A Phase 2 Double-Blind Clinical Trial to Examine the Comparative Effects on Osteoarthritic Knee Pain of CGS-200-1 (1% Capsaicin Topical Liquid), CGS-200-5 (5% Capsaicin Topical Liquid), and CGS-200-0 (Vehicle, No Capsaicin) – Statistical Analysis Plan

Protocol Number: VZU00025

Protocol Date: 24-January -2019

NCT03528369

A Phase 2 Double-Blind Clinical Trial to Examine the Comparative Effects on Osteoarthritic Knee Pain of CGS-200-1 (1% Capsaicin Topical Liquid), CGS-200-5 (5% Capsaicin Topical Liquid), and CGS-200-0 (Vehicle, No Capsaicin)

Protocol Number:	VZU00025
Protocol Version:	5.0
Protocol Date:	01AUG2018

STATISTICAL ANALYSIS PLAN

Version 2.0

A Phase 2 Double-Blind Clinical Trial to Examine the Comparative Effects on Osteoarthritic Knee Pain of CGS-200-1 (1% Capsaicin Topical Liquid), CGS-200-5 (5% Capsaicin Topical Liquid), and CGS-200-0 (Vehicle, No Capsaicin)

STATISTICAL ANALYSIS PLAN
Version 2.0

Author: [redacted] 24JAN2019
Biostatistician, CTI — Date

CTI Biostatistics Reviewer: [redacted] 24 JAN 2019
Director, Biostatistics, CTI — Date

Sponsor Approval: [redacted] 22-JAN-2019
CSO and Head of Regulatory
Vizuri Health Sciences Inc. — Date

[redacted] 24 January 2019
Clinical Operations Consultant
Vizuri Health Sciences Inc. — Date

Table of Contents

LIST OF ABBREVIATIONS

Abbreviation	Definition
ACR	American College of Rheumatology
AE	Adverse Event
ANCOVA	Analysis of Covariance
ANSI	American National Standards Institute
AQL	Acceptable Quality Level
ATC	Anatomical, Therapeutic, and Chemical
AUC	Area under the concentration/time curve
BMI	Body mass index
BP	Blood pressure
BSP	Burning-stinging pain
CDA	Clinical Data Associate
CRO	Contract Research Organization
DLT	Dose-limiting toxicity
ECG	Electrocardiogram
eCRF	Electronic case report form
ESR	Erythrocyte-sedimentation rate
FSH	Follicle-stimulating hormone
ITT	Intent-to-treat
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent-to-Treat
NRS	Numeric Rating Scale
OA	Osteoarthritis
OAKP	Osteoarthritis knee pain
PP	Per Protocol
PT	Preferred term
QC	Quality control
RCTC	Rheumatology Common Toxicity Criteria
RDC	Remote data capture
RF	Rheumatoid factor
SAE	Serious adverse event

Abbreviation	Definition
SAF	Safety
SAP	Statistical Analysis Plan
SOC	System organ class
TEAE	Treatment-emergent adverse event
TESAE	Treatment-emergent serious adverse event
VAS	Visual Analogue Scale
WHO	World Health Organization
WOMAC	Western Ontario and McMaster Universities Index of Osteoarthritis

SAP REVISIONS

The following table details the changes made to the SAP version 1.0 (04OCT2018) to provide clarifications on the efficacy analysis.

Protocol Version #	SAP Section	Modification	Description and Rationale
5.0	List of Abbreviatons	Added "PP" to the list	Updated for the analysis
5.0	4.2	Added the definition of Per Protocol analysis set and the corresponded analysis.	Updated to be consistent with the planned analysis for the study.
5.0	6.3	Updated the definition of durable clinical response.	Updated to provide clarifications.

1. INTRODUCTION

This statistical analysis plan (SAP) is based on the Protocol # VZU00025, titled "A Phase 2 Double-Blind Clinical Trial to Examine the Comparative Effects on Osteoarthritic Knee Pain of CGS-200-1 (1% Capsaicin Topical Liquid), CGS-200-5 (5% Capsaicin Topical Liquid), and CGS-200-0 (Vehicle, No Capsaicin)." See the study protocol for full details.

This document details the statistical methods planned to perform the final analyses of the study. This study will involve application of CGS-200-1 and CGS-200-5 (which has 1% and 5% capsaicin respectively) or of a Vehicle CGS-200-0 (which has no capsaicin in it). Study analysis will include comparison of the treatment effect on osteoarthritis (OA) knee pain as assessed by the Western Ontario and McMaster Universities Index of Osteoarthritis (WOMAC) score for OA signs and symptoms.

2. OBJECTIVES AND ENDPOINTS

2.1 Objectives

The objectives of the present study are the following:

- To develop pain relief dose-response data for osteoarthritis knee pain (OAKP), compared to Vehicle, for the following concentrations of Capsaicin in CGS-200 vehicle: 1% and 5% using the WOMAC pain subscale (VAS formatted) as the primary pain index for relief of OAKP.
- To confirm, in a U.S. based population, the high safety profile for CGS-200-5 that was reported in Studies VZU00021 and VZU00023 in a Dominican Republic population.
- To develop data for relief of signs and symptoms of osteoarthritis using the WOMAC subscales for pain, stiffness and function and total WOMAC score.
- To develop OAKP relief data using the 100 mm VAS scale.

2.2 Endpoints

2.2.1 Primary Efficacy Endpoint

The primary efficacy endpoint of this study will be to examine the extent of reduction in the WOMAC pain score, relative to baseline, provided by once daily, one-hour application of Vehicle (CGS-200-0), CGS-200-1 and CGS-200-5 at the following assessment times:

Baseline: ≤ 30 minutes prior to first daily application

Day 35: 31 Days (+/- 3 days) after fourth daily application

and based thereon to determine whether the responses for treatment with CGS-200-1, and CGS-200-5 are significantly different from Vehicle (CGS-200-0) and from each other at the primary endpoint time.

2.2.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints will be:

(1) To examine the extent of reduction in the WOMAC pain score, relative to baseline, provided by once daily, one-hour application of Vehicle (CGS-200-0), CGS-200-1, and CGS-200-5 at the following assessment times:

Baseline: ≤ 30 minutes prior to first daily application

Day 5 24 hrs (+/- 3) after the last of the 4 daily applications

Day 19 15 days (+/- 3) after the last of the 4 daily applications

Day 64 60 days (+/- 3) after the last of 4 daily applications

Day 94 90 days (+/- 3) after the last of 4 daily applications

(2) Day 5, 19, 35, 64, and 94 WOMAC scores on the stiffness and function subscales as well as total WOMAC score (including the WOMAC pain score).

2.2.3 Primary Safety Endpoints

The primary safety endpoints will be:

(1) The application of the study drug does not produce skin reactions (erythema, scaling, pruritus, or other) to a degree that is clinically of concern. Scoring of erythema and scaling at the application sites and of pruritus will be assessed per the scoring systems in the Protocol Appendix G.

(2) No SAE's either possibly, probably or definitely associated with study treatments.

(3) Safety labs (hematology, serum chemistry and urinalysis) do not produce other than minimal – mild toxicities (Grade 1 or Grade 2). Scoring of clinical, hematological and urinalysis out of range findings for toxicity grading will be per the Protocol Appendix B.

2.2.4 Secondary Safety Endpoints

The secondary safety endpoints will be:

(1) The distribution by study arms of specific Adverse Events (AEs) (i.e.: local application site reactions), and other AEs (inclusive of findings for hematology, serum chemistry and urinalysis).

(2) Evaluate the amount of concomitant pain medications used overtime.

2.2.5 Tolerability Endpoints

The tolerability endpoints (which are secondary in nature) will be:

(1) The application of study drug does not produce burning-stinging pain ("BSP") at the application site to a degree that is not acceptable to the subject. BSP at the application site, this will be assessed using a 0 – 10 Numeric Rating Scale (NRS) scale without guideposts. "Acceptability" of BSP will be queried per subject at the end of each application period.

(2) The application of the study drug does not produce pruritus during application or in the 24 hrs post-application to a degree that is not acceptable to the subject or, if bothersome to the subject, cannot be managed by application of ice or a cold pack or compress.

3. INVESTIGATIONAL PLAN

3.1 Study Design

This is a multi-center, randomized, double-blind clinical trial to examine the comparative effects on OAKP of CGS-200-1 (1% Capsaicin content) (N=40), CGS-200-5 (5% Capsaicin content) (N=40), and CGS-200 Vehicle (no Capsaicin) (N=40) in subjects with OA of the knees according to the 1986 American College of Rheumatology (ACR) criteria. Assigned doses will be applied at the clinic for 60 minutes on each of four consecutive days.

Subjects who meet the inclusion/exclusion criteria specified in the Protocol, Section 4.1 will be randomized to one of the three Arms in this study: CGS-200-1 or CGS-200-5 or CGS-200-0 (Vehicle) in a 1:1:1 ratio. All subjects will receive 4 consecutive days of treatment and will then be followed up until the Day 94 visit.

Even though both knee(s) will receive application of study test materials, with regard to reduction in WOMAC pain and VAS pain score associated with study treatments, only one knee will be indicated as the "Study Knee". This will be the knee with the highest WOMAC pain score at baseline. If both knees have equal WOMAC pain scores at baseline, then the right knee will be considered the "Study Knee" with regard to WOMAC pain and Visual Analogue Scale (VAS) pain score reduction.

Data will be collected from Day 1 through Day 5 and then again on Days 19, 35, 64 and 94 for efficacy, tolerability, and safety measures. See Appendix C for the Schedule of Events.

3.2 Treatment

3.2.1 Randomization Scheme and Treatment Arm Assignment

Subjects will be randomly assigned to receive CGS-200-1, CGS-200-5 or CGS-200-0 in a 1:1:1 ratio. Subjects that do not receive 4 complete consecutive daily doses and/or drop out before completing the Day 35 visit may be replaced if the reason for dropping out is other than lack of drug product tolerability. A dose will be considered complete when at least 80% has been delivered on any study drug administration day; dosing will be considered complete overall when a total of at least 80%, in the aggregate, has been delivered across all study treatment days. Subjects that drop out or who cannot be followed up from after the Day 5 visit until the Day 35 visit may be replaced. Up to 10 subjects per arm may be replaced.

3.2.2 Blinding

Treatment assignment for all subjects enrolled in this study will be double-blind; the subject, all site personnel, and Clinical Research Organization (CRO) personnel (with the exception of the Medical Monitor providing safety oversight) involved in this study will not know the treatment given to any subject. Every effort should be made to maintain the study blind.

3.2.3 Dosing Schedule

The study medication is for external use only and will be administered at the clinic by the subject under the supervision of trained site personnel. Wearing gloves, supplied by the clinic staff, a single dose of the assigned treatment will be topically applied to each knee on each of four dosing days. Both knees will be treated for all subjects, if possible, however only one knee will be the

study knee on which assessments are performed. The first application of medication will be at Visit 2 on Day 1 (initiated at Time 0). Dosing on each of the three subsequent treatment days should be performed at the same time of day ±2 hours in relation to the start time of the initial dose on Day 1.

3.2.4 Subject Compliance

Subjects will be free to withdraw from treatment or from the study at any time for any reason, or they may be withdrawn/removed, if necessary (see reasons for withdrawal below). It is understood by all concerned that an excessive rate of withdrawals can render the study uninterpretable; therefore, unnecessary withdrawal of subjects should be avoided.

Subjects will be removed from treatment for the following reasons:

- Unacceptable acute toxicities defined as:
  - Occurrence of a Dose-limiting toxicity (DLT)

Subjects will be removed from the study for the following reasons:

- Withdrawal of consent
- Non-compliance/Lost to follow-up
- Pregnancy

If treatment with study drug is interrupted and permanently stopped, subjects should remain on study and will be followed for scheduled safety and study assessments through Day 35. If the subject goes onto other treatment, this should be noted in the source documents and the concomitant medications electronic case report form (eCRF).

4. GENERAL CONSIDERATIONS FOR DATA ANALYSIS

Unless otherwise specified, continuous variables will be summarized using descriptive statistics (n, mean, standard deviation, median, minimum, and maximum). Categorical variables will be summarized showing the number and percentage of subjects within each category.

Summary results will be provided for each treatment group. All tabulations will be based on pooled data across centers.

All efficacy and tolerability statistical tests will be two-sided and tested at the 5% level of significance.

Analyses will be performed using SAS for Windows statistical software, version 9.4 or higher (SAS, Cary, NC), except where other software may be deemed more appropriate.

CTI will perform all efficacy and safety final analyses described in this SAP.

Subject data will be listed, sorted by treatment group and subject number.

4.1 Data Quality Assurance

The study sponsor, Vizuri Health Sciences, or its designee, will monitor the study. Study monitors representing the sponsor will visit study sites routinely throughout the trial. The sponsor/designee will review eCRFs and compare them with source documents to verify accurate and complete

collection of data and confirm that the study is being conducted according to the protocol. Auditors representing the sponsor may also similarly evaluate the study and its monitors.

A study coordinator at the study site will enter subject data into a remote data capture (RDC) database by completing eCRFs. All information recorded in the eCRFs for this study must be consistent with the investigator's source documentation for the study subjects. The study site will make available source documents to CTI personnel monitoring the study. The study monitor will verify consent of all subjects to participate in the study and will perform 100% source document verification of the eCRF data.

A CTI Clinical Data Associate (CDA) will review the data for discrepancies via programmed electronic consistency checks, data listings, or manually. Any discrepancies discovered via the data review process will be issued as queries in the RDC system to the study site for resolution. Once all the source verification is complete, all queries are resolved, and the database has been updated appropriately, the database will be locked and made available to CTI Biostatistics for final analyses.

All SAS programs used to create analysis datasets and output will be validated by ensuring that the ".log" files are void of all errors, warnings and notes indicative of problems. Additionally, each program will be checked to ensure that it performs according to the program specification. All programs are developed and validated by separate members of the CTI Biostatistics Department.

At the time of analysis, a quality control (QC) review of database values listed in SAS output will be compared to the database. The sample size of fields to undergo QC review will be determined by utilizing American National Standards Institute (ANSI) sampling procedures[1]. Sampling procedures are conducted using "normal" inspection criteria (Inspection Level II, Single, and Normal) and an Acceptable Quality Level (AQL) of 0.010%. The following shows the sampling criteria:

Single Normal sampling procedure for Acceptable Quality Level (AQL) 0.010%

Number of Fields	Sample Size	Accept/Reject Criteria
2-8	2	0/1
9-15	3	0/1
16-25	5	0/1
26-50	8	0/1
51-90	13	0/1
91-150	20	0/1
151-280	32	0/1
281-500	50	0/1
501-1,200	80	0/1
1,201-3,200	125	0/1
3,201-10,000	200	0/1
10,001-35,000	315	0/1
35,001-150,000	500	0/1
150,001-500,000	800	0/1
500,001-up	1,250	0/1

4.2 Analysis Sets

The following three analysis sets will be defined for this study:

The intent-to-treat (ITT) set will be defined as all randomized subjects, whether or not they receive study drug.

The modified ITT (mITT) set will be defined as the subset of the ITT subjects that receive at least one dose of study medication and have at least one evaluable post-dosing efficacy endpoint. The mITT set will be defined based on randomized treatment rather than treatment actually received.

All efficacy analyses will be carried out using the mITT set.

The safety (SAF) set will be based on the ITT subjects who have received at least one dose of study medication (based on the treatment actually received rather than randomized treatment). All safety analyses will be carried out using the SAF set.

The Per Protocol (PP) set will include all mITT subjects in whom there were no major protocol deviations. Major protocol deviations for defining the PP analysis set will be identified by Vizuri prior to the database lock. If deemed appropriate, the analyses on primary and secondary efficacy endpoints listed in Section 6.1 and 6.2 will be repeated using the PP set.

4.3 Assessment Windows

For the purpose of listing and summarizing data, the time-in-study for each subject observation will be defined using study days as defined in the Schedule of Events. No analysis windows are planned for the study regarding data collected outside the protocol specified windows. All data will be included in the analysis based on the visit as it is recorded in the database.

Baseline is defined as the last non-missing value prior to the first dosing.

If more than one assessment exists for a single visit, the value closest to the protocol study visit will be used for summary and analysis purposes. Data from all assessments will be listed.

4.4 Handling of Dropouts or Missing Data

Missing values will not be substituted by estimated values, but treated as missing in the statistical evaluation, except missing data imputation for WOMAC subscale scores, which is detailed in Section 4.6. All data from all subjects dosed in the study will be included in all listings, plots, and statistical analyses when appropriate.

4.5 Multiple Comparisons

Because of the exploratory nature of this study, there will be no adjustment for multiple comparisons. Therefore, individual statistically significant comparisons among the three treatments for any of the outcome measures should be interpreted with caution.

4.6 Data Derivations and Transformations

The WOMAC is composed of 3 subscales (A: Pain, B: Stiffness, C: Physical Function). Each subscale contains components that are scored using VAS ranging from 0 to 100 mm to capture the subject's response to each of the questions. There are 5 components for subscale A (Pain), 2 components for subscale B (Stiffness), and 17 components for subscale C (Physical Function).

A WOMAC subscale score is calculated by adding the component scores within the subscale. The total WOMAC score is computed as the sum of the 3 WOMAC subscale responses. In the case of missing data for a component of the WOMAC subscale, the average score of the non-missing components within the subscale will be used to impute for the missing values. However, the subscale score will be set to missing according to the following criteria: if ≥ 2 Pain components are missing, if both Stiffness components are missing, if ≥ 4 Physical Function components are missing. The total WOMAC score will be set to missing if any of the WOMAC subscale scores are set to missing.

5. STUDY SUBJECTS

5.1 Disposition of Subjects

A table of frequency counts and percentages of all screened subjects will be provided. Subject disposition including screen failures, subjects enrolled, study completion status, study medication completion status, and reasons for early termination will be tabulated by treatment group and overall. A by subject listing will also be provided.

5.2 Protocol Deviations

Distribution for the types of protocol deviations and the number of subjects that deviate from the protocol will be tabulated for the treatment groups in the ITT analysis set. A listing of all protocol deviations will be provided.

5.3 Demographic Characteristics

Descriptive statistics will be used to summarize the demographic characteristics (age, gender, race, ethnicity, height, weight, and Body Mass Index (BMI) for the ITT analysis set. A by subject listing will also be provided.

5.4 Baseline Characteristics

Baseline characteristics of ITT subjects include electrocardiogram (ECG), serum and urine pregnancy tests, follicle-stimulating hormone (FSH) test, erythrocyte sedimentation rate (ESR), IgM Rheumatoid Factor (RF), and exploratory assessment of OAKP. Clinical laboratory test and ECG results will be summarized using descriptive statistics for each treatment group. OA average and worst daily knee pain assessment for each knee will be provided in a listing. The knee pain scales will also be summarized by Study Knee and Non-study Knee using descriptive statistics for each treatment group.

5.5 Medical History

All medical conditions and surgical procedures will be classified by system organ class (SOC) and preferred term (PT) using the Medical Dictionary for Regulatory Activities (MedDRA). The number and percent of subjects with each medical condition and surgical procedure will be presented for each SOC and PT for the ITT analysis set.

5.6 Prior and Concomitant Medications

All concomitant medication collected from screening through the end of the study will be classified

by drug class and preferred terms according to the World Health Organization (WHO) Drug Dictionary. Prior medications are any medications taken prior to the first dose of study medication (i.e., start date and end date occurs before the first study medication dose date). If the medication start date is completely missing, then the medication will be considered as concomitant unless it can be determined that the medication end date occurred prior to the first dose of study medication. If the medication start date is partially missing and the partial date is not sufficient to determine if the medication was taken after the first dose of study medication, then the medication will be considered concomitant for the study unless it can be ruled out by the partial date and/or medication end date.

The number and percent of ITT subjects using concomitant medications will be tabulated by default Anatomical, Therapeutic, and Chemical (ATC) class and by preferred name. All prior and concomitant medications will be listed. Indication for the medications used to treat OAKP will also be presented in the listing.

6. EFFICACY AND TOLERABILITY ANALYSES

All efficacy analyses will be carried out on mITT subjects. Tolerability analysis will be based on the safety analysis set. Only one knee will be indicated as the "Study Knee" for efficacy analyses. The Study Knee will be the knee with the highest WOMAC pain score at baseline. If both knees have equal WOMAC pain scores at baseline, then the right knee will be considered the Study Knee with regard to WOMAC pain and VAS pain score reduction. Data from both knees will be presented in the listings.

6.1 Primary Efficacy Endpoint and Analysis

The primary efficacy endpoint is the change from baseline in WOMAC pain score at Day 35. WOMAC pain score reduction at Day 35 will be evaluated using the "Study Knee".

Change from baseline in WOMAC pain score on the Study Knee between each Active Treatment Arm and the Vehicle Arm at Day 35 will be evaluated using analysis of covariance (ANCOVA) model. The ANCOVA model will be fitted with the treatment as a fixed effect, baseline value as a covariate and treatment-by-baseline interaction. The adjusted group mean (least squares mean) changes will be comapared at a 2-sided significance level of 0.05. The *p* value along with the 95% confidence interval for the treatment difference will be reported.

6.2 Secondary Efficacy Endpoints and Analyses

Secondary efficacy endpoints will include: 1) the WOMAC pain scores at Day 5, 19, 64 and Day 94; 2) WOMAC scores on the stiffness and function subscales as well as total WOMAC score at Day 5, 19, 35, 64, and 94; 3) VAS OAKP assessments.

Data from WOMAC pain, stiffness, physical function subscales and total score, as well as VAS assessments will be listed and summarized by presenting descriptive statistics of raw data and change from baseline at each visit for each treatment group. Tabulation will be based on Study Knee and Non-study Knee.

WOMAC stiffness, function and total score reductions at Day 35 will be evaluated using the Study Knee. Change from baseline in WOMAC stiffness, function and total score on the Study Knee between each Active Treatment Arm and the Vehicle Arm at Day 35 will be evaluated using

ANCOVA model as described in Section 6.1.

The group mean reduction in OAKP score on a 100 mm VAS scale at Day 35 for each Active Treatment Arm will be compared to the Vehicle Arm using ANCOVA model as described previously.

For dose-response estimation, the difference of adjusted group mean between the 5% and the 1% Active Treatment Arms for the WOMAC pain, stiffness, function and total score at Days 35, 64 and 94 will be tested using the same method as described in Section 6.1.

6.3 Clinical Response

Clinical response will be measured by reduction of WOMAC pain score on the Study Knee. In order for a subject to be considered to have successfully obtained a clinical response, the following criteria must be met: WOMAC pain score on the Study Knee reduced to 50% or less of baseline at any of the stated assessment times.

A subject with a missing WOMAC pain score on the Study Knee at a visit will be considered to have failed to achieve a clinical response for that visit.

The durability of clinical response up to Day 94 will be evaluated by a tabulation of clinical response at Days 19, 35, 64, and the Day 94 visit by treatment group. Subjects who have successfully met the above criteria (reduction of a least 50% in WOMAC pain score) at the Day 5 visit and who remain at least at this reduction of pain score or lower at the Days 19, 35, 64, and the Day 94 visit will be considered to have a durable clinical response through Day 94. Subjects who at no more than one of the post Day 5 visits have less than 50% reduction in WOMAC pain score will be considered to have a durable response through the last day at which reduction in WOMAC pain score is at least 50%. Patients who have less than 50% WOMAC pain score reduction on two or more of the post Day 5 visits will be considered to have failed to achieve a durable clinical response.

Each subject's clinical response status (Responder vs. Non-responder) at each visit and the durability of clinical response up to Day 94 will be summarized for each treatment group. Percent change from baseline in WOMAC pain score on both knees will be listed at each visit.

6.4 Exploratory Analyses

Post hoc tests may be performed to test for differences in the efficacy endpoints between the CGS-200-0 (Vehicle), CGS 200-1 and CGS-200-5 treatment groups for exploratory purposes.

6.5 Tolerability Analyses

The primary tolerability endpoint will be tolerability of the formulations on each of Days 1, 2, 3, and 4. On each of these days, subjects will report an NRS score for burning / stinging at the application site pre-treatment and during treatment, and a pre-treatment and during treatment pruritus score. Pre-treatment and during treatment erythema and scaling scores will be recorded by clinic staff assigned to the study. This will be done at 0, 15, 30, 60, and 90 minutes after application to capture time points during treatment and after washing the application site. The score for each of the above tolerability variables will be entered as a value at each assessment time.

For each subject, an area under the curve (AUC) calculation will be performed in 90 minute intervals for each score value. The AUC for a specified time interval will be calculated using the

trapezoidal rule. If we have n+1 measurements yi at times ti (i=0, 1, …, n), then the AUC is calculated as $AUC = \frac{1}{2}\sum_{i=0}^{n-1}(t_{i+1} - t_i)(y_i + y_{i+1})$.

For each of Days 1 through 4, the average of AUC from left and right knees will be calculated for each of the tolerability variables. The calculated average of AUC from all the tolerability variables will then be combined and compared between each Active Treatment Arm (CGS-200-5 and CGS-200-1) and the Vehicle Arm (CGS-200-0) using analysis of variance (ANOVA).

A listing will be provided for all the tolerability variables at each visit.

7. SAFETY ANALYSIS

Safety assessments will include assessment of DLT, AEs, chemistry, hematology (complete blood count and differential), coagulation profile, urinalysis, and vital signs. All safety summaries (or analyses if applicable) will be conducted using the SAF set. No formal hypothesis testing will be performed to compare differences between treatment groups.

7.1 Extent of Exposure

The estimated percentage of study drug applied to each knee, the reason for full dose not being applied, and the duration of study drug application on each knee at each visit will be summarized for each treatment group.

7.2 Adverse Events

The AE monitoring period begins with the start of study drug administration on Day 1 and continues through Day 94.

An AE is any unfavorable or unintended sign, symptom, or disease that is temporally associated with the use of an investigational drug, but it is not necessarily caused by the investigational drug. This includes worsening (e.g., increase in frequency or severity) of pre-existing conditions. AEs will be classified by SOC and PT using the most recent version of Medical Dictionary for Regulatory Activities (MedDRA).

7.2.1 Treatment-emergent Adverse Events

A treatment-emergent AE (TEAE) is defined as an event that first occurs or worsens in intensity after the administration of study drug.

7.2.2 Adverse Event Severity

Severity will be graded using the Rheumatology Common Toxicity Criteria (RCTC), v2.0[2]. The RCTC is provided in Appendix B of the protocol. In the event that an AE does not have a RCTC code, the following severity classifications will be used:

- Grade 1-Mild: Causing no limitation of usual activities
- Grade 2-Moderate: Causing some limitation of usual activities
- Grade 3-Severe: Causing inability to carry out usual activities
- Grade 4-Life Threatening: Potentially life threatening or disabling

7.2.3 Adverse Event Relationship to Study Medication

The investigator is required to provide an assessment of causality or relationship of AEs to the study drug based on 1) temporal relationship of the event to the administration of study drug; 2) whether an alternative etiology has been identified; and 3) biological plausibility. The causality assessment categories that will be used for this study are "Not Related" (referring to "None" or "Unlikely"), and "Related" (referring to "Possible", "Probable", or "Definite"). See the study protocol, Section 8.1.3.3 for full details.

7.2.4 Serious Adverse Events

A serious adverse event (SAE) is an AE resulting in any of the following outcomes:

- Death
- Life-threatening (**immediate** risk of death)
- Inpatient hospitalization
- Prolongation of existing hospitalization
- Persistent or significant disability or incapacity
- Congenital anomaly/birth defect
- Medically important

7.2.5 Adverse Event Summaries

All AEs (serious and non-serious) occurring after study drug administration and before the end of study, regardless of relationship to study drug, will be included and classified by SOC and PT using MedDRA.

For TEAEs, the following will be summarized and presented for the SAF set:

i. An overall summary of TEAEs, which includes:
    a. the number and percentage of subjects experiencing a TEAE
    b. the number and percentage of subjects experiencing a TEAE by strongest relationship to study medication
    c. the number and percentage of subjects experiencing a TEAE by highest RCTC toxicity grade
    d. the number and percentage of subjects experiencing a DLT
    e. the number and percentage of subjects experiencing a treatment emergent SAE (TESAE)
    f. the number and percentage of subjects experiencing a TEAE leading to death
    g. the number and percentage of subjects experiencing a TEAE leading to study withdrawal

ii. the number and percentage of subjects experiencing a TEAE by SOC and PT.

iii. the number and percentage of subjects experiencing a TEAE by SOC, PT and the strongest

relationship to study medication

iv. the number and percentage of subjects experiencing a TEAE by SOC, PT and the highest RCTC toxicity grade

v. the number and percentage of subjects experiencing a DLT by SOC and PT

vi. the number and percentage of subjects experiencing a TESAE by SOC and PT

vii. the number and percentage of subjects experiencing a TEAE leading to study withdrawal by SOC and PT

In the overall summary of TEAEs table (i), besides tabulating the number and percentage of subjects, the total number of TEAE episodes will also be provided. If a subject has repeated episodes of a particular TEAE, all episodes will be counted in the summary table.

In the remaining summary tables, the incidence of TEAEs will be calculated by dividing the number of subjects who have experienced the event by the total number of subjects in the SAF set. Thus, the incidence of TEAEs is shown in terms of the total number of subjects and not in terms of the total number of episodes. If a subject has repeated episodes of a particular TEAE, only the most severe episode, or the episode with the strongest causal relationship to study drug, will be counted in the summary tables.

A subject with more than one type of TEAE in a particular SOC will be counted only once in the total of subjects experiencing TEAEs in that particular SOC. Since a subject could have more than one type of TEAE within a particular SOC, the sum of subjects experiencing different TEAEs within the SOC could appear larger than the total number of subjects experiencing TEAEs in that SOC. Similarly, a subject who has experienced a TEAE in more than one SOC will be counted only once in the total number of subjects experiencing AEs in all SOCs.

All occurrences of all AEs and SAEs will be listed for each subject, grouped by treatment group. The listing will contain the following information: treatment group, verbatim term, SOC, PT, toxicity grade, relationship to study medication, date and day of onset, date and day of resolution, treatment given to treat the adverse event, the outcome, whether the event was an SAE, whether it led to study withdrawal, whether the event was a DLT, and whether it is a TEAE. Listings will be sorted by treatment group, subject identification number, onset date, SOC, and PT. If the onset date is completely missing, then these events will be presented first. If the onset date is missing a month or a day, then these events will be presented before any complete dates.

7.3 Clinical Laboratory Assessments

Clinical laboratory tests will be obtained at the time points presented in Appendix A. Continuous clinical laboratory values will be summarized by presenting descriptive statistics of raw data and change from baseline values at each time point for each treatment group. Qualitative results at each time point measured will be summarized by presenting the number and percentage of subjects for each category.

Assessments from selected chemistry and hematology parameters can be graded based on the RCTC grading for laboratory data presented in the protocol Appendix B. A frequency distribution of the RCTC v2.0 severity at baseline, at each post-baseline assessment, and the worst value after baseline will be prepared. For each assessment after the baseline value the table will also provide the frequency of subjects with data, the frequency of subjects with Grade 3 or 4, and the frequency

of subjects with Grade 3 or 4 adjusted for baseline. For the latter frequency, subjects must have a baseline value and a value at the post-baseline time point, and subjects with the same or worse toxicity grade at baseline are not considered. Results will be tabulated and presented using counts and frequencies.

A listing of lab results with grade 1 or higher toxicity will also be provided.

7.4 Vital Signs

Descriptive summaries of the vital signs (both raw and change from baseline values) including systolic and diastolic blood pressure, heart rate, respiratory rate, and body temperature will be prepared for each treatment group by visit. Descriptive statistics for quantitative variables will include n, mean, standard deviation, median, minimum, and maximum. The 95% confidence interval of the mean will be constructed if it is appropriate (when there is more than 1 subject in a treatment group).

Vital sign values will be classified as low, normal or high based on the cut-off points defined below, and the data will be summarized by presenting shift tables from baseline to each post-baseline visit. Unscheduled visits will be excluded from the summaries but will be included in the data listings.

For Baseline vital sign measurements:

Measurement	**Lower**	**Upper**	**Unit**
Temperature	<35.0	>38.0	C
Systolic Blood Pressure (BP)	<110	>135	mmHg
Diastolic BP	<65	>85	mmHg
Heart Rate	<50	>110	beats/min
Respiratory Rate	<12	>20	breaths/min

For Post-baseline vital sign measurements:

Measurement (unit)	**Low**	**High**
Temperature (C)	<35	>38
Systolic Blood Pressure (mm Hg)	≥ 20 point decrease from baseline	≥ 20 point increase from baseline
Diastolic Blood Pressure (mm Hg)	≥ 20 point decrease from baseline	≥ 20 point increase from baseline
Pulse (beats per minute (bpm))	< 50 **and** decrease from baseline ≥ 15	> 100 **and** increase from baseline ≥ 15
Respiratory Rate (breaths per min)	<12	>20

8. INTERIM ANALYSES

The scope of this document details the final analysis to be performed by CTI, and does not describe any interim analysis that may be performed.

9. SAMPLE SIZE AND POWER CALCULATIONS

A sample size analysis has been performed and suggests that a sample size of 40 evaluable subjects per each of the three study arms will appropriately power this study. WOMAC pain subscale includes 5 components ranging from 0 to 500 points. The table below shows that 40 subjects per treatment group would provide 80% power to detect a difference of 63 between the treatment arms with respect to change from baseline in WOMAC pain score, assuming the SD of the changes from baseline in WOMAC A is 100.

Power to Detect Difference Δ Between Treatments

With Respect to Changes from Baseline (2-sided α = 5%, n=40/Group)

	Δ = Treatment Difference Between Mean Changes from Baseline									
SD (CFB†)	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**63**	**65**
100	20%	27%	35%	43%	52%	61%	69%	77%	**80%**	83%
105	19%	25%	32%	40%	48%	57%	65%	72%	77%	79%
110	17%	23%	30%	37%	45%	53%	61%	68%	73%	75%

†CFB = Change from Baseline, SD(CFB) = Standard Deviation of CFB

10. REFERENCES

1. American National Standards Institute. Sampling Procedures and Tables for Inspection by Attributes, ANSI/ASQC Z1.4-1993

2. Woodworth T, Furst DE, Alten R, Bingham C, Yocum D, Sloan V, Tsuji W, Stevens R, Fries J, Witter J, Johnson K, Lassere M, Brooks P. Standardizing assessment and reporting of adverse effects in rheumatology clinical trials II: the Rheumatology Common Toxicity Criteria v.2.0. J Rheumatol. 2007;34(6):1401-14.

11. APPENDIX

11.1 Appendix A: Schedule of Events

Assessment	SCR	Treatment Days				Follow-up Period Days					
TIMEFRAME	Within 28 days of Day 1	D1	D2	D3	D4	D5 (Post Treat Day 1)	D19± 3d (Post Treat Week 2)	D35± 3d (Post Treat Week 4)	D64± 3d (Post Treat Week 8)	D94± 3d (Post Treat Week 12)	Unscheduled, Early Termination, SAE[l]
Clinic visit	x	x	x	x	x	x	x	x	x	x	x
Informed Consent	x										
Inclusion / Exclusion	x										
Demographics	x										
Medical & Medication History	x										
Randomization		x									
Physical exam	x										x
Height and Weight[a]	x										
Vital Signs[b]	x	x	x	x	x	x				x	x
Knee X-rays[c]	x										
12-lead ECG[d]	x										
Clinical lab tests[e]	x					x		x		x	x
Pregnancy or FSH Test[f]	x	x									
Application Site Pain, Erythema, Scaling, Pruritus Assessment[g]	x	x	x	x	x	x	x	x	x	x	x
Exploratory Assessment of OA Knee Pain[h]	x	x									
VAS OA Knee Pain Assessment[i]		x	x	x	x	x	x	x	x	x	x
WOMAC[j]	x	x				x	x	x	x	x	x
Drug administration		x	x	x	x						
Assess. Of Concomitant Medications[k]	x	x	x	x	x	x	x	x	x	x	x
AE Assessments		x	x	x	x	x	x	x	x	x	x
Study Exit										x	

Schedule of Events (Continued)

a. Height and weight obtained at Screening.

b. Vital signs include temperature, blood pressure, heart rate and respiratory rate. Performed at Screening, on Days 1-4 at pre-dose and at 90 minutes (± 5 min) after applying study drug to knees, and at Follow-up on Day 5 and 94.

c. Knee X-rays of both knees must be obtained unless films/images performed within 6 months before screening. Copies of film and results will be requested.

d. Standalone 12-lead ECG performed at Screening.

e. Clinical laboratory tests include (Appendix A):

Hematology: full blood count including RBC, hemoglobin, hematocrit, MCV, MCH, MCHC, WBC, differential white cell count, platelet count (Screening, Day 5, Day 35 and Day 94);

Erythrocyte sedimentation rate (Screening only);

Chemistry: BUN, creatinine, uric acid, bilirubin (total), sodium, potassium, calcium, magnesium, phosphorous (inorganic), chloride, bicarbonate, ALK, AST (SGOT), ALT (SGPT), LDH, GGT, CPK, albumin, total protein, and glucose (Screening, Day 5, Day 35 and Day 94);

IgM RF (Screening only);

Urinalysis: pH, protein, glucose, ketones, bilirubin, blood, urobilinogen, specific gravity, leucocytes (Screening, Day 5, Day 35 and Day 94). Clinically significant abnormal dipstick results will require microscopic analysis.

f. Serum pregnancy test at Screening and urine pregnancy test at baseline on Day 1 for premenopausal women; FSH (follicle stimulating hormone) test at Screening to confirm postmenopausal status (females only) for up to 2 years after cessation of menses.

g. Application site pain will be assessed on NRS (0-10); application site erythema/edema will be assessed using modified categorical Draize test (0-3). Scaling and pruritus will also be scored on a similar categorical scale (0-3). Scoring for application site pain (not related to knee pain) will be performed on drug administration Days 1-4 (pre-application [≤ 30 min], and at 15 (± 5 min), 30 (± 5 min), 60 (± 5 min) and 90 minutes (± 5 min) after application) and on Days 5, 19, 35, 64 and 94. Application site skin reactions (erythema, edema, scaling, pruritus) will be assessed on drug administration Days 1-4 at pre-application (≤ 30 min), and at 15 (± 5 min), 30 (± 5 min), 60 (± 5 min) and 90 minutes (± 5 min) after application), and on Days 5 (24 hrs [+/- 4 hrs]), 19 (+/- 3 days), 35 (+/- 3 days), 64 (+/- 3 days) and 94 (+/- 3 days). (Appendix E and Appendix F)

h. Exploratory assessment of OA knee pain (average daily pain and daily worst pain) at Screening and Day 1 (pre-application [≤30 min]) (Appendix D).

i. OA Knee pain to be scored on 100 mm VAS scale (Appendix G) pre-treatment (≤ 30 min) on Days 1, 2, 3 and 4 and also on clinic visits at Days 5 (24 hrs [+/- 4 hrs]), 19 (+/- 3 days), 35 (+/- 3 days), 64 (+/- 3 days) and 94 (+/- 3 days).

j. WOMAC Pain, Stiffness, and Physical Function will be performed at Screening and on Day 1 prior to drug administration (≤ 30 min) and on Days 5 (24 hrs [+/- 4 hrs]), 19 (+/- 3 days), 35 (+/- 3 days), 64 (+/- 3 days) and 94 (+/- 3 days).

k. Document all medications taken within 90 days of the Screening Visit through Day 94.

l. Refer to Section 6.5 for assessments to be completed at time of SAE, early termination or unscheduled visit.