STATISTICAL ANALYSIS PLAN SM04690-OA-02

(Version 1.0, October 14, 2016)

A Phase 2, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Safety, Tolerability, and Efficacy of SM04690 Injected in the Target Knee Joint of Moderately to Severely Symptomatic Osteoarthritis Subjects

NCT02536833

STATISTICAL ANALYSIS PLAN SM04690-OA-02

Study Title:	A Phase 2, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Safety, Tolerability, and Efficacy of SM04690 Injected in the Target Knee Joint of Moderately to Severely Symptomatic Osteoarthritis Subjects
Protocol Number:	SM04690-OA-02
Protocol Version:	Amendment 04, October 13, 2016
SAP Version:	1.0
Investigational Product:	SM04690 Injectable Suspension
Clinical Phase:	Phase 2

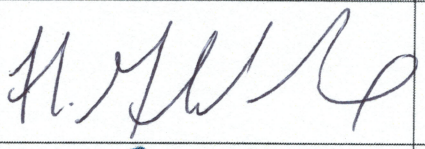
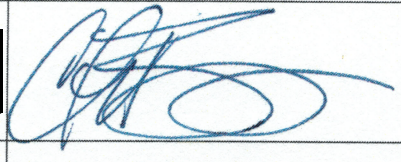
Sponsor: Samumed, LLC.
9381 Judicial Dr., Suite 160
San Diego, CA 92121
(858) 926-2900

SAP SIGNATURE PAGE

A Phase 2, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Safety, Tolerability, and Efficacy of SM04690 Injected in the Target Knee Joint of Moderately to Severely Symptomatic Osteoarthritis Subjects

Version: 1.0

Date: October 14, 2016

Name & Title	Signature	Date
[redacted]		14 Oct 2016
[redacted]		14 Oct 2016
[redacted]		Oct 14 2016

Samumed commits to satisfying the requirements of the ICH-GCP Guidelines regarding the responsibilities of the Sponsor, the US Code of Federal Regulations 21 CFR parts 50, 54, 56, 312, and 314 and Good Clinical Practice Guidelines, as applicable.

TABLE OF CONTENTS

1. INTRODUCTION AND OBSERVATIONS

Osteoarthritis (OA) is the most common form of arthritis that causes degenerative structural change and results in pain and decreased mobility. Samumed, LLC is developing SM04690, a small molecule inhibitor of the Wnt pathway, for the treatment of OA. The purpose of this study is to assess the efficacy, safety and tolerability of three different strengths of SM04690 administered by intra-articular injection into the target knee joint of moderately to severely symptomatic OA subjects.

1.1. Primary Objectives

The primary objective of this study will be to evaluate the change from baseline OA pain in the target knee as assessed by the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain subscore at Week 13.

1.2. Secondary Objectives

The secondary objectives of this study include:

- Evaluate change from baseline OA pain in the target knee as assessed by the WOMAC pain subscore at Week 26
- Evaluate change from baseline OA function in the target knee as assessed by the WOMAC function subscore at Weeks 13 and 26
- Evaluate change from baseline OA disease activity as assessed by the Patient Global Assessment at Weeks 13 and 26
- Evaluate change from baseline in medial joint space width (JSW) as documented by X-ray of the target knee at Week 26
- Evaluate the safety and tolerability of SM04690 by monitoring for treatment-emergent adverse events (TEAEs)

1.3. Exploratory Objectives

Additional exploratory objectives of this study include:

- Evaluate change from baseline in WOMAC total score for the target knee at Weeks 4, 13, 26, 39 and 52
- Evaluate change from baseline OA pain in the target knee as assessed by the WOMAC pain subscore at Weeks 4, 39 and 52
- Evaluate change from baseline OA function in the target knee as assessed by the WOMAC function subscore at Weeks 4, 39 and 52
- Evaluate change from baseline OA disease activity as assessed by the Patient Global Assessment at Weeks 4, 39 and 52
- Evaluate change from baseline in medial JSW as documented by X-ray of the target knee at Week 52
- Evaluate change from baseline in lateral JSW as documented by X-ray of the target knee at Weeks 26 and 52

- Determine the percentage of Outcome Measures in Rheumatology Clinical Trials (OMERACT)-Osteoarthritis Research Society International (OARSI) responders and "strict" responders at Weeks 4, 13, 26, 39 and 52
- Evaluate change from baseline health-related quality of life (HRQOL) as assessed by the 36-Item Short Form Health Survey (SF-36) at Weeks 4, 13, 26, 39 and 52
- Evaluate change from baseline OA disease activity as assessed by the Physician Global Assessment at Weeks 4, 13, 26, 39 and 52

2. STUDY DESIGN

This study will be a multicenter, randomized, double-blind, placebo-controlled, parallel group study of three different strengths of SM04690 injected into the target knee joint of moderately to severely symptomatic osteoarthritis subjects.

2.1. Sample Size

A sample size of approximately 445 subjects will be randomized at a ratio of 1:1:1:1 (0.03 mg active per 2 mL injection : 0.07 mg active per 2 mL injection : 0.23 mg active per 2 mL injection : placebo). The sample size for this study was based upon accepted statistical practice to establish an acceptable level of precision with respect to treatment effect estimation (Piantadosi 1997).

To explore possible power expected from this study design, an analysis of covariance regression model was simulated via a Monte Carlo simulation of 1000 replicates. The simulation model estimated a change in WOMAC Pain from baseline comparison between SM04690 and placebo adjusting for baseline WOMAC Pain. The following assumptions were made for the power simulation:

- Sample size of 100 subjects per group,
- Type 1 error rate of 5%,
- WOMAC Pain Baseline average is 11 points [scale 0-20] with a standard deviation of 3,
- WOMAC Pain Baseline is the same for both treatment and placebo groups,
- Treatment change from baseline is -5.8 points, leading to a final score of 5.2 points [scale 0-20]
- Historical placebo change from baseline is -3.5 points, leading to a final score of 7.5 points [scale 0-20]
- Final pain score has standard deviation of 4 and is the same for both treatment and placebo groups, and
- Correlation between baseline and final pain score is 0.1.

The WOMAC Pain subscore parameters were based upon observed data from the 0.07 mg group in SM04690-01. Based upon these assumptions, SM04690-OA-02 is estimated to have 95.8% power.

A simulation model for WOMAC Function was also estimated. The following assumptions were made for the power simulation:

- Sample size of 100 subjects per group,
- Type 1 error rate of 5%,
- WOMAC Function Baseline average is 39 points [scale 0-68] with a standard deviation of 11,
- WOMAC Function Baseline is the same for both treatment and placebo groups,
- Treatment change from baseline is -19.5 points, leading to a final score of 19.5 points [scale 0-68]
- Historical placebo change from baseline is -13.5 points, leading to a final score of 25.5 points [scale 0-68]
- Final function score has standard deviation of 15 and is the same for both treatment and placebo groups, and
- Correlation between baseline and final function score is 0.1.

The WOMAC Function subscore parameters were also based upon observed data from the 0.07 mg group in SM04690-01. Based upon these assumptions, SM04690-OA-02 is estimated to have 78.5% power.

No prior data exist for Patient Global Assessment in SM04690-01; no formal power analysis is possible at this time.

2.2. Randomization

Subjects will be randomized 1:1:1:1 (0.03 mg active per 2 mL injection : 0.07 mg active per 2 mL injection : 0.23 mg active per 2 mL injection : placebo) to each dose group using Medidata Balance with a permuted block design stratified by study site. A block size of 8 has been selected for this study.

2.3. Study Medication Dosing

Three strengths of SM04690 injectable suspension and placebo will be used in this study:

- SM04690 0.03 mg in 2-mL Injectable Suspension
- SM04690 0.07 mg in 2-mL Injectable Suspension
- SM04690 0.23 mg in 2-mL Injectable Suspension
- SM04690 0 mg in 2-mL Injectable Suspension; placebo (phosphate buffered saline)

3. STUDY OUTCOME MEASURES

3.1. Safety Outcome Measures

TEAEs will be presented in summary tables and categorized by seriousness and toxicity severity grade. While all TEAEs will be considered as related to study product, the investigator's assessment of relatedness will be analyzed for informational purposes. TEAEs will also be summarized according to MedDRA dictionary terms, including system organ class and preferred term.

3.2. Efficacy Outcome Measures

Knee pain and function will be assessed by the subject using the WOMAC Index. Disease activity will be assessed by the Patient Global Assessment of Disease Activity and the Physician Global Assessment of Disease Activity. Overall HRQOL will be assessed with the SF-36 questionnaire which provides scores for eight scales (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, mental health) and two summary measures (physical health, mental health). JSW will be assessed with X-Ray imaging of the target knee at Week 26 and 52.

3.2.1. Western Ontario and McMaster Universities Arthritis (WOMAC) Index

The WOMAC NRS3.1 instrument is a patient reported outcome measure to assess the symptoms of OA. Subjects will be asked to complete the WOMAC instrument based on their symptoms in the target knee.

The WOMAC consists of 24 questions in three domains: physical function (17 questions), pain (5 questions) and stiffness (2 questions). The response for each question in the NRS format ranges from 0 to 10. Each domain subscore as well as a total score are calculated by adding together the numerical responses. Subsequently, the WOMAC NRS3.1 ranges from 0 to 240 total points.

If any response is missing, the score for that domain (i.e. physical function, pain, or stiffness) as well as the WOMAC Total score will not be able to be calculated.

3.2.2. OMERACT-OARSI Response

The OMERACT-OARSI response (Pham et al. 2003) is defined as follows:

A. High improvement from baseline in pain or function greater than or equal to 50% and absolute change of at least 20 points [scaled on a 0-100 range], OR

B. Improvement from baseline in at least two of the following

   - pain greater than or equal to 20% and absolute change of at least 10 [scaled on a 0-100 range]

- function greater than or equal to 20% and absolute change of at least 10 [scaled on a 0-100 range]
- Patient Global Assessment greater than or equal to 20% and absolute change of at least 10 [scaled on a 0-100 range]

Additionally, the OMERACT-OARSI "strict" response is defined by any response satisfying the above condition A only.

Both OMERACT-OARSI response and OMERACT-OARSI "strict" response are efficacy outcome measures in this study. WOMAC Pain and WOMAC Function subscores will be utilized and scaled appropriately to calculate the OMERACT-OARSI response.

3.2.3. Patient Global Assessment of Disease Activity

The patient assessment will be a 100 mm VAS scale adapted from the Patient Assessment Form © 1999, American College of Rheumatology. The subject will rate how well they are doing, considering all the ways in which illness and health conditions may affect him/her. The VAS scale will be anchored by "Very Well" on the left and "Very Poorly" on the right.

3.2.4. Physician Global Assessment of Disease Activity

The physician assessment will be a 100 mm VAS scale adapted from the Physician Assessment Form © 1999, American College of Rheumatology The investigator will rate the subject's disease activity. The VAS scale will be anchored by "Very Good" on the left and "Very Bad" on the right.

3.2.5. Joint Space Width (JSW)

JSW in the target knee will be assessed by radiograph. A radiograph may be repeated if the image quality of the original radiograph is deemed unsatisfactory by the imaging vendor, Medical Metrics, Inc. (MMI), according to Image Review Charter. The radiograph with the best image quality (where applicable) will be used for analysis as long as it was taken within the visit window. If two or more radiographs for the same study visit are assessed as having the same image quality, the earliest radiograph will be considered for that visit. Medial and lateral joint space width measurements will be provided by the vendor.

3.2.6. Health-Related Quality of Life

Overall HRQOL will be assessed with the SF-36 Version 2 questionnaire which consists of 36 questions that are asked of the subject. The subject's responses will be used to calculate norm-based scores for eight scales (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, mental health) and two component summary measures (physical health, mental health). Scoring software provided by the SF-36 vendor, Optum, will be utilized to calculate the standardized scores.

4. ANALYSIS SETS

4.1. Intent-to-Treat Analysis Set

The Intent-to-Treat (ITT) Analysis Set includes all subjects who were randomized.

4.2. Modified Intent-to-Treat Analysis Set

The Modified Intent-to-Treat (mITT) Analysis Set includes all ITT subjects who received a protocol specified dose of SM04690 or placebo, analyzed as treated. Subjects incorrectly receiving doses not prescribed by the protocol are excluded from this analysis set.

4.3. Per Protocol Analysis Set

The Per Protocol (PP) Analysis Set includes all ITT subjects who complied with all study procedures and evaluations and did not have major protocol deviations.

4.4. Safety Analysis Set

The Safety Analysis Set includes all subjects who were exposed to study product.

5. STATISTICAL CONSIDERATIONS

5.1. Interim Analysis

By definition, "[a]n interim analysis is any analysis intended to compare treatment arms with respect to efficacy or safety at any time prior to formal completion of a trial" (ICH E9 Guidance 1999). Analysis of the specified efficacy endpoints (Section 8.2) will be conducted after all subjects complete the Week 26 visit but prior to the formal completion of the trial at 52 weeks. Thus, by definition, these efficacy analyses will be described as an interim analysis. However, no trial adaptation(s) will be made based upon the results of the interim analysis.

The datasets supporting this analysis will be validated from a snapshot of the raw study database. Source data verification of the raw data supporting the interim analysis as well as query resolution will be completed before unblinding for the interim analysis. All available data supporting the Week 26 JSW analysis (i.e. radiographs) will be measured by the imaging vendor and included in the snapshot dataset prior to unblinding for the interim analysis. In addition, all available safety events occurring up to and including the Week 26 visit will also be included in the snapshot dataset prior to unblinding and summarized in the interim analysis report.

All available data in support of the interim analysis will be dated and sequestered from the final analytical databases. Any changes made to the sequestered data in the raw database after the interim analysis that impacts any outcome within the interim efficacy analysis will be summarized in the clinical study report.

5.2. Handling of Missing Data

Multiple imputation will be utilized for the primary and secondary analyses. Missing data will be imputed based upon the observed data for each outcome under the missing at random (MAR) assumption following the paradigm first developed by Rubin (Schafer 1999). The imputation will adjust for the outcome's baseline data in a regression model with data imputed independently for each timepoint; any missing baseline data will also be imputed. A total of 10 imputation datasets will be created and analyzed based upon accepted convention (Schafer 1999). Error estimates of the multiple imputation itself as well as an overall summary of the efficacy analysis will be averaged across the 10 imputed datasets based upon Rubin's paradigm.

Specifically, let Q denote the imputed mean of an efficacy outcome Y subject to missing data at a given timepoint, where the estimate of that mean for any i^{th} imputation would be defined as function of both observed and missing data such that

$$\widehat{Q^{(i)}} = E\left(Q^{(i)} | Y_{observed}, Y_{missing}^{(i)}, Y_{Baseline}, I(Group)\right), i = 1, \ldots, 10.$$

After all 10 datasets have been imputed, the overall estimate of Q is a simple average defined as

$$\bar{Q} = \frac{\sum_{i=1}^{10} \widehat{Q^{(i)}}}{10}.$$

For this analysis using 10 imputed datasets, the variance of $\bar{Q}$ is defined as

$$Var(\bar{Q}) = 1.1B + \bar{U}$$

where B is the between imputation variance and U is the within imputation variance:

$$B = \frac{1}{9}\sum_{i=1}^{10}\left(\widehat{Q^{(i)}} - \bar{Q}\right)^2;$$

$$\bar{U} = \frac{\sum_{i=1}^{10} U^{(i)}}{10}.$$

Imputations will be performed with a seed of 201610141.

5.2.1. Tipping Point Analysis

The MI paradigm, as proposed by Rubin, does not require nor assume any specific assumption about the nature of the missing data. Further, imputations can be (theoretically) created under any assumption of missing mechanism. A sensitivity analysis will be conducted using the tipping point approach to evaluate the robustness of the primary analysis with multiple imputation under the MAR assumption. With the tipping point approach, several additional multiple imputation datasets are created, each departing slightly from the original assumptions used in the primary analyses. The intended result is to establish a certain point in which the new shifted imputations have "tipped", or overturned, the original conclusions from the primary analysis (i.e. a statistically significant result is now not statistically significant). The magnitude of the shift required to tip the analysis are then evaluated for plausibility; an implausible shift will lead to the conclusion that the multiple imputation analysis is robust to the departure from MAR assumption.

Tipping point analysis will be conducted only for the dose group(s) that is/are statistically significant under the MAR assumption. Based on the SM04690 Phase I study, the placebo change in WOMAC pain from baseline was observed at -3.5 points. When scaled to the NRS format, this translates to a -8.75 reduction in WOMAC pain. The primary analysis will be reassessed by shifting the mean change in WOMAC pain in the treatment group by intervals of +0.5 until the historical placebo threshold of -8.75 (or the placebo response observed in SM04690-OA-02, whichever is more conservative) is reached and/or surpassed. Intervals of +0.1 may further be explored in order to determine a more precise shift parameter for the tipping point. An implausible shift parameter is defined as one that results in the magnitude of change in WOMAC pain in the treatment group to equal or worse than historical (or SM04690-OA-02, again whichever is more conservative) placebo change.

Imputations for the tipping point analysis will be performed with a seed of 201610142.

5.3. Baseline

Baseline is defined as the last value recorded for any given parameter prior to study medication injection. However, for radiographs used to evaluate JSW, baseline is defined as the assessment prior to study medication injection with the best image quality as determined by the imaging vendor, where applicable. If two or more radiographs are assessed with the same image quality, the earliest radiograph will be considered for baseline.

5.4. Early Termination

If a subject discontinues the study, early termination assessments will be performed according to the protocol. If these assessments occur within the window of a scheduled visit (±3 days for Weeks 4, 13 and 39 and -10 days for Week 26), they will be associated with that visit for the purposes of ITT and mITT analysis. All other visits will be considered missing data and treated as described in section 5.2.

6. STUDY SUBJECTS

6.1. Disposition of Subjects

Subject disposition will be presented in a summary table detailing the number and percentage of subjects who were consented, randomized, treated, completed the study or discontinued (e.g. screen failure, subject decision, etc.) by treatment group and site. The disposition for individual subjects will be listed along with additional information on discontinued subjects.

6.2. Demographic and Baseline Characteristics

Demographic and baseline characteristics, including gender, race, ethnicity, age, weight, height, body mass index (BMI), Kellgren-Lawrence (KL) grade, presence of bilateral OA and

Widespread Pain Index (WPI) score will be summarized by treatment group and listed for individual subjects.

6.3. Medical History

Medical history will be collected at screening and reassessed at the Week 52 (EOS) / Early Termination visit to record any changes. A summary of reported medical history will be provided by system category for each treatment group. A subject-level listing will provide further information on each event.

6.4. Protocol Deviations

For SM04690-OA-02, a protocol deviation is defined as any change, divergence or departure from the study design or procedures of a research protocol that is under the Investigator's control and that has not been approved by the site's Institutional Review Board.

Deviations are summarized into one of the following categories:

- Those who were randomized to the study even though they did not satisfy the eligibility criteria
- Those who developed withdrawal criteria during the study but were not withdrawn
- Those who received the wrong treatment or incorrect dose
- Those who received an excluded concomitant treatment
- Those who missed or received an incorrect protocol procedure

Deviations are categorized as major or minor.

- A major deviation is defined as a divergence from the protocol that materially (a) reduces the quality or completeness of the data, (b) makes the informed consent inaccurate, or (c) impacts a subject's safety, rights or welfare.
- A minor deviation is defined as a divergence from the protocol that deviates from the procedures and guidelines outlined in the protocol, but is not classified as a major deviation (i.e. the deviation does not materially (a) reduce the quality or completeness of the data, (b) make the informed consent inaccurate, or (c) impact a subject's safety, rights or welfare).

Protocol deviations will be summarized by site, category and classification, and listed for each subject.

7. SAFETY EVALUATION

The analysis of safety outcome measures will be performed on the Safety Analysis Set. The interim report will include all available safety information up to and including Week 26. The final report will include a cumulative assessment safety.

7.1. Treatment-Emergent Adverse Events

TEAEs will be presented in an overview summary table depicting the number and percent of subjects in each treatment group experiencing TEAEs by seriousness and toxicity grade severity. While all TEAEs will be considered as related to study product, the investigator's assessment of relatedness will be included in the summary table for informational purposes. TEAEs will also be summarized by MedDRA system organ class and preferred term, including the number of TEAEs and the number and percent of unique subjects experiencing each TEAE.

Separate subject-level listings will be provided for all serious and non-serious TEAEs.

7.2. Clinical Laboratory

All chemistry, hematology and urinalysis results from the central lab will be summarized into shift tables as normal, non-clinically significant abnormal and clinically significant abnormal. Assessments of clinical significance for abnormal values will be made by the investigator on results that are outside of the normal range or have a toxicity grade of 1 or greater. The shift tables will compare the number and percent of assessments from each visit to baseline values for each treatment group. Abnormal results for each subject will be provided in listings that will include assay name, result, normal range and an explanation for clinically significant values.

Abnormal urine microscopy and manual differential assessments will be listed for each subject.

7.3. Vital Signs

Weight and vital signs, including systolic blood pressure, diastolic blood pressure, pulse rate, respiratory rate and body temperature, will be summarized for each treatment group. A statistical description of each parameter at baseline will be provided along with the change from baseline at each subsequent visit. A subject-level listing will also be provided.

7.4. Concomitant Medications

The World Health Organization Drug Dictionary (WHODD) will be used to classify prior and concomitant medications by Anatomical Main Group (Anatomical Therapeutic Chemical, ATC, Level 1), Therapeutic Subgroup (ATC Level 2), and preferred term. Prior and concomitant medication usage will be summarized by the number and percentage of subjects receiving each medication by treatment group.

Subject-level listings containing prior and concomitant medications (WHODD coding), and procedures and non-drug therapies (MedDRA coding) will be provided.

8. EFFICACY EVALUATION

All efficacy analysis will be performed on the ITT analysis set in support of regulatory requirements. Efficacy analysis will also be performed on the mITT and PP Analysis Sets.

8.1. General Considerations

For continuous variables within each treatment group, the outcome measure at each visit, as well as absolute change (outcome – baseline), will be summarized using descriptive statistics. Categorical variables will be summarized with frequency tables.

Subject-level listings will be provided for each outcome measure in the final analysis report.

8.1.1. ANCOVA Model for Continuous Efficacy Outcomes

ANCOVA models will be used for all continuous efficacy outcome measures summarized by change from baseline in order to test the following three hypotheses:

$$H0: (\beta i - \beta_0) = 0$$

$$HA: (\beta i - \beta_0) \neq 0, \quad \text{where } i = 1,2,3$$

In the statement above, β is the least squares estimate in the change in the continuous efficacy outcome from baseline at each timepoint (where β_0 is the estimate for placebo), and *i* represents each of the three SM04690 treatment groups.

Least squares estimate of difference between each treatment group and placebo in the change in the continuous efficacy outcome from baseline at each timepoint, adjusted for baseline value, will be reported along with unadjusted 95% confidence intervals and P values.

Continuous efficacy measures to be analyzed by the ANCOVA model include:

- WOMAC Pain
- WOMAC Function
- WOMAC Total
- Patient Global Assessment
- Physician Global Assessment
- Medial and Lateral Joint Space Width
- SF-36

8.1.2. Logistic Regression Model for Binary Efficacy Outcomes

Logistic regression will be used to analyze the proportion of positive responses within each treatment group compared to placebo at each timepoint for all binary efficacy outcomes. Odds ratios as well as unadjusted 95% confidence intervals and P values will be presented.

Binary efficacy measures to be analyzed by logistic regression include:

- OMERACT-OARSI “strict” response
- OMERACT-OARSI response

8.2. Interim Analysis of Efficacy Outcomes

Analysis of all endpoints in support of the primary and secondary objectives will be conducted after all subjects complete the Week 26 visit but prior to the formal completion of the trial at 52 weeks.

This will include:

- Evaluate change from baseline OA pain in the target knee as assessed by the WOMAC pain subscore at Weeks 13 and 26
- Evaluate change from baseline OA function in the target knee as assessed by the WOMAC function subscore at Weeks 13 and 26
- Evaluate change from baseline OA disease activity as assessed by the Patient Global Assessment at Weeks 13 and 26
- Evaluate change from baseline in medial joint space width (JSW) as documented by X-ray of the target knee at Week 26

Additionally, endpoints in support of exploratory objectives will also be analyzed, including:

- Evaluate change from baseline OA pain in the target knee as assessed by the WOMAC pain subscore at Week 4
- Evaluate change from baseline OA function in the target knee as assessed by the WOMAC function subscore at Week 4
- Evaluate change from baseline in WOMAC total score for the target knee at Weeks 4, 13 and 26
- Evaluate change from baseline OA disease activity as assessed by the Patient Global Assessment at Week 4
- Evaluate change from baseline in lateral JSW as documented by X-ray of the target knee at Week 26
- Determine the percentage of Outcome Measures in Rheumatology Clinical Trials (OMERACT)-Osteoarthritis Research Society International (OARSI) responders and "strict" responders at Weeks 4, 13 and 26
- Evaluate change from baseline OA disease activity as assessed by the Physician Global Assessment at Weeks 4, 13 and 26

8.2.1. Familywise Error Rate Control

The familywise error rate for the interim efficacy analyses will be controlled in the strong sense using the closed, fixed sequence testing method (Dmitrienko et al 2010). All hypothesis tests will be evaluated in the pre-specified sequential order that matches clinical inference from prior SM04690 studies regarding the relative therapeutic benefit of each dose. If a test fails to meet the

critical value of $\alpha = 0.05$, all subsequent tests will not be evaluated for the interim efficacy analysis and only be considered as exploratory analysis.

For this study, the pre-specified, fixed hypothesis hierarchy is:

H1. Change in WOMAC Pain at Week 13, 0.07 mg SM04690 vs Placebo
H2. Change in WOMAC Pain at Week 13, 0.03 mg SM04690 vs Placebo
H3. Change in WOMAC Pain at Week 13, 0.23 mg SM04690 vs Placebo
H4. Change in WOMAC Function at Week 13, 0.07 mg SM04690 vs Placebo
H5. Change in WOMAC Function at Week 13, 0.03 mg SM04690 vs Placebo
H6. Change in medial JSW at Week 26, 0.07 mg SM04690 vs Placebo
H7. Change in medial JSW at Week 26, 0.03 mg SM04690 vs Placebo
H8. Change in WOMAC Function at Week 13, 0.23 mg SM04690 vs Placebo
H9. Change in medial JSW at Week 26, 0.23 mg SM04690 vs Placebo
H10. Change in Patient Global Assessment at Week 13, 0.07 mg SM04690 vs Placebo
H11. Change in Patient Global Assessment at Week 13, 0.03 mg SM04690 vs Placebo
H12. Change in Patient Global Assessment at Week 13, 0.23 mg SM04690 vs Placebo
H13. Change in WOMAC Pain at Week 26, 0.07 mg SM04690 vs Placebo
H14. Change in WOMAC Pain at Week 26, 0.03 mg SM04690 vs Placebo
H15. Change in WOMAC Function at Week 26, 0.07 mg SM04690 vs Placebo
H16. Change in WOMAC Function at Week 26, 0.03 mg SM04690 vs Placebo
H17. Change in WOMAC Pain at Week 26, 0.23 mg SM04690 vs Placebo
H18. Change in WOMAC Function at Week 26, 0.23 mg SM04690 vs Placebo
H19. Change in Patient Global Assessment at Week 26, 0.07 mg SM04690 vs Placebo
H20. Change in Patient Global Assessment at Week 26, 0.03 mg SM04690 vs Placebo
H21. Change in Patient Global Assessment at Week 26, 0.23 mg SM04690 vs Placebo

8.2.2. Subgroup Analysis

The primary and secondary efficacy analysis described above will be further explored by the following covariates:

- KL Grade 2 vs. KL Grade 3
- Age Groups: age < 65, $65 \leq$ age < 75, age ≥ 75
- BMI Groups: BMI < 30, $30 \leq$ BMI < 35, BMI ≥ 35
- Bilateral knee vs. unilateral knee OA

- WPI: WPI < 13, WPI ≥ 13

Each covariate will be added to the original ANCOVA model along with a term for the interaction between treatment group and the covariate at Week 13 for WOMAC pain, WOMAC function and Patient Global, and Week 26 for medial JSW. The significance of each interaction term will be determined using a likelihood ratio test that compares the original model to the model with the interaction. If the significance level of the interaction is $P < 0.10$ at Week 13 for WOMAC pain, WOMAC function and Patient Global, and Week 26 for medial JSW, the interaction model will be evaluated at those timepoints as well as subsequent timepoints.

8.3. Final Analysis of Efficacy Outcomes

Interim efficacy analyses will not be repeated but will be included in the final report. Any changes to the data that was sequestered and analyzed for interim will be summarized and explained. In addition, the following objectives will be explored:

- Evaluate change from baseline in WOMAC total score for the target knee at Weeks 39 and 52
- Evaluate change from baseline OA function in the target knee as assessed by the WOMAC function subscore at Weeks 39 and 52
- Evaluate change from baseline OA in the target knee as assessed by the WOMAC Total at Weeks at Weeks 39 and 52
- Evaluate change from baseline OA disease activity as assessed by the Patient Global Assessment at Weeks 39 and 52
- Evaluate change from baseline in medial and lateral JSW as documented by X-ray of the target knee at Week 52
- Determine the percentage of Outcome Measures in Rheumatology Clinical Trials (OMERACT)-Osteoarthritis Research Society International (OARSI) responders and "strict" responders at Weeks 39 and 52
- Evaluate change from baseline OA disease activity as assessed by the Physician Global Assessment at Weeks 39 and 52
- Evaluate change from baseline health-related quality of life (HRQOL) as assessed by the 36-Item Short Form Health Survey (SF-36) at Weeks 4, 13, 26, 39 and 52

9. ANALYSIS SOFTWARE

All data processing, summarization and analyses will utilize SAS® Version 9.4. QualityMetric Health Outcomes Scoring Software 4.5 will be used to score SF-36 questionnaires.

10. REFERENCES

Dmitrienko, A., A. C. Tamhane, and F. Bretz. (2010). Multiple testing problems in pharmaceutical statistics. Boca Raton, Chapman & Hall.

ICH E9 Guidance. (1999). "ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group." Stat Med **18**(15): 1905-1942.

Pham, T., D. Van Der Heijde, M. Lassere, R. D. Altman, J. J. Anderson, N. Bellamy, M. Hochberg, L. Simon, V. Strand, T. Woodworth, M. Dougados and O. Omeract (2003). "Outcome variables for osteoarthritis clinical trials: The OMERACT-OARSI set of responder criteria." J Rheumatol **30**(7): 1648-1654.

Piantadosi, S. (1997). Clinical Trials: A Methodologic Perspective. New York, Wiley-Interscience.

Schafer, JL. (1999) “Multiple imputation: a primer.” Stat Methods Med Res **8**: 3-15.