Official Title: A Phase 2, Randomized, Double-blind, Placebo-controlled Study to Evaluate the Efficacy and Safety of ACP-044 in Subjects With Pain Associated With Osteoarthritis of the Knee

NCT Number: NCT05008835

Document Date: 21 September 2022

STATISTICAL ANALYSIS PLAN

Protocol No.:	ACP-044-005
Protocol Title:	A Phase 2, Randomized, Double-blind, Placebo-controlled Study to Evaluate the Efficacy and Safety of ACP-044 in Subjects With Pain Associated With Osteoarthritis of the Knee
Drug:	ACP-044
Sponsor:	Acadia Pharmaceuticals Inc. [redacted]
Version No. and Date	Version 1.0, 21 September 2022

SIGNATURE/APPROVAL PAGE

AUTHOR

See appended electronic signature page

[redacted]
Senior Director, Biostatistics
Acadia Pharmaceuticals Inc.

Date

APPROVERS

See appended electronic signature page

[redacted]
Executive Director, Clinical Science
Acadia Pharmaceuticals Inc.

Date

See appended electronic signature page

[redacted]
Executive Director, Biostatistics
Acadia Pharmaceuticals Inc.

Date

Signature Page for ACP-044-005 SAP v1.0

Approve	[redacted] 26-Sep-2022 18:25:27 GMT+0000

Approve	[redacted] 26-Sep-2022 21:02:13 GMT+0000

Approve	[redacted] 30-Sep-2022 22:32:35 GMT+0000

Signature Page for ACP-044-005 SAP v1.0

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AE	adverse event
ANCOVA	analysis of covariance
ATC	Anatomical Therapeutic Chemical
BID	twice daily
BMI	body mass index
BPI-sf	Brief Pain Inventory - Short Form
COVID-19	Coronavirus disease 2019
C-SSRS	Columbia-Suicide Severity Rating Scale
ECG	electrocardiogram
eCRF	electronic case report form
EOT	end of treatment
ET	early termination
GSD	Guidance for Site Documentation and Data Management Querying of Data Impacted by COVID-19
HADS	Hospital Anxiety and Depression Scale
K-L	Kellgren-Lawrence
MedDRA	Medical Dictionary for Regulatory Activities
msec	milliseconds
NRS	numerical rating scale
OA	osteoarthritis
PCI	potentially clinically important
PGIC	Subject Global Impression of Change
QID	four times daily
QTcB	QT interval corrected for heart rate using Bazett's formula
QTcF	QT interval corrected for heart rate using Fridericia's formula
SAE	serious adverse event
SAP	statistical analysis plan
SAS®	Statistical Analysis System
SD	standard deviation
SE	standard error
SF-12	12-item Short Form Survey
SOC	system organ class
TEAE	treatment-emergent adverse event
TESAE	treatment-emergent serious adverse event
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index

1 INTRODUCTION

This statistical analysis plan (SAP) provides a technical and detailed elaboration of the statistical analyses of safety and efficacy data as described in the study protocol Amendment 4 dated 16 February 2022. This trial was terminated on 8 August 2022 with 61 patients randomized.

Specifications for tables, figures, and listings are contained in a separate document. Statistical analyses for population pharmacokinetic (PK) and PK/pharmacodynamics (PD) modeling will be presented in a separate report and therefore will not be included in this SAP.

This plan should be read in conjunction with the study protocol and the electronic case report forms (eCRF)

2 OBJECTIVES

2.1 Primary Objective

The primary objective of this study is to evaluate the efficacy of ACP-044 compared with placebo in the treatment of pain associated with osteoarthritis (OA) of the knee.

The primary endpoint is:

The primary endpoint is change from Baseline to Week 4 in the weekly average of the daily average Numeric Rating Scale (NRS) pain intensity scores.

2.2 Secondary Objective

The secondary objective of this study is to evaluate the efficacy of ACP-044 compared with placebo in the treatment of pain associated with osteoarthritis (OA) of the knee.

The key secondary endpoints are:

- Patient Global Impression of Change (PGIC) at Week 4 with reference to baseline status
- Change from Baseline to Week 4 in Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) total score and individual subscale scores for pain, stiffness, and physical function
- Proportion of subjects who are responders as defined by 30% and 50% reduction from Baseline to Week 4 in the weekly average of the daily average NRS pain intensity scores

Additional secondary endpoints are:

- Change from Baseline to Week 4 in:
  - Weekly average of the daily worst NRS pain intensity scores
  - Brief Pain Inventory–Short Form (BPI-sf)

- 12-item Short Form Survey (SF-12)

2.3 Exploratory Objective

The exploratory objective of this study is to evaluate the use of rescue medication in subjects with pain associated with OA of the knee treated with ACP-044 compared with placebo.

2.4 Safety Objective

The safety objective of this study is to evaluate the safety and tolerability of ACP-044 compared with placebo in the treatment of pain associated with OA of the knee.

2.5 Pharmacokinetic Objective

The PK objective of the study is to characterize the PK profile of ACP-044 in subjects with pain associated with OA of the knee.

2.6 Pharmacokinetic/Pharmacodynamic Objective

The pharmacokinetic/pharmacodynamic (PK/PD) objective of the study is to characterize the exposure-response relationship using appropriate modelling and simulation methods.

3 STUDY DESIGN

3.1 General Study Design

This is a Phase 2, randomized, double-blind, placebo-controlled, parallel-group study to evaluate the efficacy and safety of ACP-044 in subjects with pain associated with radiographically confirmed OA of the knee. The study will compare active treatment groups, receiving either 800 mg ACP-044 BID or 400 mg ACP-044 QID, with a placebo group. The Sponsor, subjects, and Investigators will be blinded to treatment assignment.

The study periods are:

- Screening: up to 4 weeks
- Baseline Pain Assessment: a 7-day period during which open-label placebo will be taken and baseline pain assessments will be recorded. Additional 3-day window may be added, if needed.
- Double-blind treatment: 4 weeks
- Safety follow-up: approximately 7 days after end of treatment (EOT)/early termination (ET) and approximately 30 days after the last dose of study drug

Screening Period (Up to 4 weeks)

During the Screening period, subjects will be assessed for study eligibility. All subjects must have radiographically confirmed OA of the knee, including review of imaging and

confirmation of a Kellgren-Lawrence (K-L) score of 2 or 3. Subjects who have had an X-ray of the index knee within 12 weeks of Screening may provide a historical image for central reading to confirm eligibility, or imaging can be done during Screening at protocol-trained radiological sites. Randomization visits cannot occur until there is confirmation from the central reader that there are no exclusionary findings on the knee joint images. The index knee is defined as the most painful knee at Screening.

Prohibited analgesic medications must be discontinued during the Screening period. Investigators must not withdraw a subject's prohibited medication other than analgesics solely for the purpose of enrolling them into the study. Medications will be discontinued only if it is deemed clinically appropriate to do so and in consultation with the prescribing physician.

All pain medication, except for the study-provided rescue medication (acetaminophen), will be discontinued 7 days or 5 half-lives, whichever is longer, prior to the Baseline Pain Assessment Period. In the event of inadequate pain relief, acetaminophen may be taken with a maximum daily dose of 2500 mg.

Rescreening of a subject will be allowed, with the approval of the Medical Monitor.

Baseline Pain Assessment Period (1 Week)

Following eligibility at screening, subjects will complete a Baseline Pain Assessment Period that will occur approximately 7 days before randomization; an additional 3-day window may be added, if needed. During the Baseline Pain Assessment Period, subjects will be instructed in the use of a handheld device (electronic diary [eDiary]) to record their daily pain scores and rescue medication use. Compliance with recording of pain score and use of rescue medication will be assessed for continued subject eligibility.

Subjects must discontinue taking study-provided rescue medication during the Baseline Pain Assessment Period and for at least 24 hours prior to the start of study visits. Rescue medication use during the Baseline Pain Assessment Period through Visit 3 and within 24 hours of the study visits (Visits 5, 6, and 8) will be documented in the electronic data capture (EDC).

Subjects who enter the Baseline Pain Assessment Period will not be allowed to rescreen. Subjects who experience technical issues with the device and their pain scores were not assessed may be rescreened and the Baseline Pain Assessment Period may be extended, if needed and with permission of the Sponsor.

Treatment Period (4 Weeks)

At Baseline (Week 0), eligible subjects will be randomly assigned in a 1:1:1 ratio to receive ACP-044 800 mg BID, ACP-044 400 mg QID, or placebo, according to a randomization schedule. Randomization will be stratified by sex and baseline K-L score (2 or 3).

All efficacy assessments will be completed and a PK blood sample will be taken at the Baseline visit prior to administration of the first dose of study drug. Subjects will continue daily entry into the eDiary of pain scores, compliance with ingestion of study medication, and rescue medication use from Week 0 through to Week 4/EOT/ET.

Subjects will receive their first dose at the study site after all Baseline assessments have been completed and a PK blood sample will be taken approximately 1 hour after dosing. Dosing is four times a day.

During the treatment period (Week 0 through Week 4), subjects will be permitted to use only acetaminophen as rescue medication. Subjects will record their use of acetaminophen in their eDiary. Subjects must discontinue acetaminophen use for at least 24 hours prior to the start of study visits in order to minimize the confounding effects of the rescue medication on efficacy measurements.

A PK sample will be taken at each study visit according to the schedule described in the Pharmacokinetic Assessments section.

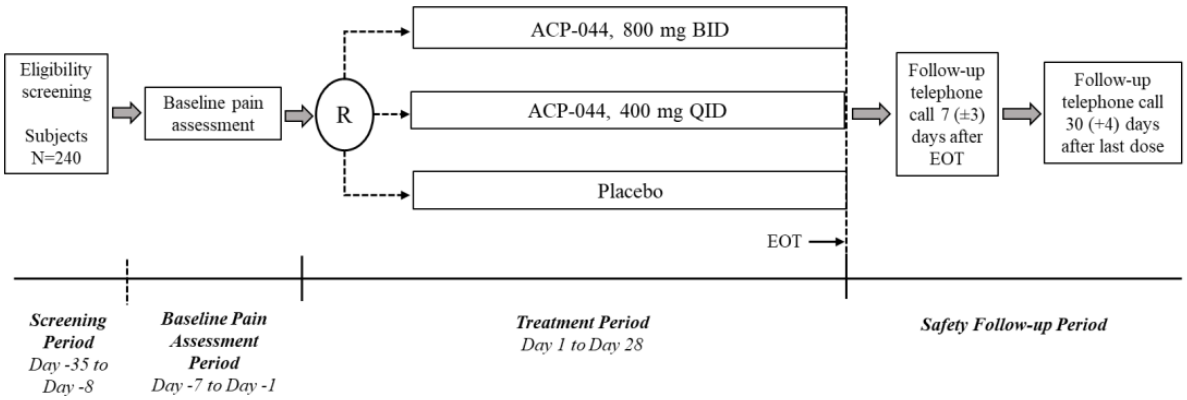
At the Week 4 visit, end of treatment (EOT) procedures will be completed.

Safety Follow-up Period (30 Days)

Safety will be further assessed during a follow-up visit to occur in the clinic approximately 7 (±3) days after EOT/ET procedures. During this in-clinic visit, pain intensity, adverse events (AEs), and concomitant medications will be assessed, since the EOT/ET. Then, a follow-up telephone call will occur approximately 30 (+4) days after the last dose of study drug to assess AEs and concomitant medications.

The study design schematic is presented in Figure 1. The schedule of assessments is provided in Appendix A.

Figure 1 Schematic of Study Design



Abbreviations: BID=twice daily; EOT=end of treatment; QID=four times daily; R=randomization

3.2 Schedule of Assessments

Schedule of events and assessments can be found in Appendix A.

3.3 Randomization

On Day 1 eligible subjects who meet inclusion and do not meet exclusion criteria will be randomized in a 1:1:1 ratio to receive ACP-044 800 mg BID, ACP-044 400 mg QID, or placebo, according to a randomization schedule. Randomization will be stratified by sex and baseline K-L score.

Approximately 50 United States (US) sites will participate in this study.

3.4 Blinding

Treatment assignments will be blinded to all study subjects, Investigators, site personnel, and Sponsor personnel.

In order to maintain the double-blind design of the study, all subjects will receive the same numbers of tablets at the same intervals, regardless of the treatment group to which they are randomized. Placebo tablets will be size- and color-matched to the ACP-044 tablets.

The investigational products are ACP-044 400 mg tablets, or matching placebo (size- and color-matched) tablets. Study drug doses are to be administered orally for 4 weeks.

Doses to be studied are:

- ACP-044 800 mg dose delivered two times a day (provided as 2×400 mg ACP-044 tablets)
- ACP-044 400 mg dose delivered four times a day (provided as 1×400 mg ACP-044 tablet)

- Placebo, size- and color-matched to ACP-044 tablets

Blinded drug will be provided in blister cards. All subjects will take the same number of tablets per day in order to maintain the blinded study design.

Dosing is to occur in the morning (2 tablets), noon (1 tablet), evening (2 tablets), and night (1 tablet) and should occur at the same time each day over the duration of the study.

Table 1 Blinded Daily Dosing Regimen

	First Daily Dose	Second Daily Dose	Third Daily Dose	Fourth Daily Dose
Placebo	PP	P	PP	P
ACP-044 400 mg QID (four times daily)	XP	X	XP	X
ACP-044 800 mg BID (twice daily)	XX	P	XX	P

Abbreviations: P=ACP-044 placebo table; X=ACP-044, 400 mg tablet

3.5 Determination of Sample Size

Approximately 240 subjects will be randomized in a 1:1:1 ratio to ACP-044 800 mg BID dose, ACP-044 400 mg QID dose, or placebo. Assuming a treatment difference of 0.85 and a common standard deviation of 1.7 between an ACP-044 dose arm and placebo, 64 subjects per arm is needed to provide approximately 80% power to detect this treatment difference using a 2 sample t-test at 2-sided alpha level of 0.05. Assuming a discontinuation rate of 20%, 80 subjects per treatment group will be randomized.

The planned sample size is 240 subjects randomized (80 subjects per treatment group).

3.6 Coronavirus Disease 2019

In March, 2020, the emerging coronavirus disease 2019 (COVID-19) pandemic resulted in implementation of urgent safety measures designed to ensure subject safety. Mechanisms to record information on the potential impact of the COVID-19 pandemic on data itself, as well as data collection and integrity, were implemented (as detailed in the "Guidance for Site Documentation and Data Management Querying of Data Impacted by COVID-19" [GSD] in the Data Management Plan).

The impact of COVID-19 on the statistical analysis is discussed in each of the relevant sections of this SAP.

Subjects who remained in this study who could not attend clinic visits were allowed to participate in remote visits via telephone or video, or site staff conducted the visit in a subject's home. Sites were required to document details of all visits that are administered remotely.

4 ANALYSIS SETS

Safety Analysis Set

The Safety Analysis Set includes all randomized subjects who received at least one dose of study drug (ACP-044 or placebo). Subjects will be analyzed based on the treatment that they actually received. The Safety Analysis Set will be used for all safety analyses.

Full Analysis Set

The Full Analysis Set includes all randomized subjects who received at least one dose of study drug and with baseline pain assessments and at least one post-baseline pain assessment. The Full Analysis Set will be used for the analysis of all efficacy endpoints. Subjects will be analyzed based on their randomized treatment assignment.

Pharmacokinetics Analysis Set

The Pharmacokinetics Analysis Set includes all subjects with at least one measurable ACP-044 plasma concentration. The Pharmacokinetics Analysis Set will be used for ACP-044 plasma concentration summaries.

5 DATA HANDLING CONVENTIONS

All data collected in the study will be listed.

5.1 General Data Reporting Conventions

Continuous variables will be summarized using the following descriptive statistics: number of subjects, mean, median, standard deviation (SD), standard error (SE), minimum, maximum and quartiles when necessary. Unless specified otherwise, means, medians, and confidence intervals (CIs) will be presented to one more decimal place than the raw data, and the standard deviations and standard errors will be presented to two more decimal places than the raw data. In general, the maximum number of decimal places is 4 and values will be truncated to 4 decimal places in situations where there are more than 4 decimal places. Wherever possible data will be decimal aligned.

Categorical variables will be summarized by the number of subjects and the percent of subjects in each category; the number of subjects and the percentage of subjects with missing data will be summarized for demographic and baseline characteristics (if applicable). Categories with zero counts will not have zero percentages displayed. For demographic summaries, percentages will be calculated by using the total number of subjects in the given treatment group as the denominator. Percentages will be presented with one decimal place.

Duration in months will be calculated as ([the number of days / 365.25] *12).

Unless specified otherwise, all statistical tests will be 2-sided hypothesis tests performed at the significance level of 5% for main effects and all CIs will be 2-sided 95% CIs. P-values will generally be presented to 4 decimal places; values less than 0.0001 will be presented as <0.0001.

Values that are collected with "<" or ">" signs will generally be analyzed as the numerical value without the sign in tables and figures. In listings, these data will be reported as collected with the sign.

5.2 Derived Variables

In general, the assessment scale scores will be directly taken from the electronic case report form (eCRF).

5.2.1 Numeric Rating Scale (NRS) of Pain Intensity Scores

A 0 to 10 NRS will be used to assess the subject's pain throughout the study. Subjects will be required to assess and record daily in the eDiary their knee pain intensity and daily use of rescue medication. Subjects will be rating both the average daily pain as well as their worst daily pain.

Knee pain intensity will be assessed by describing their knee pain at the present time from 0 to 10 where "0" means "no pain at all" and "10" means "the worst pain imaginable".

Pain intensity assessments are to be recorded in the evening prior to their 4th dose of the day.

Subjects are encouraged to refrain from acetaminophen use for at least 24 hours prior to the start of study visits in order to minimize the confounding effects of the rescue medication on efficacy measurements. At study visits, pain assessments should be performed before blood sample collection, ECG, and other procedures

5.2.2 Patient Global Impression of Change (PGIC)

The self-report measure Patient Global Impression of Change (PGIC) reflects a subject's belief about how their pain has changed since start of study treatment. PGIC is a 7-point scale depicting a subject's rating of overall improvement. Subjects rate their change as "very much improved" = 1, "much improved" =2 "minimally improved"=3 "no change" = 4 "minimally worse" = 5 "much worse" = 6 or "very much worse" = 7.

The PGIC will be completed at Week 1, Week 2, and Week 4 (EOT)/ET. At study visits, the PGIC should be performed before blood sample collection, ECG, and other procedures.

5.2.3 Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)

The WOMAC is widely used in the evaluation of hip and knee osteoarthritis (Collins et al. 2011) and rates pain during the last 48 hours. It is a self-administered questionnaire consisting of 24 items divided into three subscales:

- Pain (5 items): during walking, using stairs, in bed, sitting or lying, and standing upright
- Stiffness (2 items): after first waking and later in the day
- Physical Function (17 items): using stairs, rising from sitting, standing, bending, walking, getting in/out of a car, shopping, putting on/taking off socks, rising from bed, lying in bed, getting in/out of bath, sitting, getting on/off toilet, heavy domestic duties, light domestic duties

The test questions are scored on a scale of 0-4, which correspond to: None (0), Mild (1), Moderate (2), Severe (3), and Extreme (4). The scores for each subscale are summed up, with a possible score range of 0-20 for Pain, 0-8 for Stiffness, and 0-68 for Physical Function. The sum of the scores for all three subscales provides the total WOMAC score. Higher scores on the WOMAC indicate worse pain, stiffness, and functional limitations.

The WOMAC will be completed at Baseline (Week 0), Week 1, Week 2, and Week 4 (EOT)/ET. At study visits, the WOMAC should be performed before blood sample collection, ECG, and other procedures.

5.2.4 Brief Pain Inventory – Short Form (BPI-sf)

The BPI is a self-report measure used to assess the severity of pain and its impact on functioning in subjects with chronic diseases. The short form version uses a 24-hour recall period (i.e., assesses pain and its impact over the previous 24 hours).

The BPI uses a 0 (no pain) to 10 (pain as bad as you can imagine) numeric rating scale to measure how much pain has interfered with seven daily activities, including general activity, walking, work, mood, enjoyment of life, relations with others, and sleep.

The BPI will be completed at Baseline (Week 0), Week 1, Week 2, and Week 4 (EOT)/ET. At study visits, the BPI should be performed before blood sample collection, ECG, and other procedures.

5.2.5 12-item Short Form Survey (SF-12)

The SF-12 survey, version 2 (SF-12v2), is a shortened form (12 items) of the SF-36 Health Survey (Ware et al. 1996). It is a self-reported general assessment of health-related quality of life from the subject's perspective. The SF-12 has eight domains including physical health

summary measures: physical functioning, role–physical, bodily pain, and general health and mental health summary measures: vitality, social functioning, role–emotional, and mental health. From the evaluation of these domains, algorithms are used to generate physical and mental health composite summary scores, PCS and MCS, respectively, for comparison to normative data. In normative data, the mean score is set to 50; thus, scores >50 indicate better physical or mental health than the mean scores and scores < 50 indicate worse health. The timeframe for this scale is the past 4 weeks.

Scales of SF-12 (health survey) Physical Component Summary (PCS) and Mental Component Summary (MCS) will be derived using the weights provided in the SF-12 manual developed by J. Ware, M. Kosinski and S. Keller.

For variable 12 (Physical Component Summary (PCS) and Mental Component Summary (MCS) of SF-12), PCS and MCS are created from the SF-12 survey in this way. First, subjects with missing answer to any one of the 12 items will be ignored, i.e., will have their PCS-12 and MCS-12 scores missing. Indicator variables (1/0) will be created for all choices of each the 12 items, except that the choice associated with the best health status under each item. Thus totally 35 indicator variables will be created. For example, for item “In general, would you say your health is”, indicator variables will be created for each of the choices “Poor”, “Fair”, “Good”, and “Very Good”, but not the choice “Excellent”. This means that if a subject marks “Good” under this item, the indicator variable for “Good” will have a value 1, and the indicator variables for “Poor”, “Fair” and “Very Good” will have value 0. If a subject marks “Excellent” for this item, all indicator variables under this item will have value 0. Then all these indicator variables will be weighted based on the SF-12 Manual for the PCS-12 and MCS-12 separately. Summing of these indicator variables based on their weight, and plus the constant (56.57706 for PCS-12 and 60.75781 for MCS-12) will provide the PCS-12 and MCS-12 scores for each subject.

The SF-12 will be completed at Baseline (Week 0) and Week 4 (EOT)/ET.

5.2.6 Hospital Anxiety and Depression Scale (HADS)

The HADS was developed to detect states of anxiety or depression (Zigmond and Snaith 1983). The questionnaire focuses on non-physical symptoms so that it can be used to diagnose depression in people with significant physical ill-health. The questionnaire comprises seven questions for anxiety and seven questions for depression, which are each scored separately. Each item on the questionnaire is scored from 0 (No, not at all) to 3 (Yes definitely), for a total score between 0 to 21 for either anxiety or depression. The timeframe for this scale is the past week.

The HADS will be completed at Screening, Baseline (Week 0), and Week 4 (EOT)/ET.

5.2.7 Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is assessed at Baseline, Day 7, Day 14, and Day 28 (EOT/ET).

The C-SSRS version assessing information since the last visit will be completed at all visits (including the Baseline visit).

There are 5 questions about suicidal ideation, representing 5 types of suicidal ideation: wish to be dead; non-specific active suicidal thoughts; active suicidal ideation with any methods (not plan) without intent to act; active suicidal ideation with some intent to act, without specific plan; active suicidal ideation with specific plan and intent. If a subject answers "yes" to any of these 5 questions, this subject will be counted as having suicidal ideation.

There are 5 questions about suicidal behavior, representing 5 types of suicidal behavior: preparatory acts or behavior; aborted attempt; interrupted attempt; actual attempt; suicide. If a subject answers "yes" to any of these 5 questions, this subject will be counted as having suicidal behavior.

Suicidality is defined as a subject who reported at least 1 occurrence of suicidal ideation or at least 1 occurrence of suicidal behavior at any post-Baseline visit including unscheduled and out of window visits.

Missing C-SSRS item scores will not be imputed.

5.3 Data Presentation Conventions

- 1 year = 365.25 days. Year is calculated as (days/365.25) and will be rounded to 1 digit after the decimal point (tenths) for presentation purposes.
- 1 month = 30.4375 days. Month is calculated as (days/30.4375) and will be rounded to 1 digit after the decimal point (tenths) for presentation purposes.
- 1 pound = 0.454 kg and 1 kg = 2.2 pounds.
- 1 inch = 2.54 cm and 1 cm = 0.3937 inches.
- Body mass index (BMI) calculated as [weight (kg)/height $(m)^2$]

5.4 Study Day

If the date of assessment occurs on or after the first dose date of double-blind study medication, then study day will be calculated as (date of assessment – date of first double-blind dose) + 1. If the date of assessment occurs prior to the first double-blind dose date, then study day will be calculated as (date of assessment – date of first double-blind dose). There is no study day 0.

5.5 Baseline Definition

Baseline data are defined as data collected which are prior to the administration of the first dose. If there is more than one value on or prior to Study Day -1, the value closest to and prior to the receipt of the first dose, whether scheduled or unscheduled, will be used as the Baseline value.

The baseline of the average of the daily average NRS pain intensity score will be the average of the scores during the baseline pain assessment period which is 7 consecutive days prior to the first dose of study medication.

5.6 Analysis Visit Windows

Assessment nominal visit window will be applied.

5.6.1 Unscheduled Assessments

Both scheduled and unscheduled assessments, including the assessments at early termination visits, will be considered for planned timepoint summaries based on the above analysis visit windowing rules. All assessments will be presented in data listings.

5.6.2 Multiple Measurements within Visit Windows

In the event that more than one assessment falls within a given window, the assessment closest to the target study day will be selected for the by-visit analysis. If two assessments are equidistant from the target study day, then the chronologically last assessment will be used. Exceptions may be made for incomplete assessments, in which case, more complete assessments may be given priority. Details are provided in a separate programming conventions document.

For safety analyses where the most extreme values should be selected, e.g. overall post-baseline minimum, overall post-baseline maximum, and potentially clinically important values for overall post-Baseline summaries, all non-missing post-Baseline values should be considered, regardless of whether the value is selected for the by-visit summaries. All results will be presented in data listings.

5.7 Missing or Incomplete Date for Last Dose of Study Drug

For subjects with completely missing last dose date, the last dose date will be imputed by the last expected dosing date, defined as the earliest of the following dates: last dose administer date + scheduled dosing interval per protocol, EOT/ET date.

For subjects with partial missing last dose date, the imputation will be compared against the last expected dosing date as defined above. Detailed algorithms will be documented in a separate programming specifications document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

5.8 Missing or Incomplete Dates for Concomitant or Post-Treatment Medications

Missing or incomplete medication start or end dates will be imputed for the purpose of determining whether the medications are taken concomitantly (see Section 11 for definition). When the chronological order of medication use relative to the study drug treatment period is unclear due to missing or incomplete date(s), the medication will be considered as concomitant. The imputation algorithms will be detailed in the analysis dataset specification document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

5.9 Missing or Incomplete Dates for Adverse Events

Missing or incomplete adverse event (AE) start dates will be imputed for the purpose of determining whether the AEs are treatment-emergent (see Section 14.1 for definition). When the chronological order of an AE onset relative to the study drug treatment period is unclear due to missing or incomplete date(s), the AE will be considered as treatment-emergent. The imputation algorithms will be detailed in the analysis dataset specification document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

5.10 Missing Severity Assessment for Adverse Events

If the severity is missing for a treatment-emergent AE, a severity of “Severe” will be assigned. The imputed values for severity assessment will be used for incidence summaries, and the actual values will be presented in data listings.

5.11 Missing Relationship to Study Drug for Adverse Events

If the relationship to study drug is missing for a treatment-emergent AE, a causality of “Related” will be assigned. The imputed values for relationship to study drug will be used for incidence summaries, and the actual values will be presented in data listings.

5.12 Character Values of Clinical Laboratory Variables

If the reported value of a clinical laboratory variable cannot be used in a summary due to, for example, a character string reported for a numeric variable, an appropriately determined coded value will be used in the summary. The coding algorithms will be detailed in the analysis dataset specification document. The actual values as reported in the database will be presented in data listings.

6 SUBJECT DISPOSITION

For subjects who participate in the screening phase but are not randomized (screen failures), their demographics information (including age, sex, and primary race), screen failure reasons

(the specific inclusion/exclusion criterion (or criteria) not met or other reasons including the reason due to the COVID-19 public health emergency (PHE) and protocol version will be listed. If a subject is re-screened, then the re-screening subject ID and the final enrollment status (whether eventually enrolled) will also be displayed in this listing. In addition, the frequency that the screen failure reasons are reported will be summarized. Note that one subject may be deemed ineligible for multiple inclusion/exclusion criteria and may be allowed to rescreen with the permission of the Medical Monitor, provided the screen failure was due to a temporary condition that subsequently resolved.

The number of sites that screened at least 1 subject, number of sites that randomized at least 1 subject, number of subjects screened, and number of unique subjects screened will be tabulated. In addition, the number of subjects enrolled at each site will also be tabulated by Analysis Set and by treatment group and overall.

The number and percentage of subjects who completed the study, discontinued early (all discontinued and by discontinuation reasons including reason due to the COVID-19 PHE), and the reason for discontinuation will be summarized by treatment group using the Randomized and Full Analysis Sets. All subjects excluded from the Safety, Full Analysis Sets, and the reason(s) for exclusion will be listed.

7 PROTOCOL DEVIATIONS

Protocol deviations will be reviewed periodically over the course of the study. The review process, definition of the deviation categories, and the classification of a deviation as major or minor are detailed in the Protocol Deviation Management Plan. Protocol deviations will also be assessed with respect to relationship to the COVID-19 PHE. A summary of the number and percentage of subjects with major protocol deviations for each deviation category will be presented by treatment group for the Randomized Analysis Set in three ways: all protocol deviations, COVID-19-PHE related protocol deviations, and non-COVID 19-PHE related protocol deviations. Three data listings of all protocol deviations, COVID-19-PHE related protocol deviations, and non-COVID-19-PHE related protocol deviations will be provided.

8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographics and Baseline characteristics will be summarized for the Safety Analysis Set using descriptive statistics. Variables include age, age group (18-40 years and >40 years), sex, primary race, ethnicity, height, weight, and body mass index (BMI).

Race will also be categorized by White vs. Non-White. The eCRF reported age reflects a subject’s age at the Baseline visit date.

9 MEDICAL HISTORY

Medical history data will be coded using Medical Dictionary for Regulatory activities (MedDRA) version 23.0. The subject incidence will be summarized for each system organ class (SOC) and preferred term for the Safety Analysis Set. A subject will be counted only once per SOC or per preferred term for the summary.

A listing of the SOC, preferred term, body system, verbatim term for the medical history condition/event, start and stop dates (when available), and an indicator for whether or not the condition is ongoing will be provided.

10 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

Summaries of exposure and compliance to study drug will be provided for the Safety Analysis Set.

10.1 Exposure to Study Drug

The number of doses dosed, the number of morning doses dosed, the number of missed doses, the number of missed morning doses, and the number of total tablets dosed will be summarized.

10.2 Treatment Compliance

Study drug compliance will be calculated as the number of tablets taken, divided by the expected number of tablets taken*100.

Treatment compliance will be summarized as a categorical variable. The number and percentage of subjects within each of the following compliance levels will be tabulated: <80%, 80 to 120% and >120%.

The number of missed doses and the number of missed morning doses will be summarized as a categorical variable.

11 CONCOMITANT AND POST-TREATMENT MEDICATION

Prior medication is defined as any medication with stop dates prior to the date of the first dose of study drug. Concomitant medication is defined as any medications that are ongoing at the first dose of study drug or with a start date between the dates of the first and last doses of study drug, inclusive. Post-treatment medication is defined as any medication with a start date after the date of the last dose of study drug. Medications will be coded using WHO Drug Global Dictionary March 2020 or newer version.

The number and percentage of subjects taking prior, concomitant and post-treatment medications will be tabulated separately by each drug class (ATC Level 3) and medication preferred term, treatment group and overall for Safety Analysis Set. Multiple medication

usage by a subject in the same category will be counted only once. Listings of the prior, concomitant and post-treatment medications will also be provided.

COVID-19 Public Health Emergency (PHE) Related Medications

Concomitant medication analyses described above will also be tabulated and listed by relationship to the COVID-19 PHE (Not related to the COVID-19-PHE vs. Related to the COVID-19-PHE).

12 EFFICACY ANALYSES

Unless otherwise specified, all efficacy analyses will be performed using the planned treatment assignments based on the randomization schedule for the Full Analysis Set.

12.1 Efficacy Variables

Primary Efficacy Endpoint

- Change from Baseline to Week 4 in the average of the daily average Numeric Rating Scale (NRS) pain intensity scores

Key Secondary Efficacy Endpoints

- Patient Global Impression of Change (PGIC) at Week 4 with reference to baseline status
- Change from Baseline to Week 4 in Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) total score and individual subscale scores for pain, stiffness, and physical function
- Proportion of subjects who are responders as defined by 30% and 50% reduction from Baseline to Week 4 in the average of the daily average NRS pain intensity scores

Additional Secondary Efficacy Endpoints

Change from Baseline to every study visit in:

- Weekly average of the daily worst NRS pain intensity scores
- Brief Pain Inventory–Short Form (BPI-sf)
- 12-item Short Form Survey (SF-12)

Exploratory Endpoints

- Comparison of average daily rescue medication use
- Hospital Anxiety and Depression Scale (HADS)

12.2 Adjustment for Covariates

The baseline average of daily average pain score as a covariate in the model.

12.3 Handling of Missing Data

Weekly Mean Pain Intensity Scores

If a subject 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 subject 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 13. Sensitivity analyses will be conducted exploring different patterns for missing data.

WOMAC Index Total Score

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.

12.4 Multiple Comparisons / Multiplicity

The primary efficacy endpoint is change from baseline to week 4 in the average of the daily average Numeric Rating Scale (NRS) pain intensity scores. For the primary efficacy analysis, the null hypothesis is that there is no treatment effect between 400 mg QID and placebo during the study versus the alternative hypothesis that there is a treatment effect. Similar hypothesis applies to the comparison of 800 mg BID to placebo.

The overall Type I error is controlled by a hierarchical testing procedure. The procedure is to test the difference of the treatment effect between 400 mg QID versus placebo first, at significant level $\alpha=0.05$. If this test is significant, then comparison for 800 mg BID versus placebo will be performed at $\alpha=0.05$ level.

12.5 Examination of Subgroups

No subgroup analyses for primary endpoint are planned.

13 METHODS OF EFFICACY ANALYSES

13.1 Primary Efficacy Endpoints

13.1.1 Primary Efficacy Endpoints

Estimand

The primary estimand has the following attributes:

- Population: all randomized subjects who received at least one dose of study drug and with baseline pain assessments and at least one post-baseline pain assessment.
- Endpoint: change from baseline to week 4 in the average of the daily average Numeric Rating Scale (NRS) pain intensity scores
- Intercurrent events: premature discontinuation from the study, the hypothetical estimand is to evaluate the pharmacological effect as if no withdrawals had occurred.
- Treatment condition: one of the randomized treatment groups of the following per protocol: 1600 mg total dose (provided as 2×800 mg tablets), 1600 mg total dose (provided as 4×400 mg tablets), or placebo.
- Population-level summary: difference in mean of endpoint between treatment conditions.

Statisitical Hypotheses

The primary endpoint is the change from baseline to week 4 in the average of the daily average Numeric Rating Scale (NRS) pain intensity scores.

The mean change from baseline to week 4 daily average NRS score will be compared between each of the ACP-044 dose groups (400 mg QID, 800 mg BID) to the placebo group.

The null hypothesis is the mean change from baseline to week 4 in the average of the daily average Numeric Rating Scale (NRS) pain intensity scores using ACP-044is equal to that of placebo in subjects with pain associated with osteoarthritis of the knee.

The alternative hypothesis is the mean change from baseline to week 4 in the average of the daily average Numeric Rating Scale (NRS) pain intensity scores using ACP-044 is not equal to that of placebo in subjects with pain associated with osteoarthritis of the knee.

Primary Estimator

The primary efficacy variable will be analyzed using an analysis of covariance model (ANCOVA) based on the FAS with treatment, sex, and baseline K-L score (randomization strata) as factors and baseline weekly average of daily average pain score as a covariate in the model. The primary comparison will be between each ACP-044 dose group versus the

placebo group. The comparison between the average of the means of the two ACP-044 dose groups versus the mean of placebo group will also be conducted. Missing data will be imputed using multiple imputation based on missing at random assumption (MAR) with an exception of missing data from subjects who are treated with ACP-044 and discontinue the study due to lack of efficacy or adverse events. Their missing data will be imputed using the data from placebo group.

13.1.2 Sensitivity Analyses

No sensitivity analyses are planned for primary efficacy endpoint.

13.1.3 Supportive Analyses

No supportive analyses are planned for primary efficacy endpoint.

13.2 Key Secondary Efficacy Endpoints

Patient Global Impression of Change (PGIC) at Week 4 with reference to baseline status

The PGIC will be summarized as both a categorical and continuous variable (including 95% CI for the mean) at every study visit week. This endpoint will be analyzed using analysis of variance (ANOVA) with treatment, sex, and baseline K-L score as factors on LOCF data. The comparison at week 4 will be viewed as key secondary.

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, sex and baseline K-L score as factors. Additionally, this proportion will be compared using a Cochran-Mantel-Haenzel test stratified by baseline K-L score and sex.

Change from Baseline to Week 4 in Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) total score

The change from baseline in the WOMAC Index total score will be analyzed using analysis of covariance (ANCOVA) with baseline WOMAC score as covariate, treatment, sex, and baseline K-L score as main effects on LOCF data.

Proportion of subjects who are responders as defined by 30% and 50% reduction from Baseline to Week 4 in the average of the daily average NRS pain intensity scores will be evaluated by the Cochran-Mantel-Haenszel (CMH) General Association Test, missing data considered as "non-respond", controlling for baseline K-L score and sex.

13.3 Secondary Efficacy Endpoints

Change from Baseline to every study week visit in:

Weekly average of the daily worst NRS pain intensity scores will be evaluated using ANCOVA with baseline SF-12 score as covariate, treatment, baseline K-L score and sex as main effects on LOCF data.

Brief Pain Inventory–Short Form (BPI-sf) will be evaluated using ANCOVA with baseline SF-12 score as covariate, treatment, baseline K-L score and sex as main effects on LOCF data.

12-item Short Form Survey (SF-12) The change from baseline to week 4 will be evaluated using ANCOVA with baseline SF-12 score as covariate, treatment, baseline K-L score and sex as main effects on LOCF data.

13.4 Exploratory Efficacy Endpoints

Comparison of average daily rescue medication use

The average daily number of acetaminophen tablets used during the study will be compared between treatments using a Wilcoxon Rank Sum Test stratified by baseline K-L score and sex.

Hospital Anxiety and Depression Scale (HADS) The change from baseline to week 4 will be evaluated using ANCOVA with baseline HADS score as covariate, treatment, baseline K-L score, sex as main effects on LOCF data.

14 SAFETY ANALYSES

All safety analyses will be performed using the actual treatment for the Safety Analysis Set.

14.1 Adverse Events

Adverse events will be coded using MedDRA dictionary, Version 24.0 or newer.

An AE (classified by preferred term) will be considered a treatment-emergent AE (TEAE) if started after first dose administration and no later than last dose date + 30 days. AEs reported on Day -1 based on Baseline (pre-dose) findings (e.g., clinically significantly abnormal vital signs, laboratory test results, or electrocardiogram parameters) will not be considered as TEAEs.

The event counts, the number, and percentage of subjects reporting TEAEs in each treatment group will be tabulated by system organ class (SOC) and preferred term; by SOC, preferred term, and maximum severity; and by SOC, preferred term, and relationship to study drug. If more than one TEAE occurs with the same preferred term for the same subject, then the subject will be counted only once for that preferred term using the most severe and most related occurrence for the summarization by severity and by relationship to study drug. The display in these tables will be sorted alphabetically by SOC and then by descending subject frequency for the preferred terms (counts from ACP-044 treated group) within each SOC.

The event counts, the number and percentage of subjects with any TEAEs will also be tabulated by preferred term without SOCs. This table will be sorted by descending subject frequency within ACP-044 treated group.

The incidence of most frequently reported (preferred terms reported by ≥ 5% of subjects in any treatment group) TEAEs, SAEs reported after treatment start, TEAEs leading to discontinuation of study drug, and TEAEs related to study drug will be summarized by SOC, preferred term, and treatment group. The tables will be sorted alphabetically by SOC and then by descending frequency within each SOC in ACP-044 treated group. In addition, the incidence of fatal treatment-emergent AEs (i.e., events that cause death) will be summarized separately by SOC and preferred term.

The incidence of adverse events of interest, will be summarized separately by SOC and preferred term. The adverse events of interest evaluation will be based on the following prespecified preferred terms of aspartate aminotransferase increased, alanine aminotransferase increased, gamma-glutamyltransferase increased, blood alkaline phosphatase increased, blood bilirubin increased, hyperbilirubinaemia, liver function test increased, blood creatine phosphokinase increased, rash, hypotension, blood pressure decreased , presyncope, hypotension, orthostatic hypotension, pre-syncope, syncope, and bradycardia.

These summary tables except for the most frequently reported TEAEs tables may also be presented for COVID-19 and non-COVID-19 related events, if applicable.

An AE listing by subject will display all events, including those which occur during screening, and will include the verbatim term in addition to the MedDRA SOC and preferred term. This listing will also include all relevant eCRF data associated with the event: date of onset, date resolved, date of last dose, severity, frequency, outcome, relationship to study drug, and action taken with study drug. A listing by subject will display all TEAE, excluding AE that are not treatment-emergent. Separate listings will be presented for subjects with treatment-emergent SAEs, related TEAEs, TEAEs leading to discontinuation, fatal TEAEs (if any), and COVID-19 related TEAEs (if applicable).

14.2 Clinical Laboratory Variables

Clinical laboratory tests are performed at Baseline and Week 4 (EOT/ET).

The laboratory evaluations will include, but are not limited to, the following:

- Clinical chemistry serum tests (CHEM)
    - Sodium (Na), potassium (K), chloride (Cl), phosphorus (P), calcium (Ca),carbon dioxide (CO2), blood urea nitrogen (BUN), creatinine (CR), uric acid,
    - Estimated glomerular filtration rate (eGFR) should only be performed at Visit 1 (Screening)

  - Alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transpeptidase (GGT), alkaline phosphatase (ALP), total bilirubin (TBIL), lactate dehydrogenase (LDH)
    - INR to be evaluated, as needed
  - Creatine kinase (CK)/creatine phosphokinase (CPK)
  - Lipid panel should only be performed at Visit 1 (Screening):
    - Total cholesterol, HDL-cholesterol, triglycerides, LDL-cholesterol, cholesterol/HDL ratio, non-HDL cholesterol; should only be performed at Visit 1 (Screening)
  - HbA1c should only be performed at Visit 1 (Screening)
  - Glucose
  - Albumin (ALB) should only be performed at Visit 1 (Screening)
  - Total protein should only be performed at Visit 1 (Screening)
- Hematology tests
  - Complete blood count (CBC) including:
    - White blood cell (WBC) count
    - Complete differential (relative and absolute)
    - Hematocrit (Hct), hemoglobin, red blood cells (RBC), platelets
    - Reticulocyte count
- Endocrinology
  - Thyroid stimulating hormone (TSH) and free T4 (Screening)
- Pregnancy test
  - A serum pregnancy test should only be performed at screening for women of childbearing potential
  - A urine pregnancy test should be performed at Baseline (Week 0) and Week 4 (EOT)/ET for women of child-bearing potential
    - If urine cannot be obtained in women of childbearing potential, a serum pregnancy test should be done in its place
- COVID-19 test:

  - COVID-19 diagnostic PCR test at Screening

- Urinalysis (UA)
  - Blood, RBCs, WBCs, protein, glucose, ketones, specific gravity, pH, nitrates
- Urine drug screen (UDS) with reflex testing
  - A urine drug screen will test for controlled substances at Screening, Baseline Pain Assessment Period, and at Week 4/EOT/ET. A dipstick urine drug screen will be performed at Baseline (Week 0) for rapid detection of controlled substances and for certain prescription medications in order to determine if the subject continues to meet subject eligibility. The following controlled substances will be tested for with a urine drug screen and/or urine dipstick according to the schedule presented in Table 6–1 in protocol: amphetamine, barbiturates, benzodiazepines, buprenorphine, cocaine, ecstasy (MDMA), methadone, morphine/opiates, methamphetamine, marijuana (THC), phencyclidine (PCP), propoxyphene, tricyclic antidepressants. Negative drug screens are required for study eligibility.
    - If there is a positive urine drug screen for an analgesic medication that is not allowed per protocol, the subject must washout of medication, if appropriate, prior to the Baseline Pain Assessment Period.
    - If there is a positive urine drug screen for a medication that is allowed per protocol (e.g., metformin leading to a false positive for amphetamine), the subject must have a valid prescription for such medication, and if the subject denies use of the positive prohibited medication (e.g., amphetamine), they can be included in the study and it will not be considered a protocol deviation.

Clinical laboratory values (in Système International [SI] units) and the change from Baseline values will be summarized by treatment group at each post-Baseline visit using descriptive statistics. The overall minimum, maximum as well as the last post-Baseline observed and change from Baseline values will also be summarized. For urinalysis with categorical results, the number and percentage of subjects will be tabulated by category at Baseline and each post-Baseline visit, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group.

Laboratory values will also be summarized in shift tables by treatment group, to determine the number and percentage of subjects with values classified as below, within, and above normal ranges at each post-Baseline visit relative to the same classification at the Baseline

visit. For the by-visit shift summary, the denominator is the number of subjects with non-missing values at Baseline and the given visit for the given parameter and treatment group. For the shift to the overall post-Baseline minimum or maximum, all post-Baseline values will be considered, including unscheduled and out of window values and the denominator is the number of subjects with non-missing Baseline value and at least 1 post-Baseline value for the given parameter and treatment group.

Clinical laboratory values are potentially clinically important (PCI) if they meet either the low or high PCI criteria listed in Table 2 and Table 3. The number and percentage of subjects with post-Baseline PCI values for each of the categories in Table 3 and Table 4 will be summarized by treatment group for selected parameters. For the overall post-Baseline summaries of PCI values, all post-Baseline values will be considered, including unscheduled and out of window values. Subjects with multiple PCI values for a given parameter will be counted only once for that parameter. For the overall post-Baseline summary, the numerator of the percentage is the number of subjects with at least 1 post-Baseline PCI laboratory value for the given parameter and treatment group, and the denominator is the number of subjects with at least 1 post-Baseline laboratory value for the given parameter and treatment group.

Table 2 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry

Analyte	Conventional Unit	Low PCI Criteria	High PCI Criteria	SI Unit	Low PCI Criteria	High PCI Criteria
Hematology (whole blood)						
Hemoglobin (male)	g/dL	<11	>18	g/L	<110	>180
Hemoglobin (female)	g/dL	<10	>17	g/L	<100	>170
Hematocrit (male)	%	<30	>55	L/L	<0.3	>0.55
Hematocrit (female)	%	<30	>50	L/L	<0.3	>0.5
Leukocyte (White Blood Cell Count)	x 10^3/uL	≤2.8	≥15	x 10^9/L	≤2.8	≥15
Neutrophils	x 10^3/uL	≤1.5	No upper limit	x 10^9/L	≤1.5	No upper limit
Platelet Count	x 10^3/uL	≤75	≥700	10^9/L	≤75	≥700
Chemistry (serum or plasma)						
ALT (SGPT)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
AST (SGOT)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Total Bilirubin	mg/dL	No lower limit	≥1.5 ULN	umol/L	No lower limit	≥1.5 ULN
BUN	mg/dL	No lower limit	≥30.0	mmol/L	No lower limit	≥10.71
Creatine Kinase (CK)	U/L	No lower limit	≥3 ULN	U/L	No lower limit	≥3 ULN

Table 2 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry (Continued)

Analyte	Conventional Unit	Low PCI Criteria	High PCI Criteria	SI Unit	Low PCI Criteria	High PCI Criteria
Sodium	mEq/L	≤125	≥155	mmol/L	≤125	≥155
Potassium	mEq/L	≤3.0	≥5.5	mmol/L	≤3.0	≥5.5
Calcium, total	mg/dL	<8.0	>11.0	mmol/L	<2.0	>2.75
Lactate Dehydrogenase (LDH)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Alkaline Phosphatase	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Uric acid (male)	mg/dL	No lower limit	≥10.5	umol/L	No lower limit	≥624.75
Uric acid (female)	mg/dL	No lower limit	≥8.5	umol/L	No lower limit	≥505.75
Albumin	g/dL	≤2.6	≥6.0	g/L	≤26	≥60
Total Protein	g/dL	≤5.0	≥10.0	g/L	≤50	≥100
Chloride	mEq/L	≤85	≥120	mmol/L	≤85	≥120
Glucose (random)	mg/dL	≤45.1	≥200.0	mmol/L	≤2.48	≥11
Serum Creatinine	mg/dL	Not Applicable	>1.5 ULN	umol/L	Not Applicable	>1.5 ULN
Triglycerides	mg/dL	Not Applicable	>300	mmol/L	Not Applicable	>3.39
Gamma-Glutamyl Transferase (GGT)	U/L	Not Applicable	≥3 ULN	U/L	Not Applicable	≥3 ULN

Table 3 Criteria for Potentially Clinically Important Laboratory Values - Urinalysis

Urinalysis (qualitative dipstick)	Low PCI Criteria	High PCI Criteria
Blood (occult blood)	Not Applicable	≥ Moderate
Protein	Not Applicable	≥ 100 mg/dL
Glucose	Not Applicable	≥ 500 mg/dL

Clinical laboratory data will be displayed in data listings with date and study day of collection. All units will be displayed according to SI conventions for units. Out of range values will be flagged in the data listings (i.e., 'L' or 'H'). A separate listing will be provided for a subset of the chemistry, hematology, and urinalysis analytes with values classified as PCI.

The pregnancy results (positive or negative) for female subjects will be presented in a listing.

In addition, the following additional PCI criteria for liver function tests will be flagged, listed, and summarized:

- ALT (SGPT) ≥ 2x ULN U/L

- AST (SGOT) ≥ 2x ULN U/L
- Gamma Glutamyl Transferase ≥ 2x ULN U/L
- Alkaline Phosphatase ≥ 2x ULN IU/L
- Creatine Kinase (CK) ≥ 5x ULN

14.3 Vital Signs

Vital signs will include body temperature, resting respiration rate, supine and standing systolic and diastolic blood pressure, and pulse rate. Blood pressure and pulse rate measurements will be taken after the subject rests for 5 minutes in the supine position. The subject will then be asked to stand and blood pressure and pulse rate will be taken after the subject stands for 1 minute and then again after 3 minutes. Subjects unable to stand may be assessed while sitting upright. A drop in SBP of ≥20 mmHg or in DBP of ≥10 mmHg or if the subject is experiencing lightheadedness or dizziness is considered abnormal and orthostasis should be considered.

Vital signs will be evaluated at all designated study visits. A 15-minute window is permitted for vital sign measurements.

Vital signs will be summarized using descriptive statistics at Baseline and all scheduled post-Baseline visits/timepoints. The change from Baseline values will also be summarized at the scheduled post-Baseline visits/timepoints.

Vital sign values will be considered PCI if they meet the criteria listed in Table 4. The number and percentage of subjects with post-Baseline vital signs that are PCI will be summarized by treatment group at each post-Baseline visit and for overall post-Baseline. For the overall post-Baseline summaries, all post-Baseline values will be considered, including unscheduled and out of window values. Subjects with multiple PCI values for a given parameter will be counted only once for that parameter. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-Baseline PCI vital sign for the given parameter, visit and treatment group, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. For the overall post-Baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-Baseline PCI vital sign for the given parameter and treatment group, and the denominator is the number of subjects with at least 1 post-Baseline vital sign for the given parameter and treatment group. A listing of overall and of subjects with any PCI vital sign values will be provided.

Table 4 Criteria for Potentially Clinically Important Vital Signs

Vital Sign Parameter	Unit	Criteria[a]		
		Observed Value	And/Or	Change Relative to Baseline
Systolic blood pressure (supine)	mmHg	≥180	And	Increase of ≥20
		≤90	And	Decrease of ≥20
Systolic blood pressure (standing or sitting)	mmHg	≥180	And	Increase of ≥20
		≤90	And	Decrease of ≥20
Diastolic blood pressure (supine)	mmHg	≥105	And	Increase of ≥15
		≤50	And	Decrease of ≥15
Diastolic blood pressure (standing or sitting)	mmHg	≥105	And	Increase of ≥15
		≤50	And	Decrease of ≥15
Systolic blood pressure (supine to standing)	mmHg			Decrease ≥ 20 mmHg
Diastolic blood pressure (supine to standing)	mmHg			Decrease ≥ 10 mmHg
Pulse (supine)	bpm	≥120	And	Increase of ≥15
		≤50	And	Decrease of ≥15
				Decrease of ≥7%
Pulse (standing or sitting)	bpm	≥120	And	Increase of ≥15
		≤50	And	Decrease of ≥15
				Decrease of ≥7%

[a] A post-baseline value is considered as a PCI value if it meets both criteria for observed value and change from baseline.

In addition, the following additional PCI criteria for orthostatic hypotension, significant bradycardia will be flagged, listed, and summarized:

- Orthostatic hypotension: decrease in systolic blood pressure (SBP) ≥20 mmHg or diastolic blood pressure (DBP) ≥10 mmHg after transitioning from the supine position (after resting for 5 minutes) to standing at 1 or 3 minutes with or without symptoms such as light headedness or dizziness and not explained by dehydration, illness, or medications.Significant bradycardia: defined as heart rate of ≤45 beats per minute (BPM) on an ECG, or pulse decrease of ≥25% from Baseline with resulting heart rate < 60 BPM.

14.4 Electrocardiogram (ECG)

12-lead ECGs are collected at Baseline (Week 0) both predose and 1 hour postdose (the first dose of the day), Week 1, Week 2, and Week 4 (EOT)/ET. Observed values of ECG variables (e.g., heart rate, PR interval, QRS interval, QT interval, and QTc interval) and the changes

from Baseline at each assessment time point will be summarized by treatment group.

QTcB and QTcF will also be categorized into the following categories (msec) and the number and percentage of subjects in each category will be summarized by treatment group at each visit and for the overall post-Baseline maximum:

- Observed: ≤450, 451 - ≤480, 481 - ≤500, and >500; >450; >480.
- Change from Baseline: ≤10, 11 - 30, 31 - 60, and >60; >30.
- QTcF result >470 ms for males, >480 ms for females.
- QTcF increases ≥60 ms from Baseline.

For cardiologist's interpretations, the number and percentage of subjects with ECG results that are determined as normal or abnormal will be summarized at scheduled visits. The overall post-baseline worst interpretation will also be summarized (i.e. if a subject has one or more post-baseline ECG results that is/are considered as abnormal, this subject will be counted in the abnormal category). Cardiologist's interpretations will also be summarized in a shift table. The shifts from Baseline to overall post-Baseline worst interpretation will also be presented. For the by-visit shift summary, the denominator is the number of subjects with non-missing cardiologist's interpretation at Baseline and the given visit for the given treatment group. For the summaries of shift from Baseline to the overall post-Baseline worst interpretation, the denominator is the number of subjects with non-missing Baseline and at least 1 post-Baseline cardiologist's interpretation for the given treatment group.

Electrocardiogram variable values will be considered PCI if they meet the criteria listed in Table 5. The number and percentage of subjects with post-baseline PCI values will be summarized by treatment group at each post-Baseline visit and for overall post-Baseline. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-Baseline PCI ECG for the given parameter, visit and treatment group, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. For the overall post-Baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-Baseline PCI ECG for the given parameter and treatment group, and the denominator is the number of subjects with at least 1 post-Baseline ECG value for the given parameter and treatment group. A listing of overall and of all subjects with any PCI ECG values will be provided.

Table 5 Criteria for Potentially Clinically Important ECG Values

ECG Parameter	Unit	High PCI Criteria
QRS Interval	msec	≥120
PR Interval	msec	≥220
QTcB or QTcF	msec	>500
QTcB or QTcF: change from baseline	msec	>60

14.5 Physical Examination

Physical examinations are performed at Baseline (Week 0), Week 4 (EOT)/ET. Physical examination results (normal, abnormal, and not done) will be summarized in a frequency table by treatment group, body system and visit. A listing of overall physical examination data will be listed.

14.6 Other Safety Variables

14.6.1 Suicidality Based on C-SSRS

The event counts and the number and percentage of subjects reporting any post Baseline suicidal ideation, suicidal behavior, or suicidality will be tabulated by treatment group. The event counts and the number and percentage of subjects reporting any post-Baseline non-suicidal self-injurious behavior will also be tabulated. For calculating the percentages, the denominator will be the number of subjects in the Safety Analysis Set.

15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

15.1 Pharmacokinetic Endpoints

The PK endpoints of the study are:

- Plasma concentrations of ACP-044 using sparse sampling on the following study days: Baseline (Week 0), Week 1, Week 2, and Week 4 (EOT)/ET
- ACP-044 PK parameters determined using a population PK approach

Plasma concentration of ACP-044 will be summarized by treatment group using the Pharmacokinetic Analysis Set.

PK parameter summary and analysis will be discussed in a separate plan.

15.2 Pharmacokinetic/Pharmacodynamic Endpoints

- ACP-044 exposure-response relationship:
    - for efficacy using the primary endpoint and selected secondary endpoints

- for safety using selected TEAEs

PK/PD endpoint summary and analysis will be discussed in a separate plan.

16 INTERIM ANALYSIS

No interim analyses are planned for this study.

17 DATA MONITORING/REVIEW COMMITTEE

There is no data monitoring/review committee in this study.

18 COMPUTER METHODS

Statistical analyses will be performed using Version 9.4 (or newer) of SAS® (SAS® Institute, Inc., Cary, North Carolina) on a suitably qualified and validated environment.

Validation and quality control of the tables, listings and figures containing the results of the statistical analyses will follow appropriate standard operating procedures.

19 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

Due to early termination of the ACP-044 program, the changes to the planned analyses are:

- Primary efficacy endpoint will be analyzed using an analysis of covariance model (ANCOVA) based on the FAS with treatment, sex, and baseline K-L score (randomization strata) as main factors and baseline weekly average of daily average pain score as a covariate in the model. The missing data will be imputed using baseline observed data carried forward (BOCF).
- Key secondary efficacy endpoints, the observed data will be provided in descriptive statistics including point estimates and listings.
- Secondary efficacy endpoints, the observed data will be provided in descriptive statistics and listings, no statistical inferences will be applied.
- Exploratory efficacy endpoints, the observed data will be provided in descriptive statistics and listings, no statistical inferences will be applied.

20 REFERENCES

Centers for Disease Control and Prevention. 2018 Annual Surveillance Report of DrugRelated Risks and Outcomes — United States. Surveillance Special Report 2. Centers for Disease Control and Prevention, U.S. Department of Health and Human Services. Published August 31, 2018.

Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function. Arthritis Care Res. 2011;63(S11):S208-S228.

FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic Guidance for Industry, Investigators, and Institutional Review Boards (March 2020)

ICH E9(R1): Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials, Adopted on November 2019

Points to consider on implications of Coronavirus disease 4 (COVID-19) on methodological aspects of ongoing clinical 5 trials, March 25 2020.

Rubin, D. B. Inference and Missing Data. Biometrika. 1976; 63:581-92.

Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983;67(6):361-370

21 APPENDICES

Appendix A Schedule of Events and Assessments

	Screening Period	Baseline Pain Assessment Period	Double-Blind Treatment Period							Safety Follow-up Period	
										Clinic visit	Telephone call
Visit Day/Week	**Days -35 to -8**	**Days -7 to -1**	**Baseline Week 0**	**Day 3**	**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Unscheduled Visit[r]**	**7 days after EOT**	**30 days after last dose**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8 (EOT/ET)**			
Visit window (# days)		**+3**	**0**	**±1**	**±3**	**±3**	**±3**	**±3**		**±3**	**+4**
Informed consent	X										
Inclusion/exclusion criteria	X	X	X								
Medical, medication, and surgical history and demographics	X										
Physical examination	X		X					X			
Vital signs (including orthostatic changes)[a]	X	X	X		X	X		X	X		
Height, weight, and BMI calculation[b]	X		X					X			
12-lead ECG[c]	X		X[c]		X	X		X			
Clinical laboratory tests	X		X	X[d]	X	X	X[d]	X			
Pregnancy test[e]	X		X					X			
Urine drug screen	X	X	X[f]					X			
COVID-19 test[g]	X										
C-SSRS	X	X	X		X	X		X	X		
Beck Depression Inventory–II	X										
Verified Clinical Trials Subject Registry	X										
Central Imagingh	X										
Issue eDiary and train on usei		X									
Collect eDiary								X			

Appendix A Schedule of Events and Assessments (Continued)

	Screening Period	Baseline Pain Assessment Period	Double-Blind Treatment Period							Safety Follow-up Period	
										Clinic visit	Telephone call
Visit Day/Week	**Days -35 to -8**	**Days -7 to -1**	**Baseline Week 0**	**Day 3**	**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Unscheduled Visit[r]**	**7 days after EOT**	**30 days after last dose**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8 (EOT/ET)**			
Visit window (# days)		**+3**	**0**	**±1**	**±3**	**±3**	**±3**	**±3**		**±3**	**+4**
Randomization			X								
PK blood sample collection[j]			X[k]		X[l]	X[l]		X[l]			
eDiary completion: Numeric Rating Scale Daily Pain Score (Average and Worst)[m]		X	X	X	X	X	X	X		X	
PGIC[n]					X	X		X			
WOMAC[n]			X		X	X		X			
BPI-sf[n]			X		X	X		X			
SF-12			X					X			
HADS	X		X					X			
Participant Scorecard[o]					X	X		X			
Accurate Symptom Reporting (ASR) Training – subject facing	X		X		X[p]	X[p]		X[p]			
Placebo Response Reduction (PRR) Training – subject facing	X		X		X[p]	X[p]		X[p]			
Study drug dispensing/accountability			X		X	X		X			
eDiary completion: Rescue medication dispensing/accountability[q]		X	X	X	X	X	X	X			
Prior and concomitant medications	X	X	X	X	X	X	X	X	X	X	X
Assessment of adverse events	X	X	X	X	X	X	X	X	X	X	X

Abbreviations: ASR=accurate symptom reporting; BMI=body mass index; BPI-sf=Brief Pain Inventory–Short Form; COVID-19=coronavirus disease 2019; C-SSRS=Columbia–Suicide Severity Rating Scale; ECG=electrocardiogram; eDiary=electronic diary; EOT=end of treatment; ET=early termination; HADS=Hospital Anxiety and Depression Scale; PGIC=Subject Global Impression of Change; PK=pharmacokinetic; PRR=placebo response reduction; SF-12=12-item Short Form Survey; WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index

[a] Vital signs will include body temperature, resting respiration rate, supine and standing systolic and diastolic blood pressure, and pulse rate. Blood pressure and pulse rate will be measured after the subject rests for 5 minutes in the supine position. The subject will then be asked to stand and blood pressure and pulse rate will be taken after standing for 1 minute and then again after 3 minutes. Subjects unable to stand may be assessed while sitting upright. The same position and arm should be used each time vital signs are measured for a given subject.

[b] Height will only be measured at the Screening visit, and BMI will be calculated based on height and weight measurements at Screening. Weight will also be measured at any unscheduled visit.

[c] A single 12-lead ECG will be performed at Screening, at Baseline (Week 0) both predose and at 1 hour postdose (the first dose of the day), Week 1, Week 2, and EOT/ET. A 1-hour window is permitted for the predose ECG and a 20-minute window is permitted for the postdose ECGs. Electrocardiograms should be performed before blood sampling or at least 30 minutes after blood sampling.

[d] A full chemistry (CHEM) panel of clinical laboratory tests will be completed, excluding glucose and the CHEM tests noted in Section 6.5.5 to only be performed at Screening.

[e] Applicable to women of childbearing potential. A serum pregnancy test should be performed at Screening and a urine test must be performed at all other designated visits.

[f] At Baseline, urine drug screen will be performed using a dipstick test which must be negative for any non-permitted medications or drugs in order for the subject to continue in the study.

[g] COVID-19 diagnostic PCR test at Screening.

[h] Randomization visit cannot occur until there is confirmation from the central reader that there are no exclusionary findings.

[i] Subjects will be instructed on the use of and will complete electronic clinical outcome assessments using handheld devices during the study. An eDiary, will be used to record pain scores, compliance with ingestion of study medication, and use of rescue medication.

[j] A PK sample (approximately 4 mL) should also be taken from subjects who experience SAEs or AEs leading to discontinuation, as soon as possible after the occurrence of the event.

[k] A predose PK blood sample must be collected before administration of study drug. A postdose PK blood sample will be collected at the end of ECG assessment approximately 1 hour after study drug administration.

[l] PK samples at Visits 5, 6, and 8 will be collected at one of the following time intervals: 1) 2-5 hours after morning dosing OR 2) 6-9 hours after morning dosing OR 3) 10-12 hours after morning dosing. Every effort should be made to collect the PK samples at distinct time intervals during Visits 5, 6, and 8. However, if the interval is the same across these visits, then the collection time should vary within that interval.

[m] Pain intensity assessments are to be recorded in the evening during the Baseline Assessment Period and prior to their 4th dose of the day during the Double-blind period. Subjects will be rating both the average daily pain as well as their worst daily pain. At the Safety Follow-up Clinic Visit (7 days after EOT), subjects will rate both the average daily pain as well as their worst daily pain over the past week and the data will be entered into EDC.

[n] At study visits, PGIC, WOMAC, and BPI-sf should be performed before blood sample collection, ECG, and other procedures.

[o] Participant Scorecard is viewed by the subject and reviewed by the Investigator at the designated study visits. The Investigator or designee, will acknowledge the scorecard review on the training platform, aLearn.

[p] ASR and PPR training at Visits 5, 6, and/or 8 is on an as needed basis and the need for training will be determined by Participant Scorecard intervention recommendations.

[q] Record rescue medication use in eDiary.

[r] The Investigator may perform any additional evaluations deemed to be clinically indicated. If an unscheduled visit is required for safety reasons, a 12-lead ECG and all appropriate clinical laboratory tests are to be performed.

Appendix B Summary of Version Changes

Version No:	Document History Description of Update	Author(s)	Version Date
1.0	Original Version	[redacted]	21 September 2022