- Set any imputed values <0 or >10 to 0 and 10 respectively.

- Continue in similar fashion until no missing data with dataset including either completers or those that discontinued treatment not due to AE/LOTE.

**Step 5 – Combine datasets back together and re-structure back as per ADaM**
Combine final datasets from Steps 3 and 4. Sort by imputation and subject number. Re-structure dataset to be univariate with a Week (AVISITN = 1, 2,3 ,4 ,5 ,6 7, 8) and AVAL (using NRSWEEK1 to NRSWEEK8) and CHG (AVAL-BASE) variable. (MMRMDATA)

**Step 6 – Perform MMRM on each imputed dataset**
The primary efficacy variable will be analyzed using a mixed effect model with repeated measures (MMRM). The model will contain treatment (placebo vs CR845), treatment-by-week interaction as fixed effects, baseline value and OA joint (hip vs knee) as covariates, and patient as a random effect.

```
proc mixed data=MMRMDATA;
by _Imputation_;
class TRTP AVISITN STRATA SUBJID;
model CHG = TRTP|AVISITN STRATA BASE /ddfm=kr;
repeated AVISITN/SUBJECT=SUBJID type=un;
lsmeans TRTP*AVISITN/diff;
ods output Diffs=DIFF_MI LSMeans=LSM_MI;
run;
```

Check log to ensure that there are no convergence problems across all 100 analyses. If so then use different covariance structures in the following order until this is achieved:

```
type=toep
```

```
type=ar(1)

type=cs

type=vc
```

Extract 2 separate datasets from DIFF_MI and LSM_MI selecting the LS means (and standard errors) and difference between LS means (and standard errors) at Week 8 only. (DIFF_MI8 and LSM_MI8).

**Step 7 – Combine MMRM results using Proc MIANALYZE**

Use PROC MIANALYZE to obtain estimates of LS means and differences and 95% CIs.

```
proc mianalyze data=DIFF_MI8;
modeleffects estimate;
stderr sdterr;
ods output ParameterEstimates=DIFF_MIAN8;
run;

proc mianalyze data=LSM_MI8;
modeleffects estimate;
by TRTP;
ods output ParameterEstimates=LSM_MIAN8;
run;
```