Protocol B3741002

A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study Comparing the Analgesic Efficacy and Safety of a Single Oral Dose of a Novel Fixed-Dose Combination of Ibuprofen 400 Mg with Caffeine 100 Mg to Ibuprofen 400 Mg and to Placebo in the Treatment of Post-Surgical Dental Pain in Otherwise Healthy Subjects

Statistical Analysis Plan
(SAP)

Version: 1

Date: 23-Mar-2018

[redacted]

TABLE OF CONTENTS

LIST OF TABLES 4
1. VERSION HISTORY 5
2. INTRODUCTION 5
2.1. Study Objectives 6
2.2. Study Design 6
3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS 10
3.1. Primary Endpoint(s) 10
3.2. Secondary Endpoint(s) 10
CCI [redacted]
3.4. Baseline Variables 13
3.4.1. Stratification Variables 14
3.4.2. Covariates 14
3.5. Safety Endpoints 14
3.5.1. Adverse Events 14
3.5.2. Laboratory Data 15
3.5.3. Vital Signs 15
CCI [redacted]
4. ANALYSIS SETS 16
4.1. Full Analysis Set 16
4.2. Per Protocol Analysis Set 16
4.3. Safety Analysis Set 16
4.4. Other Analysis Sets 16
5. GENERAL METHODOLOGY AND CONVENTIONS 16
5.1. Hypotheses and Decision Rules 16
5.2. General Methods 17
5.2.1. Analyses for Continuous Data 17
CCI [redacted]
CCI [redacted]
5.2.4. Analyses for Time to Event Data 18
5.3. Time Windows and Methods to Manage Missing Data 18
[redacted]

5.3.1. Time Windows....18
5.3.2. Methods to Manage Missing Time to Event Data....19
6. ANALYSES AND SUMMARIES....19
6.1. Primary Endpoint: SPRID 0-8....20
6.1.1. Primary Analysis....20
6.1.1.1. Sensitivity/Robustness Analyses....20
6.2. Secondary Endpoint(s)....21
6.2.1. SPRID 0-t, SPID 0-t, TOTPAR 0-t, PRID, PRR, and PID....21
6.2.2. Time to Onset of "Meaningful" Relief, Time to Onset of "First Perceptible" Relief (Confirmed by "Meaningful" Relief), and Duration of Relief....22
CCI [redacted]
6.4. Subset Analyses....23
6.5. Baseline and Other Summaries and Analyses....23
6.5.1. Baseline Summaries....24
6.5.1.1. Pain Evaluations....24
6.5.1.2. Demographic Data....24
6.5.1.3. Background Information....24
6.5.1.4. Tooth Extraction and Surgery Details....25
6.5.1.5. Baseline and Background Endpoints for Subset Analyses....25
6.5.2. Study Conduct and Subject Disposition....25
6.5.3. Concomitant Medications and Non-Drug Treatments....25
6.6. Safety Summaries and Analyses....25
6.6.1. Adverse Events....25
6.6.2. Laboratory Data....26
6.6.3. Vital Signs....26
7. INTERIM ANALYSES....26
8. REFERENCES....27
9. APPENDICES....28

[redacted]

LIST OF TABLES

Table 1 Summary of Major Changes in SAP Amendments 5
Table 2 Schedule of Activities 8
Table 3 Summary of Efficacy Analyses 28
Table 4 Time Window for Scheduled Time Points 29

APPENDICES

Appendix 1. SUMMARY OF EFFICACY ANALYSES 28
Appendix 2. DATA DERIVATION DETAILS 29
Appendix 2.1. Definition and Use of Visit Windows in Reporting 29
Appendix 2.2. Definition of Endpoints NPSR, PRR, PID and PRID 30
Appendix 3. STATISTICAL METHODOLOGY DETAILS 31
Appendix 4. TARGET MEDICAL EVENT (TME) LIST 31

[redacted]

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for study B3741002 is based on the protocol dated 13JUL2016.

Table 1 Summary of Major Changes in SAP Amendments

SAP Version	Change	Rationale
1	Not Applicable	Not Applicable

Note: In this document any text taken directly from the protocol is *italicized*.

2. INTRODUCTION

Ibuprofen (IBU) is one of the most widely used nonprescription analgesic/antipyretic agents available. Serum concentrations of IBU have demonstrated a linear relationship to the level of analgesia. The 400 mg strength of IBU is considered to be the most efficacious over-the-counter (OTC) dose of IBU in terms of efficacy and safety when compared to other available OTC NSAIDs.

Caffeine, a methylxanthine, exerts a variety of pharmacological actions both peripherally and centrally in the body. Generally, when caffeine is consumed in foods and drinks, the intake of caffeine can be classified into three broad categories: low (<100 mg per day), moderate (100-400 mg per day), and high intake (>400 mg per day), with a majority of individuals falling into the moderate consumption category. Caffeine is also available as an oral pharmacologic agent at doses of 100-200 mg every 3-4 hours as needed for somnolence.

Clinical studies have demonstrated that the combination of NSAIDs with caffeine is more effective than the NSAID alone in postpartum, headache, and postsurgical (dental) pain states. Forbes et al, intended to quantify the relative incremental efficacy of the addition of caffeine 100 mg with IBU 100 mg and 200 mg using the postsurgical dental pain model. The results demonstrated the combination was 2.4 to 2.8 times more effective than IBU alone, and that the combination provides more rapid onset and longer duration of analgesic effect than IBU alone. Subsequent studies of the combination of IBU with caffeine have shown that it may provide similar enhancement of analgesic effect versus the IBU alone and placebo in dental pain, post-episiotomy pain, tension headache, migraine headache, and dysmenorrhea pain.

Pfizer Consumer Healthcare (PCH) has developed a novel fixed-dose liquid-filled gelatin capsule (subsequently referred to as liquid capsule) formulation consisting of 400 mg IBU with 100 mg of caffeine. The aim of this clinical study is to evaluate the relative analgesic efficacy of this novel formulation of IBU 400 mg with 100 mg of caffeine in a liquid capsule

[redacted]

compared to IBU liquid capsule alone, providing consumers with a novel more effective formulation of IBU 400 mg for extra-strength pain relief.

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B3741002. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives

The primary objective of this study is to assess analgesic efficacy of a single oral dose of IBU 400 mg plus caffeine 100 mg, versus IBU 400 mg alone, and placebo, in subjects undergoing dental surgery for third molar extraction.

The secondary objective of the study is to assess the safety and tolerability of IBU 400 mg with caffeine 100 mg, IBU 400 mg, and placebo.

2.2. Study Design

This will be a 3-arm, randomized, double-blind, placebo-controlled, parallel group, singlecenter, 8-hour inpatient study. Approximately 385 adolescent and adult subjects will be enrolled to ensure 346 complete the study.

Subjects will be healthy males and females aged 16-40 years, inclusive, who are experiencing post-operative pain following surgical extraction of 3 or more third molars, of which at least 2 must be a partial or complete bony mandibular impaction. Following extraction, subjects will rest quietly at the study center until they experience post-surgical pain severity of ≥5 on the 0-10 point numeric pain severity rating (NPSR) scale measuring pain severity, confirmed by a Visual Analog Pain Severity Rating Scale (VAS PSR) of at least 50 mm on a 100 mm VAS PSR scale. Upon completion of the baseline scales, eligible subjects will be randomized to receive a single oral dose of study medication under double-blind conditions and will be evaluated on-site for 8 hours following administration of study medication.

Subjects will be randomized to receive one of the following treatments in a 3:3:1 ratio:

- *IBU 400 mg with caffeine 100 mg fixed-dose combination liquid-filled gelatin capsule;*
- *Advil® Extra Strength Liqui-Gels®, IBU 400 mg liquid-filled gelatin capsule (currently marketed product in Canada);*
- *Placebo (unprinted liquid filled gelatin capsule).*

Once subjects have received the investigational product, they will complete post-baseline

[redacted]

assessments. Subjects will evaluate the time to first perceptible relief and time to meaningful relief using a double stopwatch method up to 8 hours post-dose or until the time of first rescue medication use, whichever is sooner. At 0.25, 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, and 8 hours post-dose, subjects will provide ratings of pain severity using the 11-point NPSR and a rating of PR at each time point using a 5-point categorical Pain Relief Rating (PRR) scale. At 8 hours, subjects will also complete an overall assessment using the 6-point categorical Global Evaluation of the study medication. Adverse events will be assessed during the course of the study.

For the remaining of this document these products will be referred to as: IBU/CAF (IBU 400 mg plus caffeine 100 mg), IBU (IBU 400 mg), and placebo.

[redacted]

Table 2 Schedule of Activities

	Screening [a]	Surgery	Time relative to dose (hrs.)											Day 14 (+3 days)
			0	0.25	0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0	7.0	8.0 / Early Termination	
Informed Consent	x														
Review Inclusion & Exclusion Criteria	x	x													
Demographic Information	x														
Medical History	x														
Prior Medications	x														
History of drug and alcohol use	x														
Laboratory Test [b]	x													x	
Pregnancy Test [c]	x	x													
Serum FSH (post-menopausal females only)	x														
Contraception Check [d]	x													x	x
Surgical Procedure		x													
Surgical Trauma Scale		x													
Vital Signs (e.g., HR, BP, T, RR)	x		x											x	
Randomization			x												
Dosing			x												
Pain Evaluations:															
0-100 mm VAS Pain Severity Rating (PSR) [e]			x												
11-point Numerical Pain Severity Rating Scale (NPSR) [k]			x	x	x	x	x	x	x	x	x	x	x	x	
5-point Categorical Pain Relief Rating [f]				x	x	x	x	x	x	x	x	x	x	x	
Time to 'First Perceptible' Relief [g]				→	→	→	→	→	→	→	→	→	→	→	
Time to 'Meaningful' Relief [h]				→	→	→	→	→	→	→	→	→	→	→	
6-Point Categorical Patient Global Evaluation [i]														x	
Subjects taking a rescue medication during this time will be considered:			Discontinued	Treatment Failure											
Concomitant Medications		x	x	x	x	x	x	x	x	x	x	x	x	x	

[redacted]

	Screening [a]	Surgery	Time relative to dose (hrs.)												Day 14 (+3 days)
			0	0.25	0.5	1.0	1.5	2.0	3.0	4.0	5.0	6.0	7.0	8.0 / Early Termination	
AE	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→
SAE[J]	→	→	→	→	→	→	→	→	→	→	→	→	→	→	→

a. Screening must be within 30 days of surgery.
b. Screening laboratory tests include hematology, blood chemistry, urinalysis, coagulation, serum pregnancy test, and serum FSH as described in the protocol. In addition, females of child bearing potential will be given a serum pregnancy test at screening. If Early Termination, repeat safety labs.
c. Pregnancy test will be serum based at screening and urine based thereafter, provided the test has a sensitivity of at least 25 mIU/mL Females of CBP (child-bearing potential) must be given a urine pregnancy test on the morning of (but before) the surgery.
d. Contraception will be checked at Screening, Early Termination, and Follow-up. For subjects, this is the opportunity to assess any shift to or from childbearing to non-childbearing potential and address appropriately. In addition, the investigator or his/her designee will instruct the subject to call immediately if a selected contraception method is discontinued or if pregnancy is known or suspected.
e. 100 mm VAS Pain Severity Rating Scale: none=0 to severe=100.
f. 5-point Categorical Pain Relief Rating Scale: none=0, a little=1, some=2, a lot=3, and complete=4. Completed at each time point starting at 15 minutes post-dose and immediately before rescue medication use (if it occurs before hour 8).
g. Subject is instructed to stop the first stopwatch "when you first begin to feel any pain relieving effect whatsoever of the drug"
h. Subject is instructed to stop the second stopwatch "when you have meaningful relief, that is, when the relief from the pain is meaningful to you".
i. 6-point categorical Global Evaluation score: very poor=0, poor=1, fair=2, good=3, very good=4, excellent=5. Completed at the 8-hour time point or immediately before taking rescue medication.
j. SAE/AE active reporting period starts from when subject signs the ICD through 14 calendar days after the last administration of the investigational product. A qualified site staff member will conduct a post-study follow up via telephone 14 (or up to 17) days after the last dose of investigational product and inquire about any new SAE/AE. Assess immediately before administering rescue medication.
k. The 11-point Numerical Pain Severity Rating (NPSR) will evaluate pain intensity with 0= no pain and 10=worst possible pain. Completed at each time point and immediately before rescue medication use (if it occurs before hour 8). Subject must have a baseline pain score ≥5 on the 11-point NPSR confirmed by ≥50 on theVAS to be randomized to receive study medication.

Abbreviations & symbols: → = ongoing/continuous event; CBP= child bearing potential; FSH= follicle stimulating hormone; HR = heart rate; BP = blood pressure; NPSR= Numerical Pain Severity Rating; T = temperature; RR = respiration rate; PRR= Pain Relief Rating Scale; VAS = Visual Analogue Scale; PSR= Pain Severity Rating; AE= adverse event; SAE= serious adverse event;

[redacted]

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint(s)

- Time-weighted sum of pain relief rating (PRR) and pain intensity difference (PID) from 0 to 8 hours (SPRID 0-8) for IBU/CAF versus IBU.

The primary endpoint is formulated as:

$$SPRID_{0-8} = \sum_{i=1}^{11}(PRR_i + PID_i)(t_i - t_{i-1}) \text{, where } i=1,2,\ldots,11.$$

t_i corresponds to each post-dose time point, (i.e., $t_1 = 0.25$, $t_2 = 0.5$, ..., *and* $t_{11} = 8$ *hours*), and the time of dosing is considered as $t_0 = 0$.

Additionally,

- PRR_i is a 5-point categorical pain relief rating scale measured at each post-dose time point t_i;
- PID_i is calculated as the baseline NPSR score minus the post-baseline NPSR at each time point t_i, where the baseline NPSR is measured at time 0. So a higher positive value is indicative of a greater improvement.

See Appendix 2.2 for more details.

3.2. Secondary Endpoint(s)

- Time-weighted sum of pain relief rating and pain intensity difference scores over 2 hours (SPRID 0-2), over 4 hours (SPRID 0-4), over 6 hours (SPRID 0-6) and over 8 hours (SPRID 0-8, for IBU/CAF versus placebo comparison) post dose;

These secondary endpoints are derived similarly as the primary endpoint. For example, SPRID 0-2, the time-weighted sum over 2 hours (from $t_0=0$ to $t_5 = 2$ *hours*), is formulated as:

$$SPRID_{0-2} = \sum_{i=1}^{5}(PRR_i + PID_i)(t_i - t_{i-1}) \text{, where } i=1,2,\ldots,5.$$

- Time-weighted sum of pain intensity difference scores over 2 hours (SPID 0-2), over 4 hours (SPID 0-4), over 6 hours (SPID 0-6), and over 8 hours (SPID 0-8) post dose; positive and higher scores indicate greater reduction in pain;

These secondary endpoints are derived similarly as the primary endpoint. For example, SPID 0-2, the time-weighted sum over 2 hours (from $t_0=0$ to $t_5 = 2$ *hours*), is formulated as:

$$SPID_{0-2} = \sum_{i=1}^{5} PID_i * (t_i - t_{i-1}) \text{, where } i=1,2,\ldots,5.$$

[redacted]

- Time-weighted sum of pain relief rating over 2 hours (TOTPAR 0-2), over 4 hours (TOTPAR 0-4), over 6 hours (TOTPAR 0-6), and over 8 hours (TOTPAR 0-8) post dose; a higher score indicates greater pain relief;

These secondary endpoints are derived similarly as the primary endpoint. For example, TOTPAR 0-2, the time-weighted sum over 2 hours (from $t_0=0$ to $t_5 = 2$ *hours*), is formulated as:

$$TOTPAR_{0-2} = \sum_{i=1}^{5} PRR_i * (t_i - t_{i-1}), \text{ where } i=1,2,\ldots,5.$$

- *The sum of pain relief rating and pain intensity difference scores (PRID) at 0.25, 0.5, 1, 1.5, 2, 3, 4 5, 6, 7, and 8 hours post dose;*

See Appendix 2.2 for more details.

- *Pain Relief Rating (PRR): scored on the 5-point Categorical Pain Relief Rating Scale (0=No relief to 4=Complete relief) at 0.25, 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, and 8 hours post dose;*

See Appendix 2.2 for more details.

- *Pain Intensity Difference (PID): calculated as the baseline 11-point Numerical Pain Severity Rating (NPSR) minus the post-dose NPSR at 0.25, 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, and 8 hours post dose;* a higher value is indicative of greater improvement;

See Appendix 2.2 for details.

- *Time to onset of "meaningful" relief;*

Time to "meaningful" relief is defined as the elapsed time from dosing until the subject depresses the second stopwatch labeled meaningful relief; otherwise, the subject will be considered censored (see Section 5.3.2 for missing data handling rules).

If subjects become treatment failure prior to depressing the second stopwatch, the time to treatment failure will be considered the censoring time. Treatment failure is defined as taking rescue medication, or discontinuing the study due to lack of efficacy. Thus, the censoring time (or the time to treatment failure) will be determined as the time from dosing to:

- the start date and time for the first use of rescue medication; or
- the time of early termination if the subject withdraws due to lack of efficacy, and this time can be identified from the time of the assessments conducted at the visit of withdrawal/Early Termination.

If subjects withdraw from the study due to non-efficacy related reasons (e.g., AEs) without depressing the second stopwatch, the censoring time will be the time of withdrawal.

[redacted]

Subjects who do not depress the second stopwatch by the end of the evaluation period will be considered censored at their final assessment time, or the scheduled duration of study (8 hours), whichever is earlier.

- *Time to onset of "first perceptible" relief, confirmed by "meaningful" relief;*

Time to "first perceptible" relief (confirmed by "meaningful" relief) is defined as the elapsed time from dosing until the subject depresses the first stopwatch labeled first perceptible relief, if the subject also indicated achieving "meaningful" relief by the end of the scheduled evaluation (8 hours).

If the confirmation is not achieved, the subject will be censored, for analysis purposes, at the time when he/she depresses the first stopwatch.

If the subject becomes a treatment failure before depressing the first stopwatch, the subject will be considered censored (see Section 5.3.2 for missing data handling rules), the time to treatment failure will be considered the censoring time. Time to treatment failure is defined above in time to "meaningful" relief.

If the subject withdraws from the study due to non-efficacy related reasons before depressing the first stopwatch, the subject will also be considered censored, the censoring time will be at the time of the withdrawal.

If subjects do not depress the first stopwatch by the end of the evaluation period, he/she will be considered censored at their final assessment time, or the scheduled duration of study (8 hours), whichever is earlier.

- *Duration of relief, as measured by the time to treatment failure (ie, time to first dose of rescue medication or discontinuation due to lack of efficacy).*

Duration of relief will be estimated by the time to treatment failure. As mentioned earlier, treatment failure is defined as taking rescue medication, or discontinuing the study due to lack of efficacy. If a subject becomes a treatment failure before the end of the scheduled duration of the study, the time to treatment failure is the actual time from dosing to treatment failure, which has been defined above in time to "meaningful" relief. Subjects who never become a treatment failure will be considered censored at their final assessment time, or the scheduled duration of study, whichever is earlier.

CCI [redacted]

[redacted]

CCI

[redacted]

3.4. Baseline Variables

Baseline variables are those collected prior to dosing and can occur at screening, surgery or post-surgery at hour 0 during the study period (see Table 2). For this study, screening period is defined as the time period from the screening visit until start of surgery.

Pain Evaluations: the following will be collected prior to dosing, post-surgery at hour 0 during the study period. They will be summarized and listed. No treatment comparabilities will be assessed.

- 11-point Numerical Pain Severity Rating (NPSR):

[redacted]

A scale rating from 0 to 10 [0 (None), 10 (Worst Possible Pain)]. A minimum score of 5 (≥ 5) must be indicated before the subject can enter the trial.

- 100 mm Visual Analog Scale Pain Severity Rating (VAS-PSR):
  a scale rating from 0 to 100 [0 (no pain), 100 (worst pain)]. A minimum score of 50 mm is required to confirm that the subject has at least moderate pain at baseline in order to be randomized.

Demographic Data: Sex, race, ethnicity, and age will be collected at screening. These variables will be summarized and listed. No treatment comparabilities will be assessed.

Background Information: Past and current medical history (non-drug allergies) will be collected at screening and will be summarized by treatment and total in a similar manner as adverse events (AEs) using the MedDRA system organ class (SOC) and preferred term. The corresponding listing will be also produced. Drug allergies will be also collected at this time but will only be listed. No treatment comparabilities will be assessed.

Tooth Extraction and Surgery Details: The duration of surgery, number of teeth removed, and surgical trauma will be collected during the surgery. These variables, in addition to the time to study medication since the end of surgery will be summarized as described in Section 6.5.1.4. All tooth extraction and surgery details including the date of surgery and information on the teeth removed will be listed.

3.4.1. Stratification Variables

Subjects will be stratified by gender and baseline pain severity ("moderate" or "severe"), where a score of 5 to 7 on the baseline 11-point NPSR will be classified as "moderate" and a score of 8 to 10 will be classified as "severe".

3.4.2. Covariates

The baseline value of the 11-point NPSR, will be used as a covariate in the analysis of the summary scores such as SPRID, SPID and TOTPAR from 0-2, 0-4, 0-6, and 0-8 hours, as well as PRR, PID and PRID scores at each post-dosing time points. The statistical specifications for these variables or their analysis are described in Section 5.2.1.

3.5. Safety Endpoints

Adverse event (AE) analyses will include all events which initially occurred, or worsened following treatment.

3.5.1. Adverse Events

An adverse event is considered treatment emergent relative to a given treatment (called treatment emergent adverse events [TEAE]) if:

- the event occurs for the first time during the treatment and was not seen prior to the start of treatment (i.e., receiving investigational product), or

[redacted]

- the event was seen prior to the start of treatment but increased in severity during treatment.

The time period for actively eliciting and collecting AEs and SAEs ("active collection Period") for each subject begins from the time the subject /parent(s)/legal guardian/legally acceptable representative provides informed consent, which is obtained before the subject's participation in the study (ie, before undergoing any study-related procedure and/or receiving investigational product), through and including a minimum of 14(+3) calendar days after the last administration of the investigational product.

No tier-1 events (pre-specified events of clinical importance) have been identified in the Safety Review Plan (SRP) for the product IBU/CAF. However, for this study AEs included in the Targeted Medical Event (TME) list of the SRP (See Appendix 4) will be summarized in a similar manner as Tier-1 events if any are identified . Therefore, a 3-tier approach will be used to summarize these type of AEs. Under this approach, AEs are classified into 1 of 3 tiers. Different analyses will be performed for different tiers (See Section 6.6.1).

Tier-1 events: These are pre-specified events of clinical importance and are maintained in a list in the product's TME of the SRP. For this study TME in the SRP are considered as Tier-1 events.

Tier-2 events: These are events that are not tier-1 but are "common". A MedDRA preferred (PT) is defined as a tier-2 event if its frequency is at least 2% in any treatment group.

Tier-3 events: These are events that are neither tier-1 nor tier-2 events.

3.5.2. Laboratory Data

Laboratory assessments such as hematology, coagulation, blood chemistry, urinalysis, and pregnancy will be performed at screening and results will be used for inclusion purposes only. The results will not be entered in the database.

3.5.3. Vital Signs

Each subject's vital signs (including heart rate, blood pressure, respiratory rate and temperature) will be measured and recorded at Screening, Baseline (prior to dosing), and at end of study, or at the time of withdrawal or the first use of rescue medication (if aplicable).

4. ANALYSIS SETS

Data for all subjects will be assessed to determine if subjects meet the criteria for inclusion in each analysis population prior to releasing the database and unblinding. The classifications will be documented per standard operating procedures.

4.1. Full Analysis Set

The Full Analysis Set (FAS), primary analysis population, is the intent-to-treat (ITT) subject population, defined as all randomized subjects who dosed with the study medication and provided a baseline assessment. Subjects will be assigned to the randomized treatment group regardless of the treatment received.

4.2. Per Protocol Analysis Set

The 'per-protocol' analysis set will include the subset of subjects included in the Full Analysis Set who completed the study per protocol (evaluable subjects). Thus, this analysis set will exclude any subjects who were discontinued from the study, or had a significant protocol violation. Subjects who will be excluded from the 'per-protocol' analysis set will be documented prior to breaking of the study blind. If the per protocol population consists of 90% or more of the FAS population, a per protocol analysis will not be performed.

4.3. Safety Analysis Set

The safety analysis set will include all subjects who received the study medication. Subjects will be analyzed according to the treatment they actually receive regardless of which treatment group they are randomized. A randomized but not treated subject will be excluded from the safety analyses.

4.4. Other Analysis Sets

Not applicable.

5. GENERAL METHODOLOGY AND CONVENTIONS

The final analysis of data will be performed after unblinding of the treatment codes.

5.1. Hypotheses and Decision Rules

For the analysis of primary endpoint, the statistical hypothesis to be tested is that the IBU with caffeine formulation will provide significantly greater analgesic effect than the IBU treatment. The statistical testing will be performed at the 0.05 (two sided) level of significance.

No adjustments for multiplicity will be made for the secondary endpoints.

[redacted]

5.2. General Methods

All computations will be performed using SAS® version 9.2 or higher (SAS Institute, Cary, NC). Statistically significant treatment differences will be declared if the probability of random occurrence between treatments, p, is ≤ 0.05 (two sided). Treatment differences will be considered marginally significant if $0.05 < p \leq 0.10$. All tests will be two sided.

5.2.1. Analyses for Continuous Data

Descriptive Statistics

For continuous endpoints (SPRID, SPID, TOTPAR, PRR, PID, PRID), descriptive statistics including sample size, mean, standard deviation, median, minimum, and maximum will be provided by treatment group.

Analysis of Covariance (ANCOVA)

All analyses of continuous efficacy endpoints (SPRID, SPID, TOTPAR from 0-2, 0-4, 0-6, 0-8 hours, as well as PRR, PID, PRID at each post dosing time point) will use the analysis of covariance (ANCOVA) model. The model will contain treatment, gender, and baseline 11-point numerical pain severity rating as factors. For each pairwise comparison of interest, the treatment difference based on the least-squares means (LSM), its standard error (SE), the p-value and the associated 95% confidence interval (CI) will be presented.

CCI [redacted]

[redacted]

[redacted]

CCI [redacted]

[redacted]

[redacted]

CCI [redacted]

5.2.4. Analyses for Time to Event Data

Time-to-event endpoints will be displayed graphically using the survival curves based on the Kaplan-Meier estimates. The median survival time and its corresponding 95% CI will be estimated using the method of Simon and Lee (1982) method.

The Gehan-Wilcoxon test, stratified by gender and baseline categorical pain severity terms will be used to analyze the time-to-event endpoints, including time to "meaningful" relief, time to "first perceptible" relief (confirmed by "meaningful" relief), and duration of relief (i.e., time to treatment failure) (overall treatment effect and pairwaise comparisons).

5.3. Time Windows and Methods to Manage Missing Data

The following adjustments to the data will be made, as needed, prior to computing the derived endpoints.

5.3.1. Time Windows

Efficacy assessments (NPSR and PRR) are scheduled to be completed at 0.25, 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, and 8 hours post-dose and the analyses will be done corresponding to these time points. Because some subjects may complete their assessments at times differing from the scheduled time points, each such score will be adjusted so as to reflect more accurately the score that would have been observed had it been recorded at the scheduled time point. Time windows will be created for each of the analyzed time points, the widths of ±5 minutes for time points which are less than 1 hour apart, and ±10 minutes for time points are one hour apart from previous time point. All pain relief and pain severity scores for each subject will be assigned to the scheduled time points according to these windows if the actual assessment time fall within these windows (See Table 4 Time Window for Scheduled Time Points).

All assessments completed more than a minute after a subject takes the first dose of rescue medication will be ignored. Remaining efficacy assessments subsequent to the time of rescue medication will be extrapolated (see below for **Extrapolation**).

Specifically, if there is only one, or multiple, or none efficacy assessments in a time window, the following assignment rules will be applied:

- If only one assessment is performed within a time window, the corresponding value will be assigned to that time point (e.g., an assessment performed anywhere between 115 and 125 minutes, regardless of when it was scheduled, will be assigned to the 2-hour time point).

[redacted]

- If more than one value falls within a window (this may include assessments completed at the time of rescue medication), a simple average of the values will be used.

- If at a particular time point, there is no value within its window, then a value will be assigned by either interpolation or extrapolation.

    1) **Interpolation:** If a window is empty and assessments are available before and after the time point window, then a value will be interpolated by calculating the weighted average of the values immediately proceeding, and subsequent to the time point window, with weights inversely proportional to the duration between the analyzed time point and the actual time.

    2) **Extrapolation:** If a window is empty and no subsequent scores are available, a value will be assigned to the empty window via extrapolation. The extrapolated value will depend on whether the subject took rescue medication, dropped out due to lack of efficacy, or discontinued due to an adverse event during or before the particular time point's window. If so, the worse of the preceding score (this may be the assessment completed at the time of rescue medication) and the baseline score will be assigned as the pain severity scores, and a score of 0 ('no relief') as the pain relief score. If the data being missing is not due to any of these reasons, then the last observation will be carried forward for the pain severity scores and the pain relief scores.

5.3.2. Methods to Manage Missing Time to Event Data

- For subjects who achieved "first perceptible" and "meaningful" relief, but the time to "meaningful" relief is missing (i.e. time was inadvertently not noted in the source documentation), their "meaningful" relief will be censored at the time of "first perceptible" relief. In addition, since it was recorded that the subject achieved "meaningful" relief (although the time is missing), the "first perceptible" relief will be considered confirmed.

- If a subject indicated that they achieved "first perceptible" relief, but the time is missing, time to "first perceptible" relief will be retained as missing. Similarly, if both times (time to "first perceptible" relief and time to "meaningful" relief) are missing regardless of whether they indicated that they achieved the relevant level of relief, they will both be retained as missing.

6. ANALYSES AND SUMMARIES

A summary of the primary and all secondary endpoints are presented in Appendix 1 with relevant details.

6.1. Primary Endpoint: SPRID 0-8

6.1.1. Primary Analysis

- Analysis population: Full Analysis Set;
- Analysis methodology: SPRID 0-8 scores will be analyzed using the ANCOVA model (specified in Section 5.2.1).
- Supporting objective: Primary Objective

Reporting results:

- Raw data: The sample size, mean, standard deviation, median, minimum and maximum will be presented for each treatment arm.
- The LSMs, standard error, difference between the LSMs for each pair of treatment groups, p-value, and the corresponding 95% CI will be presented. The root mean square error (RMSE) and overall p-value based on the ANCOVA model will be also presented.

Figures

- Vertical bar chart showing the LSMs for SPRID 0-8 scores plus the corresponding standard errors on the y-axis showing any statistical significant differences based on the ANCOVA model.
- A forest plot showing the treatment differences, 95% CIs and p-values based on the ANCOVA model.

6.1.1.1. Sensitivity/Robustness Analyses

To support the interpretation of the primary analysis in the event that more than 10% of the FAS subjects are excluded from the full analysis set, the following analyses will be performed:

- Analysis population: Per Protocol Analysis Set;
- Analysis methodology: SPRID 0-8 scores will be analyzed using the ANCOVA model (specified in Section 5.2.1).

Reporting results:

- The sample size, mean, standard deviation, median, minimum and maximum will be presented for each treatment arm. The LSMs, standard error, difference between the LSMs for each pair of treatment groups, p-value, and the corresponding 95% CI will be presented. The RMSE and overall p-value based on the ANCOVA model will be also presented.

[redacted]

6.2. Secondary Endpoint(s)

6.2.1. SPRID 0-t, SPID 0-t, TOTPAR 0-t, PRID, PRR, and PID

- Analysis population: Full Analysis Set;
- Analysis time point:
  - SPRID: 0 to 2, 0 to 4, 0 to 6, and 0 to 8 hours post-dose;
  - SPID: 0 to 2, 0 to 4, 0 to 6, and 0 to 8 hours post-dose;
  - TOTPAR: 0 to 2, 0 to 4, 0 to 6, and 0 to 8 hours post-dose;
  - PRID: 0.25, 0.5, 1, 1.5, 2, 3, 4 5, 6, 7, and 8 hours post dose;
  - PRR: 0.25, 0.5, 1, 1.5, 2, 3, 4 5, 6, 7, and 8 hours post dose;
  - PID: 0.25, 0.5, 1, 1.5, 2, 3, 4 5, 6, 7, and 8 hours postdose;
- Analysis methodology: the summary scores of SPID, TOTPAR and SPRID from 0-2, 0-4, 0-6 and 0-8 hours, as well as PRR, PID and PRID scores at each post-dosing time point will be analyzed using the ANCOVA model (specified in Section 5.2.1).

Reporting results:

- 11-point NPSR score: The sample size, mean, standard deviation, median, minimum and maximum at baseline and at each post-baseline time point will be presented for each treatment arm.
- For all endpoints in this section: The sample size, mean, standard deviation, median, minimum and maximum, will be presented for each treatment arm as well as the LSM and standard error based on the main effects ANCOVA model. The RMSE and overall p-value based on the main effects model will be also listed. For each pair-wise comparison of interest, the treatment differences based on the LSM, p-value and corresponding 95% CI based on the main effects model will be presented.

Figures

- Vertical bar charts showing the LSMs for SPID, SPRID, and TOTPAR scores at the different time points plus the corresponding standard errors on the y-axis will be presented as well as any significant differences based on the main effects model.
- Line charts of the mean PRR, PID, and PRID scores (y-axis) over time (x-axis) for all treatment groups plus statistical significant differences based on the main effects model will be provided.

[redacted]

6.2.2. Time to Onset of "Meaningful" Relief, Time to Onset of "First Perceptible" Relief (Confirmed by "Meaningful" Relief), and Duration of Relief

- Analysis population: Full Analysis Set;
- Analysis methodology: Time to event data will be analyzed using the Gehan-Wilcoxon test as specified in Section 5.2.4.

Reporting results:

- The median (in minutes) for the time to event endpoints, the 95% CI for the medians (see Section 5.2.4), number and percentage of subjects with event for each treatment arm, and the p-values from the Gehan-Wilcoxon test (overall, and pairwise) will be presented.

Figures:

- For the duration of relief, the estimated survival curve based on Kaplan-Meier method will be displayed with the probability of no treatment failure on the y-axis and time (in minutes) on the x-axis. Statistically significant differences, if any, will be also added.
- For the time to onset of "meaningful" relief, the estimated survival curve based on Kaplan-Meier method will be displayed with the probability of onset on the y-axis and time (in minutes) on the x-axis. Statistical significant differences, if any, will be also added.
- For the time to onset of "first perceptible" relief (confirmed by "meaningful" relief), the estimated survival curve based on Kaplan-Meier method will be displayed with the probability of onset on the y-axis and time (in minutes) on the x-axis. Statistical significant differences, if any, will be also added.

CCI

[redacted]

CCI

[redacted]

6.4. Subset Analyses

Subgroup analyses will be conducted for the primary endpoint among subjects in the full analysis set, based on dichotomized baseline pain severity and gender, respectively. Similar analysis models as described above will be used, except that the baseline pain severity (numerical or categorical) or gender will be excluded from each model, as appropriate.

6.5. Baseline and Other Summaries and Analyses

All summaries will be presented by treatment groups and total.

[redacted]

6.5.1. Baseline Summaries

6.5.1.1. Pain Evaluations

- Analysis population: Full Analysis set and/or the per protocol population when if this analysis is required;
- Analysis methodology: Descripive statistics.

Reporting results:

- The mean, standard deviation, median, minimum, and maximum values will be provided for the VAS pain intensity and 11-point numerical pain severity scores. The number and percentage of subjects within each baseline pain severity category ('moderate' and 'severe') will be presented. The sample size will be also listed.

6.5.1.2. Demographic Data

- Analysis population: Safety population;
- Analysis methodology: Descripive statistics.

Reporting results:

- The number and percentage of subjects within each category in gender, race and ethnicity will be presented while for age the overall mean, standard deviation, median, minimum, and maximum values will be provided. The sample size overall and by treatment will be also listed.

6.5.1.3. Background Information

- Endpoint: Medical Conditions;
- Analysis population: Safety population;
- Analysis methodology: None.

Reporting results:

- For past and current medical history (non-drug allergies): The sample size, number and percentage of subjects within each SOC and preferred term will be presented by treatment and overall in a similar manner as adverse events (AEs). The corresponding listing will be also produced.
- For Drug allergies: listing will be produced.

6.5.1.4. Tooth Extraction and Surgery Details

- Endpoint: The duration of surgery, the time to study medication since the end of surgery; number of teeth removed, and surgical trauma rating;
- Analysis population: Safety population, full anlaysis set, if these two populations are very different and per protocol population (if required);
- Analysis methodology: Descriptive statistics.

Reporting results:

- For the duration of surgery (in minutes) and the time to study medication since the end of surgery (in hours), the overall mean, standard deviation, median, minimum, and maximum values will be provided.
- For the categories within the number of teeth extracted and the surgical trauma rating, the number and percentage of subjects will be listed.
- All tooth extraction and surgery details including the date of surgery and information on the teeth removed will be listed appropriately.

6.5.1.5. Baseline and Background Endpoints for Subset Analyses

Demographic and baseline pain assessments will also be summarized by pain severity categories: moderate vs. severe. Descriptive statistics as listed for the overall population will also be presented for the 2 subgroups. No treatment comparabilities will be assessed.

6.5.2. Study Conduct and Subject Disposition

Subject disposition and populations groups will be summarized using frequencies and percentages. No treatment comparabilities will be assessed.

6.5.3. Concomitant Medications and Non-Drug Treatments

Prior and concomitant medications will be summarized using the preferred drug name from WHODrug for the Safety population. No treatment comparabilities will be assessed.

Concomitant non-drug treatment or procedures will be listed.

6.6. Safety Summaries and Analyses

6.6.1. Adverse Events

Adverse event analyses will include all events collected during the active collection period of the study. Adverse events will be summarized by the MedDRA SOC and preferred term, and classified according to their severity (mild, moderate, or severe) and relationship (related or not related) to study product. For the summary by severity, subjects who have multiple

[redacted]

occurrences of the same AE will be classified according to the worst reported severity of the AE. Similarly, for the summary by relationship to the study product, the AE will be classified according to the worst relationship.

All non-TEAE will be summarized according to the MedDRA preferred term.

No tier-1 events (pre-specified events of clinical importance) are identified in the SRP for the product IBU/CAF. However, as indicated earlier, for this study, AEs included in the TME list of the SRP (See Appendix 4) will be summarized in a similar manner as Tier-1 events if any are identified. Therefore, AEs will additionally be summarized using a 3-tier approach. AEs will be classified into 3 tiers as defined in Section 3.5.1. All tier-1, 2, or 3 AEs will be tabulated and listed separately in a similar manner as all other AEs. The AEs will be sorted alphabetically within a system organ class.

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. As such, safety analysis is generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation. The 3-tier approach facilitates this exploratory analysis.

Subgroup analyses will be conducted for the overall AE summary among the safety population, based on the dichotomized baseline pain severity.

6.6.2. Laboratory Data

Laboratory assessments such as hematology, coagulation, blood chemistry, urinalysis, and pregnancy will be performed at screening and early termination (see Table 2), and results will be used for inclusion or safety purposes. Laboratory data will not be entered in the database and thus will not be summarized.

6.6.3. Vital Signs

The values at baseline (pre-dose), end of study (the last post-dose) as well as the change from baseline (End of study – baseline) for the vital signs data (i.e., heart rate, blood pressure, respiratory rate and temperature) will be summarized by descripive statistics, overall and by treatment group for the safety population. The number and percentage of subjects with vital sign data out-of-normal ranges will be summarized at both pre and post dose time points.

Corresponding listing will be created and subjects with out-of-normal ranges values will be flagged.

7. INTERIM ANALYSES

No interim analysis is planned.

[redacted]

8. REFERENCES

Koch G, Carr GJ, Amara IA, Stokes ME, Uryniak TJ. Categorical data analysis. In Berry DA, ed. Statistical Methodology in the Pharmaceutical Sciences. New York: Marcel Dekker, 1990: 389-474.

Simon R, Lee YJ, Nonparametric confidence limits for survival probabilities and median survival time. Cancer Treat Rep 1982; 66:37-42.

[redacted]

9. APPENDICES

Appendix 1. SUMMARY OF EFFICACY ANALYSES

Table 3 Summary of Efficacy Analyses

Primary	**Comparisons**	**Analyses**
SPRID 0-8	IBU/CAF versus IBU	ANCOVA model
Secondary		
SPRID0-8	Overall, and - IBU/CAF versus placebo - IBU versus placebo	ANCOVA model
SPRID0-2, SPRID0-4, SPRID0-6, SPID0-2, SPID0-4, SPID0-6, SPID0-8, TOTPAR0-2, TOTPAR0-4, TOTPAR0-6, TOTPAR0-8	Overall, and - IBU/CAF versus IBU - IBU/CAF versus placebo - IBU versus placebo	ANCOVA model
PRR, PID and PRID scores at each post-dosing time point	Overall, and - IBU/CAF versus IBU - IBU/CAF versus placebo - IBU versus placebo	ANCOVA model
-Time to onset of meaningful relief, -Time to onset of first perceptible relief, -Duration of relief	Overall, and - IBU/CAF versus IBU - IBU/CAF versus placebo - IBU versus placebo	Gehan-Wilcoxon test; Figures: Kaplan-Meier method
CCI [redacted]		

[redacted]

CCI

Note: ANCOVA model will contain treatment, gender, and baseline 11-point numerical pain severity rating as factors.
Overall: treatment effect in the ANCOVA model.

Appendix 2. DATA DERIVATION DETAILS

Appendix 2.1. Definition and Use of Visit Windows in Reporting

Table 4 Time Window for Scheduled Time Points

Scheduled time point	Time window (min)
0.25 hr/15 min	[10, 20]
0.5 hr/30 min	[25, 35]
1 hr/60 min	[55, 65]
1.5/90 min	[85, 95]
2 hr/120 min	[115, 125]
3 hr/180 min	[170, 190]
4 hr/240 min	[230, 250]
5 hr/300 min	[290, 310]
6 hr/360 min	[350, 370]

7 hr/420 min	[410, 430]
8 hr/480 min	[470, 490]

Appendix 2.2. Definition of Endpoints NPSR, PRR, PID and PRID

Time point (hr)	***Numerical Pain Severity Rating (NPSR)***	***Categorical Pain Relief Rating (PRR)***	***Pain Intensity Difference (PID)***	***Sum of pain relief rating and pain intensity difference scores (PRID)***
$t_0 = 0$	$NPSR_0$			
$t_1 = 0.25$	$NPSR_1$	PRR_1	$PID_1 = NPSR_0 - NPSR_1$	$PRID_1 = PRR_1 + PID_1$
$t_2 = 0.5$	$NPSR_2$	PRR_2	$PID_2 = NPSR_0 - NPSR_2$	$PRID_2 = PRR_2 + PID_2$
$t_3 = 1$	$NPSR_3$	PRR_3	$PID_3 = NPSR_0 - NPSR_3$	$PRID_3 = PRR_3 + PID_3$
$t_4 = 1.5$	$NPSR_4$	PRR_4	$PID_4 = NPSR_0 - NPSR_4$	$PRID_4 = PRR_4 + PID_4$
$t_5 = 2$	$NPSR_5$	PRR_5	$PID_5 = NPSR_0 - NPSR_5$	$PRID_5 = PRR_5 + PID_5$
$t_6 = 3$	$NPSR_6$	PRR_6	$PID_6 = NPSR_0 - NPSR_6$	$PRID_6 = PRR_6 + PID_6$
$t_7 = 4$	$NPSR_7$	PRR_7	$PID_7 = NPSR_0 - NPSR_7$	$PRID_7 = PRR_7 + PID_7$
$t_8 = 5$	$NPSR_8$	PRR_8	$PID_8 = NPSR_0 - NPSR_8$	$PRID_8 = PRR_8 + PID_8$
$t_9 = 6$	$NPSR_9$	PRR_9	$PID_9 = NPSR_0 - NPSR_9$	$PRID_9 = PRR_9 + PID_9$
$t_{10} = 7$	$NPSR_{10}$	PRR_{10}	$PID_{10} = NPSR_0 - NPSR_{10}$	$PRID_{10} = PRR_{10} + PID_{10}$
$t_{11} = 8$	$NPSR_{11}$	PRR_{11}	$PID_{11} = NPSR_0 - NPSR_{11}$	$PRID_{11} = PRR_{11} + PID_{11}$

[redacted]

Appendix 3. STATISTICAL METHODOLOGY DETAILS

The following SAS Code will be used to calculate the Gehan-Wilcoxon Test p-values for pairwise comparisons.

```
proc lifetest data = all /*timelim = 8hour or 480 min*/;
where rx in (1,2); /* Remove this statement for overall analysis*/
time vartime*cenmean(0); /*0 = Censored*/
strata sex base / group=rx test=(wilcoxon);
run;
```

Where

rx= Treatment (1 and 2 = Treatments being compared)
vartime= time-to-event variable
cenmean= censoring variable (cenmean=0 censored)
base = dichotomized baseline variable

Appendix 4. TARGET MEDICAL EVENT (TME) LIST

BRAND or TRADE Name	TARGET MEDICAL EVENT TERM	TME Safety Area of Interest and RATIONALE FOR INCLUDING TME	LEVEL (PT, LLT, HLT, HLGT & SOC?)
Ibuprofen/caffeine	Gastrointestinal ulceration and perforation	Event of PV interest	HLGT
Ibuprofen/caffeine	Gastrointestinal haemorrhages NEC	Event of PV interest	HLGT
Ibuprofen/caffeine	Renal failure and impairment	Event of PV interest	HLT
Ibuprofen/caffeine	Cardiovascular disorders	Event of PV interest	HLGT
Ibuprofen/caffeine	Cardiac arrhythmias	Event of PV interest	HLGT
Ibuprofen/caffeine	Heart failures	Event of PV interest	HLGT
Ibuprofen/caffeine	Central nervous system vascular disorders	Event of PV interest	HLGT
Ibuprofen/caffeine	Hearing disorders	Event of PV interest	HLGT
Ibuprofen/caffeine	Hepatic and hepatobiliary disorders	Event of PV interest	HLGT
Ibuprofen/caffeine	Bullous conditions	Event of PV interest	HLT

[redacted]

Ibuprofen/caffeine	Angioedemas	Event of PV interest	HLT
Ibuprofen/caffeine	Overdoses	Event of PV interest	HLT