# Statistical Analysis Plan

| | |
|---|---|
| Sponsor: | Levicept Ltd. |
| Protocol No: | LEVI-04-21-02 |
| Protocol Title: | A Phase 2a, Randomized, Double-Blind, Placebo-Controlled Trial of the Efficacy and Safety of LEVI-04 in Patients with Osteoarthritis of the Knee |
| Document Date: | 18-Mar-2024 |
| Document Revision | Final 3.0 |

## DOCUMENT HISTORY

| Rev No | Date | Description |
|---|---|---|
| 1.0 | 01-Sep-2023 | Original document |
| 2.0 | 08-Dec-2023 | The addition of analyses for X-Ray data, standardization of the WOMAC subscale scores, and clarification on the handling of individual missing questionnaire items. |
| 3.0 | 18-Mar-2024 | Update to the derivation for daily average amount of rescue medication usage and the proportion of days on rescue medication to account for days with missing data and participants that discontinue early. |

# TABLE OF CONTENTS

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ACR | American College of Rheumatology |
| ADA | Anti-Drug Antibody |
| AE | Adverse Events |
| AESI | Adverse Events of Special Interest |
| ANOVA | Analysis of Covariance |
| BCTQ | Boston Carpal Tunnel Questionnaire |
| BLQ | Below Limit of Quantification |
| BMI | Body Mass Index |
| BT | Body Temperature |
| BP | Blood Pressure |
| CRO | Clinical Research Organization |
| eCRF | Electronic Case Report Form |
| ECG | Electrocardiogram |
| EOT | End of Trial |
| FCS | Fully Conditional Specification |
| FDA | Food and Drug Administration |
| FU | Follow-up |
| GCP | Good Clinical Practice |
| GDPR | General Data Protection Regulation |
| GLP | Good Laboratory Practice |
| GMP | Good Manufacturing Practice |
| HA | Health Authorities |
| HR | Heart Rate |
| HV | Healthy Volunteer |
| IA | Interim Analysis |

| | |
|---|---|
| IB | Investigators Brochure |
| ICF | Informed Consent Form. |
| ICH | The International Conference on Harmonization |
| IEC | Independent Ethics Committees |
| IMP | Investigational Medicinal Product |
| IRB | Institutional Review Boards |
| ITT | The Intention-to-Treat |
| IWRS | Interactive Web Responding System |
| JSW | Joint Space Width |
| KL | Kellgren Lawrence |
| LLOQ | Lower Limit of Quantification |
| mg | Milligrams |
| MI | Multiple Imputations |
| mITT | Modified Intention-to-Treat |
| mJSW | Minimum Joint Space Width |
| mm | Millimeters |
| MRI | Magnetic resonance imaging |
| NGF | Nerve Growth Factor |
| NOAEL | No Observed Adverse Effect Level |
| NRS | Numeric Rating Scale |
| NT3 | Neurotrophin-3 |
| NSAID | Non-steroidal Anti-inflammatory Drug |
| OA | Osteoarthritis |
| PCR | Polymerase Chain Reaction |
| PD | Pharmacodynamics |

| | |
|---|---|
| PDF | Portable Document Form |
| PGA | Patient Global Assessment |
| PK | Pharmacokinetics |
| PP | Per-Protocol |
| PT | Preferred Term |
| QTcF | Heart rate corrected QT interval (Fridericia) |
| RPOA | Rapidly Progressive Osteoarthritis |
| SAE | Serious Adverse Event |
| SAP | Statistical Analysis Plan |
| SIF | Subchondral Insufficiency Fractures |
| SD | Standard Deviation |
| SOC | System Organ Classes |
| SSRI | Selective Serotonin Reuptake Inhibitor |
| SSS | Symptom Severity Scale |
| StEPP | Staircase Evoked Pain Procedure |
| SUSAR | Suspected Unexpected Serious Adverse Reactions |
| TEAE | Treatment Emergent Adverse Event |
| WOMAC | The Western Ontario and McMaster Universities Osteoarthritis Index |
| WOPAIN | WOMAC Pain Subscale |

## 1. INTRODUCTION

Symptomatic knee osteoarthritis (OA) is a painful chronic joint disease estimated to affect approximately 45% of the population during a lifetime[1]. Clinical manifestations include joint stiffness, swelling, crepitus, and limited mobility, but the hallmark symptom of OA is pain[2]. Pain is the driving symptom leading patients to seek treatment and the symptom that most affects quality of life[3]. OA pain etiology, however, is complex, heterogenic, and yet to be fully understood[4].

Among several available guidelines for management of knee OA pain The European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO)[5] and Osteoarthritis Research Society International (OARSI)[6], are consistent in the majority of their recommendations: Following core treatments (e.g. patient education and exercise programs), both the OARSI and ESCEO recommend the use of topical Non-Steroidal Anti-inflammatory Drugs (NSAIDs) and short-term use of paracetamol (acetaminophen) as first-line pharmaceutical management of knee OA pain. Patients with persistent pain are recommended oral NSAIDs, however, only to be used intermittently for the shortest period and at the lowest possible dose, due to their known adverse cardiovascular, hepatic, and renal risks. If such treatments are insufficient or not well tolerated, both guidelines recommend intraarticular corticosteroid injection or/and hyaluronic acid injections. Though highly efficient in the short term (2-4 weeks) intraarticular injections with corticosteroids are associated with osteoporosis, decreased efficiency when used long- term (over 6 weeks),[7] and risk of accelerated cartilage loss[8]. As a last option, prior to recommending surgical intervention, guidelines recommend short-term treatment with opioids or treatment with the Selective Serotonin Reuptake Inhibitors (SSRI), Duloxetine.

In conclusion treatment options for sufficient long-term OA pain management is limited and/or unsustainable, possibly leading clinicians to overprescribe drugs with serious and complicated long-term effects such as NSAIDs and opioids.

Despite numerous attempts in clinical drug development to provide a durable, safe, and effective treatment option, none have currently been successful. Targeted drug development of pain medication and further research on OA pain pathology is therefore warranted.

LEVI-04's primary mode of action is inhibition of the activity of the Neurotrophins; Nerve Growth Factor (NGF) and Neurotrophin-3 (NT3) to reduce pain.

NGF is an important mediator of pain associated with OA[9]. Monoclonal antibodies against NGF showed efficacy in treating chronic pain due to OA in phase III clinical trials[10,11]. However, risk

of Rapidly Progressing Osteoarthritis (RPOA)[12] and adverse effects on the sympathetic nervous system[13] raised concerns regarding their safety.

LEVI-04 is a fusion protein and thus has a different antagonist profile to anti-NGF antibodies and differs on several points:

- Unlike anti-NGF antibodies that require plasma concentrations ~1000-fold the primary affinity to demonstrate efficacy in clinical studies[10], LEVI-04 is efficacious in preclinical models of OA at concentrations of its potency for inhibition of NGF activity.
- LEVI-04 exhibits a more gradual inhibition of NGF-induced activity compared with anti-NGF antibody mediated inhibition.
- Multiple preclinical studies, including comparative histopathology and neuropathological evaluation studies of LEVI-04 and anti-NGF antibodies, have been performed specifically to address any safety concerns (See IB for full overview). These studies confirmed the detrimental effects of anti-NGF antibodies, but found no evidence of such effects of LEVI-04 at therapeutic doses

Together, this indicates LEVI-04 not to be associated with the known risk of RPOA associated with anti-NGF antibodies. Conversely, treatment with LEVI-04 was found to be associated with improvements in the histopathological assessments compared to the control animals, which could indicate a beneficial disease-modifying effect of LEVI-04. These preclinical findings support the investigation of LEVI-04 as a novel drug for the treatment of OA.

For a full description of the modalities of LEVI-04 and study references, see the study protocol and Investigators Brochure (IB).

### 1.1 Study Design

This is a Multiple Arm, Multicenter, Prospective, Randomized, Double-blind, Placebo-controlled, Phase 2a study of LEVI-04 intravenous infusion for the treatment of knee osteoarthritis. The purpose of the trial is to evaluate the efficacy, safety, and tolerability of five monthly infusions of LEVI-04 as compared to placebo in participants with radiographic and symptomatic knee osteoarthritis. Eligible participants will be randomized to one of the four treatment arms: 0.3mg/kg LEVI-04, 1.0 mg/kg LEVI-04, 2.0mg/kg LEVI-04 or saline vehicle control (placebo) in a 1:1:1:1 ratio (Figure 1). All participants will receive five intravenous IMP infusions 4 weeks apart, i.e. at Randomization (Day 1), Visit 5 (Week 4), Visit 7 (Week 8), Visit 8 (Week 12) and Visit 9 (Week 16) with assessments as per Schedule of Events and A Follow-up/EOT visit (Visit 11) and a Follow-up Safety Phone visit (Visit 12) will take place at Week 20 and Week 30, respectively.

### 1.2 Rationale

The overall trial design is considered an appropriate study design for Phase 2 drug development and consistent with current regulatory guidance. The primary goal of the study is to establish proof of concept of the biologic activity of LEVI-04 leading to reduction in pain from knee OA compared to placebo. The trial is designed to add to the LEVI-04 safety data collected in the Phase 1 trial (ClinicalTrials.gov: NCT03227796, See IB) and establish the most suitable dose for a Phase 3 trial. The Primary endpoint is change in WOMAC pain subscale score from Randomization (Visit 3) to week 17 (Visit 10). WOMAC pain subscale is a Food and Drug Administration (FDA) recommended clinical trial efficacy endpoint for osteoarthritis and considered a reliable, validated outcome measure14,15. The study duration is 30 weeks including 4 weeks to Follow-up/EOT visit from last dose and considered sufficient to meet study objectives. Radiographs of large joints bilaterally (knee, hip, and shoulder joints) and MRI of both knees during Screening are included to determine eligibility. Radiographs of large joints, and MRI of the Target Knee will be completed at Follow-up/EOT visit (week 20). Safety assessments throughout the trial include Physical examination, Vital signs, Clinical laboratory values, electrocardiogram (ECG) data, and Adverse Events (AE).

## 2. STUDY OBJECTIVES AND ENDPOINTS

**Table 1: Objectives and mapped Endpoints**

| Objective(s) | Endpoint(s) |
|---|---|
| Primary | |
| To evaluate the efficacy of LEVI-04 (multiple doses) compared to placebo in reducing pain due to knee OA. | Change in WOMAC pain subscale score from Randomization (Visit 3) to Visit 10 (week 17) |
| Secondary | |
| To evaluate the efficacy of LEVI-04 (multiple doses) compared to placebo in improving physical function. | Change in WOMAC Physical function subscale score from Randomization (Visit 3) to Visit 6 (week 5) and Visit 10 (week 17). |
| | Change in StEPP from Randomization (Visit 3) to Visit 6 (week 5) and Visit 10 (week 17). |

| Objective(s) | Endpoint(s) |
|---|---|
| To evaluate the efficacy of LEVI-04 (multiple doses) compared to placebo in improving joint stiffness. | Change in WOMAC Stiffness subscale score from Randomization (Visit 3) to Visit 6 (week 5) and Visit 10 (week 17). |
| To evaluate the efficacy of LEVI-04 (multiple doses) compared to placebo in Patient Global Assessment (PGA). | Change in PGA from Visit 6 (week 5) to Visit 10 (week 17). |
| To evaluate the efficacy of LEVI-04 (multiple doses) compared to placebo in reducing pain. | Change in WOMAC Pain subscale score from Randomization (Visit 3) to Visit 6 (week 5). |
| | Proportion of participants achieving 30% and 50% reduction in WOMAC Pain subscale score at week 5 and week 17 using a cumulative distribution function. |
| | Change in average weekly NRS score from Randomization (Visit 3) to Visit 6 (week 5) and Visit 10 (week 17). |
| | Area under the curve of Average Daily NRS pain from Randomization (Visit 3) to Visit 11 (week 20). |
| To evaluate rescue medication use in the LEVI-04 group (multiple doses) compared to placebo. | Rescue Medication usage during the trial (average amount used per day while on study treatment, proportion of days of use while on study treatment) |
| Exploratory | |
| To evaluate the efficacy of LEVI-04 compared to placebo in time to onset of pain relief. | Time from Randomization to a decrease in average weekly (NRS) pain of 30% from Baseline NRS Pain score. |
| To explore associations between imaging biomarkers and pain severity. | Associations of imaging biomarkers and pain severity at Randomization (Visit 3) to Visit 11 (week 20). Correlations between pain variability during screening and major efficacy endpoints |
| To explore associations of biomarkers of the neurotrophin signaling pathway and pain severity | Change from baseline in blood biomarkers and associations with clinical outcomes |

| Objective(s) | Endpoint(s) |
|---|---|
| Safety | |
| To evaluate the overall safety of LEVI-04 compared to placebo. | Adverse Events (AEs) between treatment arms (Count and Severity) |
| To evaluate the safety of LEVI-04 compared to placebo in relation to the peripheral nervous system. | Peripheral Nervous System AEs between treatment arms (Count and Severity) |
| To evaluate the safety of LEVI-04 compared to placebo in relation to joint safety. | Incidences of Joint Events and subtypes of Joint Events. Events include:<br>o Osteonecrosis<br>o Destructive Arthropathy, including Rapidly Progressive Osteoarthritis (RPOA) type 2<br>o Changes in signs of inflammation on MRI of the Target Knee,<br>o Subchondral insufficiency fracture<br>o Rapid loss of joint space width (JSW) by 2 mm within 12 months or loss of 50% of JSW if less than 2 mm at baseline. This is also called RPOA type 1.<br>o Any other incidence of joint events such as acute fracture, bone marrow infiltration, etc. |
| | Incidence of surgical interventions including joint replacement. |

# 3. PRIMARY ESTIMAND

| Primary Estimand | |
|---|---|
| Primary scientific question of interest | What is the optimal dose of LEVI-04 compared to placebo in participants with knee OA? |
| Target Population | Defined through appropriate inclusion/exclusion criteria as listed in the study protocol. |
| Treatments of Interest | LEVI-04 intravenous infusion dose 0.3 mg/kg<br>LEVI-04 intravenous infusion dose 1.0 mg/kg<br>LEVI-04 intravenous infusion dose 2.0 mg/kg |
| Outcome Variable | Change in WOMAC pain subscale score from Randomization (Visit 3) to Visit 10 (week 17) |
| Summary Measure | Least square mean difference between active doses and placebo and associated p-values from ANCOVA model with fixed effects for treatment, and covariates site and Baseline WOMAC pain subscale score, adjusting for multiplicity using Dunnett's step-down procedure. |
| Intercurrent Events | Discontinuations due to:<br>• Adverse Event<br>• Lack of efficacy |

# 4. STATISTICAL METHODOLOGY

## 4.1 General Principles

All collected study data will be presented in listings. All derivations and statistical analyses will be performed using SAS® software Version 9.4. Prior and concomitant medications will be coded using the current version of the World Health Organization (WHO) Drug Dictionary available at the start of the study and will be used for the duration of the study even if a newer version of the dictionary is released. Similarly, adverse events (AEs) and medical history will be coded using the current version of the Medical Dictionary for Regulatory Activities (MedDRA) available at the start of the study and will be used for the duration of the study.

Unless stated otherwise, descriptive summaries for continuous variables will include n (number of participants with non-missing results), mean, standard deviation (SD), median, minimum, and

maximum. For categorical variables, n and percent will be provided. Percentages will be calculated out of the number of participants in the given population with a non-missing result.

Nominal outcome summaries, presented by collection day and/or by time point, will be based on scheduled assessments as planned in the protocol. Unscheduled assessments will be presented in listings and will not be included in the analyses except for determination of Baseline as described in Section 5.1. All other visit and time point collections will be as recorded on the electronic case report forms (eCRFs).

WOMAC subscale scores will be standardized to a range of 0 to 10. Imputation of missing WOMAC pain subscale scores described in Section 4.9 will be based on the original score range. Standardization will occur after imputation at the time of analysis. Listings will present the original scores.

The primary analysis will be conducted once data collection associated with this analysis has been completed. That is, no additional data will be collected that would affect the results of the primary analysis despite the study being ongoing.

Unless otherwise stated, hypothesis testing will be two-sided with an alpha level of 0.05.

## 4.2 Randomization

The screening/randomization procedure will be centrally managed through an electronic Interactive Web Responding System (IWRS) integrated in the EDC system. To enroll and randomize a new participant, the Investigator/authorized personnel will have to access the EDC system by entering their username and password and filling in the requested data. Randomization will be centrally balanced across the study (not by center).

A total of approximately 416 participants are planned to be randomized in a 1:1:1:1 to the placebo and three active arms. If needed based on the interim conditional power analysis, the sample size will increase up to 624.

## 4.3 Sample Size Considerations

This study is planned to enroll approximately 416 participants, with a possibility of adding an additional 208 participants pending a conditional power analysis done at the interim analysis. It is assumed that there is a minimum effect of 10 out of 100 in change from baseline in WOMAC pain subscale score for at least one of dose groups compared with placebo, and a standard deviation of 19.2 for change from baseline. Dropout rate is assumed to be 15%, distributed evenly over the study duration of the study. An interim analysis (IA) will be conducted when approximately 208

participants have completed the Week 5 visit. Based on the IA's conditional power analysis, the study will continue as planned with a total of 416 participants randomized, or with an increase of up to 208 participants (an additional 50%) up to a total of 624 participants. The final analysis will be conducted between the three treatment arms and placebo using a simple combination test with weights of 1/2 on the interim Z and 1/2 on the final Z, using a Dunnett's test. With this design, the power for a statistically significant difference between at least one active treatment group versus placebo assuming the study will continue to at least the originally planned sample size is:

| Effect scenario: treatment mean difference versus placebo | | | Power with sample size of 416 | Increase of sample size to 624 at the interim analysis | | |
|---|---|---|---|---|---|---|
| Arm 1 | Arm 2 | Arm 3 | SD=19.2 | SD=22 | SD=23 | SD=24 |
| 10 | 7.5 | 5 | 90% | 94% | 92% | 90% |
| 10 | 5 | 2.5 | 88% | 92% | 89% | 86% |
| 10 | 0 | 0 | 86% | 91% | 88% | 85% |

## 4.4 Multiplicity Adjustment and Sample Size Reassessment

Dunnett's step-down multiple comparison procedure will be used to compare the three active arms to placebo. Dunnett adjusted p-values for three comparisons will be determined initially (equivalent to a two-sided type-1 error of approximately 0.0188 for each comparison) to maintain an overall two-sided type-1 error of 0.05. If any of the three comparisons is found to be significant (adjusted two-sided p-value ≤ 0.05), the remaining two comparisons will be made using Dunnett adjusted p-values for two comparisons (equivalent to a two-sided type 1 error of approximately 0.0270 for each comparison). If either one is significant (adjusted two-sided p-value ≤ 0.05), the last comparison will be made at a two-sided 0.05 level of significance.

In order to protect the overall type-1 error given the sample size recalculation, a simple combination test will be used to combine the statistics for each treatment comparison from the interim and post-interim data, assigning 1/2 weight to the interim Z-value, and 1/2 to the post-interim Z-value (i.e., the Z-value of the post-interim participants are calculated from the participants not included in the interim analysis). The final Z-value will be calculated as: $Z = Z_1 (\frac{1}{2})^{0.5} + Z_2 (\frac{1}{2})^{0.5}$ where $Z_1$ is the interim Z-value, and $Z_2$ is the Z-value calculated on participants not included in the interim analysis. Z-values for the comparisons of each dose group

to placebo for the interim and post interim participants will be calculated by transforming the one-sided p-values (obtained from the ANCOVA one-sided t-statistics) for each comparison to one-sided Z-values.

## 4.5 Analysis Sets

**The Intention-to-Treat Analysis Set**

The Intent-to-Treat (ITT) Analysis Set will include all participants randomly allocated to a treatment, based on the intention to treat "as randomized" principle (i.e., the planned treatment regimen rather than the actual treatment given in case of any difference).

**Modified Intention-to-Treat Analysis Set**

The Modified Intention-to-Treat (mITT) Analysis Set will include all participants from the ITT Analysis Set who have a baseline and at least one post-treatment WOMAC pain subscale score assessment available.

**Per-Protocol Analysis Set**

The Per-Protocol (PP) Analysis Set will include all participants from the mITT Analysis Set who have been treated according to the trial protocol and fulfill the following criteria:

- Absence of major clinical trial protocol deviations with respect to factors likely to affect the efficacy of the treatment.
- Adequate compliance with trial medication: participants must have at least four study drug administrations, including an administration at Week 16.

Protocol deviations will be reviewed prior to unblinding to determine which, if any, participants are to be excluded from the Per-Protocol Analysis Set.

**Safety Analysis Set**

The Safety Analysis Set will include all participants who have been administered at least one dose of trial treatment. Participants will be analyzed according to the actual treatment they receive.

## 4.6 Participant Disposition

The number of participants screened and the number of participants not continuing beyond screening and the reason why will be summarized overall. The number of participants in each of the analysis sets will be summarized by treatment, site, and overall, where applicable.

The number and percentage of participants who completed the study will be presented by treatment, and overall for the ITT Analysis Set. Frequency and percentage of participants who withdrew or discontinued from the study and the primary reason for withdrawal, will also be summarized by treatment and overall.

## 4.7 Participant Demographics and Baseline Characteristics

Demographic data and participant characteristics at Baseline will be summarized descriptively. Medical history will be presented by MedDRA system organ class (SOC) and preferred term (PT).

## 4.8 Prior and Concomitant Medications

Prior and concomitant medications are classified as follows.

- A prior medication is any medication taken prior to the first dose of study drug.
- A concomitant medication, (i.e., concomitant with study drug) is any medication taken on or after the first dose of study drug.

A medication may be classified as both prior and concomitant. When determining prior or concomitant status for a medication, in the event that the start or end dates of the medication are unknown or incomplete, the medication will be considered as prior and concomitant unless the non-missing date information, if any, is enough to conclude that the medication could not be prior or, separately, could not be concomitant. Incidence of prior and concomitant medication use, according to medications collected on the prior/concomitant medication CRF page (i.e., excluding rescue medications from the rescue medication dispensing and accountability CRFs), will be summarized by World Health Organization (WHO) Drug dictionary coded terms Anatomical Therapeutic Chemical (ATC) classification and preferred name (PN).

## 4.9 Missing data

The WOMAC osteoarthritis index is a self-administered questionnaire consisting of 24 items divided into three subscales: pain (5 items), stiffness (2 items), and physical function (17 items). Scores for the subscales are based on the sum of the responses (each ranging from 0, no symptoms, to 10, worst imaginable) within the subscale. If participants fail to complete all questions, then missing responses will be imputed as the mean of the non-missing responses within the same subscale provided that:

- Pain Subscale: no more than one response is missing,
- Stiffness Subscale: no more than one response is missing,
- Function Subscale: no more than three responses are missing.

If the number of missing responses for a given subscale meets the specified criterion, then the overall score for the subscale is calculated according to the non-missing and imputed responses and considered as an observed response. If more than the specified number of responses is missing, the subscale is considered invalid and the subscale score is considered missing. All non-missing responses will be presented in listings regardless of the number of missing responses.

Missing data for the primary endpoint (WOMAC pain subscale score) will be imputed using multiple imputations (MI) methodology. All missing visits data will be imputed using all prior visits data as well as treatment. Data that is missing at random will be imputed using the randomized treatment assignment, whereas data that is missing due to treatment (adverse event, lack of efficacy) will be imputed using placebo as the treatment. For the MI model, 100 imputations will be generated using PROC MI of SAS®. Fully conditional specification (FCS) model using the regression method will be used with non-missing baseline and post-baseline WOMAC pain subscale, and other important covariates to impute any missing data including missed visits and missing data due to dropouts. The seed to be used is 20220420. The results from the MI will then be used to derive the value of the change from baseline in the WOMAC pain subscale score at each timepoint, and then analyzed using the analysis of covariance (ANCOVA) model. The results of the 100 analyses will be transformed into a normal statistic and combined into a single analysis using PROC MIANALYZE.

Missing WOMAC pain subscale scores will be imputed as described above at the time of the interim analysis. This imputation will be used only for the interim analysis. Missing scores will be imputed again at the end of the study for all participants in the study, including those that were included in the interim analysis using all available data at the end of the study. Both the interim and end of study imputations will be based on the same methodology and seed. Participants may be ongoing in the trial at the time of the interim analysis, so they may have a missing Week 17 score at the time of the interim analysis and a non-missing score at the end of the study. With the inclusion of the post-interim analysis participants and additional collections for some of the interim participants at the end of the study, the analysis results at the end of the study for participants included in the interim analysis may differ from the results observed at the time of the interim analysis.

The Boston Carpal Tunnel Questionnaire (BCTQ) is a measure of self-reported severity of symptoms and functional status associated with carpal tunnel syndrome. The questionnaire

consists of a Symptom Severity subscale (11 items) assessing severity of wrist or hand pain, change of sensation, and weakness in hand or wrist the past two weeks and a Functional Status subscale (8 items) assessing ability to perform various hand/wrist related tasks. Overall scores for the subscales are based on the average of the responses within the subscale. Individual missing responses will be imputed as the average of the non-missing responses. The average score of a subscale will not be calculated if more than one response is missing.

Survey for Autonomic Symptoms is a two-part questionnaire comprising 11 (for females) or 12 (for males) items, validated for assessing autonomic symptoms. For each item the participant will be asked if he/she have had the symptom the last 6 months (Yes/No) and, if Yes, how much would they say the symptom bothered them on a score ranging from 1 (“Not at all”) to 5 ("A lot”). The overall score will be calculated as the sum of the responses, where “No” is scored as “0” and “Yes” responses are analyzed according to the score associated with how much the symptom bothered them. The number of participants with exactly one, exactly two, or three or more symptoms will be summarized. For the sum, individual missing responses will be imputed as the average of the non-missing responses. The average score will not be calculated if more than one response is missing. For the summary of the number of symptoms, missing responses will not be imputed. Participants with more than one missing response will be excluded from the number of symptoms summary unless the participant has checked “Yes” for three or more of the non-missing system responses.

## 4.10 Multicenter trials

Study center may have an impact on the study outcomes and therefore will be used as a covariate in the primary and secondary efficacy analyses.

## 4.11 Primary Endpoint Analysis

The primary population of all efficacy analyses will be the ITT Analysis Set. The change from baseline in WOMAC pain subscale score will be analyzed using an ANCOVA model with main effect treatment and covariates study center and baseline score. The primary timepoint for analysis will be week 17. Since an interim analysis is to be performed, the final analysis will be conducted between each of the three treatment arms and placebo using a simple combination test (1/2 weight on interim Z and 1/2 weight on final Z) in order to maintain the overall specified alpha level. Further, p-values will be adjusted (and compared to 0.05) using a step-down Dunnett testing method for comparison of the three active doses with placebo.

The step-down testing method will first apply Dunnett adjustment for three comparisons (0.3 mg/kg, 1.0 mg/kg, and 2.0 mg/kg LEVI-04) with placebo. If the lowest adjusted p-value is non-

significant (>0.05), then all three comparisons will be considered non-significantly different from placebo. If the lowest adjusted p-value is significant (≤0.05), then the associated LEVI-04 dose level will be considered significantly different from placebo. The testing will continue by applying a Dunnett adjustment for two comparisons (the remaining two LEVI-04 dose levels) with placebo. If the lowest adjusted p-value from the two comparison Dunnett adjustment is non-significant, then the remaining two dose levels will be considered non-significantly different from placebo. If the lowest adjusted p-value from the two comparison Dunnett adjustment is significant, then the associated LEVI-04 dose level will also be considered significantly different from placebo. Testing will continue by comparing the unadjusted p-value to 0.05 for the last LEVI-04 dose level versus placebo. This procedure preserves the global type 1 error. Adjusted p-values from the step-down Dunnett may not be smaller than the smallest adjusted p-value reported from the previous step. If the smallest Dunnett adjusted p-value from one of the steps is smaller than the previous step's adjusted p-value, then the p-value from the current step will be set to the previous step's adjusted p-value. Dunnett adjustments to the p-values will be calculated using the SAS® PROBMC function. In the first step, for each of the three comparisons, the Z from the associated combination test will be utilized by the PROBMC function as follows to calculate the adjusted p-value:

adj_p3 = 1 - PROBMC("DUNNETT2", max_abs_z, . , . ,3,lam1,lam2,lam3);

where,
max_abs_z = Maximum of the absolute value of all three combination Z-statistics
n0 = n01+n02: placebo sample size at the end of the study
n1 = n11+n12: LEVI-04 0.3 mg/kg sample size at the end of the study
n2 = n21+n22: LEVI-04 1.0 mg/kg sample size at the end of the study
n3 = n31+n32: LEVI-04 2.0 mg/kg sample size at the end of the study

where nij is the sample size of the ith group (i=0,1,2,3) at the interim (j=1) and post interim (j=2) periods.

$$\text{lam1\_1} = \sqrt{\frac{n11}{n11+n01}}$$

$$\text{lam2\_1} = \sqrt{\frac{n21}{n21+n01}}$$

$$\text{lam3\_1} = \sqrt{\frac{n31}{n31+n01}}$$

$$\text{lam1\_2} = \sqrt{\frac{n12}{n12+n02}}$$

lam2_2 = $\sqrt{\frac{n22}{n22+n02}}$

lam3_2 = $\sqrt{\frac{n32}{n32+n02}}$

Correlations among the interim Z-statistics

$\rho$12_1=lam1_1*lam2_1
$\rho$13_1=lam1_1*lam3_1
$\rho$23_1=lam2_1*lam3_1

Correlations among the post-interim Z-statistics

$\rho$12_2=lam1_2*lam2_2
$\rho$13_2=lam1_2*lam3_2
$\rho$23_2=lam2_2*lam3_2

Correlations among the (final) combined Z-statistics

$\rho 12 = (\rho 12\_1 + \rho 12\_2)/2$

$\rho 13 = (\rho 13\_1 + \rho 13\_2)/2$

$\rho 23 = (\rho 23\_1 + \rho 23\_2)/2$

lam1= $\sqrt{\frac{\rho 12 * \rho 13}{\rho 23}}$

lam2= $\sqrt{\frac{\rho 12 * \rho 23}{\rho 13}}$

lam3= $\sqrt{\frac{\rho 13 * \rho 23}{\rho 12}}$

adj_p3 = Dunnett adjusted p-value for the maximum absolute value of all three combination z-statistics.

The above procedure will be similarly applied to the remaining dose groups using PROBMC:

adj_p2 = 1 - PROBMC("DUNNETT2", max_abs_z, . , . ,2,lam_i,lam_j);

where:

max_abs_z: Maximum of the absolute value of the remaining two combination Z-statistics
lam_i,lam_j: the lambdas for the remaining groups

In the third step, the remaining adjusted p-value will be adj_p1=p1, i.e., no adjustment.
Following the calculation of the p-value adjustment, the following process will ensure monotonicity of the adjusted p-values:

```
adj_p3=adj_p3;
adj_p2=max(adj_p3,adj_p2);
adj_p1=max(adj_p2,adj_p1);
```

The WOMAC pain subscale score will be tabulated per time point (visit) and treatment group, along with the absolute change from baseline where baseline is defined as the latest assessment prior to first treatment administration.
For the treatment effect testing of IMP versus placebo, the null and the alternative hypothesis will be:

H0: Mean (Placebo) = Mean (IMP)

H1: Mean (Placebo) ≠ Mean (IMP)

If the null hypothesis is rejected, the alternative hypothesis will be accepted, and it will be concluded that the treatment effect of IMP differs from placebo.

The The SAS® Mixed procedure will be utilized as follows.

```
proc mixed;
  class site trt;
  model val = trt site base;
run;
```

where:
val = dependent variable (eg. change from baseline in WOMAC pain subscale score)
site = Investigational Site,
trt = Treatment (Arms A, B, C, D),
base = Baseline WOMAC pain subscale score

To assess the robustness of the primary results, sensitivity analysis will be performed with modification of participant population (mITT and PP Analysis Sets) using observed data only. Further investigation of the robustness of the primary results (in particular, the influence of time and covariates) may be performed, if deemed necessary.

## 4.12 Secondary Endpoint Analyses

Secondary endpoints will be analyzed based on the ITT Analysis Set using observed data only. Continuous or ordinal categorical endpoints will be analyzed using the same model as specified for the primary endpoint at each visit, where applicable, unless otherwise stated. Dichotomized secondary endpoints such as the Proportion of participants achieving 30% and 50% reduction in WOMAC Pain subscale scores at week 5 and week 17, will be analyzed using a logistic regression model, with PROC GENMOD, with main effect treatment and covariates study center and baseline WOMAC pain subscale score, and where the link function is logit, and the distribution is binomial. In addition to the ITT Analysis Set, WOMAC physical function subscale score, Staircase Evoked Pain Procedure (StEPP) Pain Intensity, WOMAC stiffness subscale score, PGA, and the Numeric Rating Scale (NRS) pain score will be summarized by the mITT and PP Analysis Sets.

NRS pain score is collected daily. Weekly post-baseline average daily scores will calculated for all weeks through Week 20 regardless of whether or not there was a scheduled visit at that week. Weeks will be calculated according to actual study days. Since daily NRS pain is reported in the evenings, it will be assumed that the Day 1 assessment is collected after dosing. Therefore, Week 1 will be calculated from NRS scores collected on Days 1 through 7, Week 2 from Days 8 through 14, etc. Baseline average daily pain score will be calculated as the average of scores collected on the seven days prior to first dose of study drug (Day 1). Missing daily pain scores will not be imputed. Weekly post-baseline average daily scores will be calculated if at least four of the seven scores are non-missing. The denominator for determining the weekly score will only include days with a non-missing score.

The area under the curve for average daily NRS pain will be calculated using the linear trapezoidal method from Day 1 through Week 20 (Day 141) or the last non-missing daily NRS pain score, whichever comes first. Missing daily pain scores will not be imputed. The area under the curve for the average daily NRS pain will be analyzed using the same model as the primary endpoint.

The average amount of daily rescue medication usage while on study treatment and the proportion of days of rescue medication use while on study treatment will be calculated based on the diary, which captures daily use of rescue medication. The average amount of rescue medication used per day while on study treatment will be calculated as the total number of tablets taken on or after the first dose of study drug divided by the total number of days with known medication usage [i.e.,

specified as no usage (0 tablets) or the number of tablets taken provided]. Similarly, the proportion of days of rescue medication use will be calculated as the number days rescue medication was used divided by the total number of days with known medication usage (i.e., usage specified as "yes" or "no"). These two rescue medication endpoints will be analyzed using an ANCOVA with main effect treatment and covariates sex, baseline BMI, baseline target knee WOMAC pain subscale score, and baseline OA category (unilateral, bilateral). Bilateral OA participants have both knees with a Kellgren Lawrence (KL) grade ≥ 2 and unilateral OA participants have only one knee with a KL grade ≥ 2.

## 4.13 Exploratory Endpoint Analysis

Continuous exploratory endpoints will be analyzed using the same model as the primary endpoint for the ITT Analysis Set using observed data only. Time from baseline to a decrease in average weekly (NRS) pain of 30% will be compared between treatments using a log-rank test stratified by study center and summarized using Kaplan-Meier estimates. Correlations between pain severity and imaging biomarkers and pain severity and efficacy endpoints will be presented graphically.

## 4.14 Additional Questionnaire Analyses

The BCTQ and the Survey for Autonomic Symptoms results will be summarized for the ITT Analysis Set using observed data only. Change from baseline in the average Symptom Severity subscale scores and average Functional status subscale scores for the BCTQ will be summarized by treatment and visit. Frequency and percentage of participants reporting at least 1, 2, or 3 symptoms according to the Survey for Autonomic Symptoms will be summarized by gender, treatment, and visit. The change from baseline in total symptom impact score according to the Survey for Autonomic Symptoms will be summarized by gender, treatment, and visit. No formal testing will be performed on the BCTQ or the Survey for Autonomic Symptoms results.

## 4.15 Subgroup Analysis

Subgroup analyses (especially for the analysis of the primary endpoint) will be considered for the categories of OA at baseline (unilateral, bilateral),country, and infusions received group (received all infusions, missed one or more infusions) from the mITT Analysis Set who have a baseline and at least one post-treatment WOMAC pain subscale score available.

## 4.16 Safety Analyses

Safety will be assessed by adverse events, laboratory assessments (including pharmacokinetic data), vital sign measurements, ECG data, index knee examination, X-Ray and MRI (including but not limited to Minimum Joint Space Width (mJSW) and Incidence of Rapidly Progressive Osteoarthritis (RPOA) Types I and II), and physical examination abnormalities using the Safety

Analysis Set.

### 4.16.1 Plasma Concentration Analyses

Summaries for concentrations will be presented for the Safety Set by LEVI-04 dose using descriptive statistics (number of participants, mean, SD, CV %, median, minimum, and maximum). Concentration values below the limit of quantification (BLQ) will be treated as 0 when calculating summary statistics. Where all values are BLQ, the mean, median, minimum, and maximum concentration will be presented as BLQ and the SD and CV % will be reported as not applicable.

### 4.16.2 Adverse Events

An AE is defined as "any untoward medical occurrence in a participant administered a pharmaceutical product, regardless of causal relationship with this treatment". An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal product, whether considered related to the medicinal product or not. For surgical or diagnostic procedures, the condition/illness leading to such a procedure is considered as the AE rather than the procedure itself.

An event will be considered a treatment-emergent AE (TEAE) if the start or worsening of the event was on or after the administration of IP. Summaries of adverse events will be limited to TEAEs, but listings of all events will be provided. A serious adverse event is an event recorded as 'Serious' on the adverse event eCRF page. Adverse events will be classified by MedDRA SOC and PT.

The overall summary of TEAEs will include frequencies and percentages and will be presented for the following:

- Overview of adverse events;
- All adverse events by SOC and PT;
- Serious adverse events by SOC and PT;
- Adverse events of special interest by SOC and PT;
- Adverse events by SOC, PT, and severity;
- Adverse events related to study treatment by SOC and PT

Relationship to study drug will be classified as "unrelated" or "related". The severity of an adverse event will be rated by the investigator as "mild", "moderate", or "severe". Participants experiencing multiple instances of the same PT are only counted once within each SOC and PT combination. Participants experiencing multiple instances of the same PT will only be counted

once according to the highest severity during the treatment emergent period when summarizing severity.

Serious adverse events, if present, also will be listed using verbatim, SOC, and PT. Should any serious adverse events occur during the study, the serious adverse events are displayed in a table and a narrative for each serious adverse event included in the study report.

### 4.16.3 Vital Signs Measurements

Vital sign observed values and changes and shifts from Baseline will be summarized at each scheduled visit, where applicable. A listing of all clinically significant abnormal results will be presented.

### 4.16.4 Clinical Laboratory Tests

Clinical laboratory parameters, including observed values and changes and shifts from Baseline, will be summarized at each scheduled visit, where applicable. Laboratory parameters may be classified as low, normal, or high (quantitative parameters) or abnormal or normal (qualitative parameters) by the central laboratory according to the reference ranges provided in the laboratory data transfer. Low, high, and abnormal results will be flagged in the listings of individual participant data. Some parameters may not have normal ranges.

### 4.16.5 ECGs

The overall ECG interpretation will be summarized by presenting the number and percentage of participants with “Normal”, “Abnormal, not clinically significant”, and “Abnormal, clinically significant”. Shifts from Baseline in overall ECG interpretation will also be presented. A listing of all clinically significant abnormal overall ECG interpretations will be presented.

Continuous ECG parameter (e.g., QTcF) observed values and changes from Baseline will be summarized per schedule.

### 4.16.6 Physical Examinations

Physical examinations will include, at a minimum, assessments of the cardiovascular, respiratory, gastrointestinal, neurological, and musculoskeletal systems. Any abnormal findings will be recorded as medical history or adverse events depending on the time of assessment.

### 4.16.7 X-Rays

Minimum Joint Space Width as determined by X-Ray will be summarized by observed values and changes from Baseline at Visit 11 The incidence of Rapidly Progressive Osteoarthritis (RPOA) will be summarized.

# 5. DATA HANDLING

## 5.1 Baseline and Study Visits

The last non-missing value, including unscheduled, prior to the first dose of study drug will be used as the Baseline value. Summaries presented by nominal collection day and/or by time point will be based on the schedule of assessments as planned in the protocol and as recorded on the eCRFs.

## 5.2 Pooling of Sites

Sites contributing less than 16 participants to the PP Analysis Set will be pooled. Pooling will begin by sorting the sites based on the number of participants in the PP Analysis set. The site with the lowest number of participants will be pooled with the site with the next lowest number of participants. If necessary, pooling with subsequent sites will continue until the pooled site contains at least 16 participants in the PP Analysis set. Pooling for additional sites will continue until all sites/pooled sites have at least 16 participants in the PP Analysis set. Pooled site will be used for all analysis (e.g., ANCOVA) containing site in the model or site as a stratification factor.

## 5.3 Missing Data

WOMAC pain scores will be imputed as described in Section 4.9 for the primary analysis based on multiple imputations. All other analyses will be based on available data. Handling of plasma values that are below the limit of quantification with respect to the pharmacokinetic analyses are described in Section 4.16.1.

## 5.4 Unscheduled Data

Unscheduled results will only be considered for Baseline determination; otherwise, unscheduled data will not be used in summary tables or included in the analyses. All unscheduled data will be presented in the listings.

## 5.5 Presentation Conventions

For all parameters, where applicable:

- n: no decimal place,

- CV% and CV% geometric mean will be displayed to one decimal place,

Parameters that are reported or derived at a fixed number of decimal points (e.g., vital sign measurements, including body mass index (BMI), clinical laboratory test parameters, and plasma drug concentrations), will be reported as follows:

- Mean, median, and first and third quartiles will be displayed to one more decimal than the maximum number of decimal places reported for the original data,
- SD will be reported to two more decimal places than the maximum number of decimal places reported for the original data,
- Minimum and maximum will be reported to the maximum number of decimal places reported for the original data.

### 5.5.1 Significant Figures

When descriptive statistics are to be presented to a specific number of significant figures, the following conventions will be utilized:

- Results of "0" will be displayed as "0".
- Results that cannot be displayed to the exact specified number of significant figures (in Section 3.5) without the use of scientific notation will be displayed using the maximum number of significant figures that is less than the specified number of significant figures. E.g., "1000" and "1295" cannot be displayed to three significant figures without the use of scientific notation, so "1000" and "1300" would be presented, respectively.
- All other results will be displayed to the specified number of significant figures in Section 3.5. Examples of results displayed to three significant figures are presented below:

| **Original Result** | **Presentation Result (3 Significant Figures)** |
|---|---|
| 12.34 | 12.3 |
| 0.01234 | 0.0123 |
| 0.12 | 0.120 |
| 10 | 10.0 |
| 1234 | 1230 |

# 6. CHANGES FROM THE PROTOCOL

The secondary objective "To evaluate the proportion of responders based on various levels of reduced pain in participants receiving LEVI-04 (multiple doses) compared to placebo" was overly specific with regards to the endpoints associated with it. That is, not all associated endpoints are response rates. However, all the endpoints are measures of pain, so the objective was modified to: "To evaluate the efficacy of LEVI-04 (multiple doses) compared to placebo in reducing pain".

The protocol specifies that descriptive statistics will be provided for physical examinations. Results for individual body systems are not collected. Any abnormalities are recorded as medical history or adverse events.

Analyses associated with exploratory endpoint "to explore associations between imaging biomarkers and pain severity" will not be conducted at this time. These analyses may be generated at a later date.

The protocol specifies that the daily average amount of rescue medication usage and the proportion of days on rescue medication while on treatment will be calculated by dividing by the number of days on study treatment. If participants do not have any missing diary entries and complete the study, then using the number of days on study treatment for the denominator for these parameters will provide reasonable estimates. However, participants may not complete the study or provide responses to the rescue medication usage for all days. For these instances, rescue medication usage and the proportion of days on rescue medication would be underestimated if using the number of days on study treatment as the denominator as it would assume zero usage on those unknown days. In order to provide more accurate estimates for these two parameters, the denominator will only include those study days for which the required usage information is known for the given parameter.

# 7. INTERIM ANALYSIS

An interim analysis (IA) will be conducted on the first 208 randomized participants in the study. The interim analysis will be performed once all 208 participants have completed the Week 5 visit or have discontinued prior to Week 5. Based on the IA's conditional power analysis, the study will continue as planned with a total of 416 participants randomized, or with an increase of up to 208 participants (an additional 50%) up to a total of 624 participants. Conditional power calculations will be performed using two alternative methodologies:

1. Multiple imputations will be utilized to impute missing data. The Week 17 effect size will be determined for each imputation. The conditional power will be calculated based on these multiple imputations interim Week 17 effect sizes. The Week 17 effect size for the

post-interim participants will be assumed to be from the same distribution as its associated Week 17 effect size estimate from the interim participants.

2. Multiple imputations will be utilized to impute missing data. The Week 5 effect size will be determined for each imputation. The conditional power will be calculated based on these multiple imputations interim Week 5 effect sizes. The Week 5 effect size for the post-interim participants will be assumed to be from the same distribution as its associated Week 5 effect size from the interim participants

A single set of one hundred imputations will be determined and utilized for both methods above.

The IA summary will be shared with the IA independent statistician. Below is the mock summary of the IA:

Levicept Ltd.: Protocol LEVI-04_21_02

Table 1

Interim Analysis

Summary of Mean Change from Baseline in WOMAC Pain Subscale Score
and Associated Conditional Power
Intent-to-Treat Analysis Set – Multiple Imputations

| | **0.3 mg/kg (N=xx)** | **1.0 mg/kg (N=xx)** | **2.0 mg/kg (N=xx)** | **Placebo (N=xx)** |
|---|---|---|---|---|
| **Method 1: Change from Baseline to Week 17** | | | | |
| n | xx | xx | xx | xx |
| LS Mean (SE)[a] | xx (xx.x) | xx (xx.x) | xx (xx.x) | xx (xx.x) |
| LS Mean Difference (SE)[a] | xx (xx.x) | xx (xx.x) | xx (xx.x) | |
| **Method 2: Change from Baseline to Week 5** | | | | |
| N | xx | xx | xx | xx |
| LS Mean (SE)[a] | xx (xx.x) | xx (xx.x) | xx (xx.x) | xx (xx.x) |
| LS Mean Difference (SE)[a] | xx (xx.x) | xx (xx.x) | xx (xx.x) | |
| **Conditional Power[b]** | | | | |
| Method 1 | | | | |
| Final planned sample size (N=416) | xx.x | | | |
| 50% increase in final sample size (N=624) | xx.x | | | |
| Method 2 | | | | |
| Final planned sample size (N=416) | xx.x | | | |
| 50% increase in final sample size (N=624) | xx.x | | | |

a: From an ANCOVA model with main effect treatment and covariates study center and baseline score with missing scores imputed using multiple imputations methodology. For method 1, participants that are still in study without week 17 data will be considered as missing at random. For both methods 1 and 2, for participants who discontinued due to lack of efficacy of adverse event, imputation will be done using the placebo as the treatment group.

b: Conditional power is calculated as the probability for at least one of the comparisons against placebo is significant at the Dunnett adjusted type-1 error of 0.xxx per comparison (controlling overall type-1 error at a two-sided 0.05 level), assuming the post-interim effect size comes from the same distribution as its associated interim effect size.

The two methods of estimating the effect size and consequently, the conditional power could lead to different sample size reassessments. The independent statistician will report the most conservative (largest) sample size in the recommendation to the designated personnel at Levicept. The only recommendation communicated to Levicept will be whether the sample size will need to increase or stay the same.

**Conditional Power Calculations**

For each of the methods above, conditional power will be calculated as follows:

1. $Z_{i1}, i = 1,2,3$ will be calculated by transforming the interim p-value of each dose group comparison to placebo using the inverse normal transformation of the ANOVA t-statistics.
2. $\delta_i = Z_{i1}\sqrt{\frac{1}{n_{i1}} + \frac{1}{n_{01}}}, i = 1,2,3$ is the effect size of the three dose group comparisons to placebo from the interim participants, which is assumed to be also the effect size of the post-interim participants.
3. $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$ is the correlation matrix of the three t-statistics from the comparisons against placebo, where:

$$\rho_{ij} = \lambda_i \lambda_j, \lambda_i = \sqrt{\frac{n_{i1}}{n_{i1}+n_{01}}}$$

4. Let $\mathbf{\Lambda}$ be a lower triangular matrix from the Cholesky decomposition of $\mathbf{\Sigma}$.
   Let $X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ be a three-dimensional vector of independent standard normal random variables and let $\Delta = \begin{pmatrix} \frac{\delta_1}{\sqrt{\frac{1}{n12} + \frac{1}{n02}}} \\ \frac{\delta_2}{\sqrt{\frac{1}{n22} + \frac{1}{n02}}} \\ \frac{\delta_3}{\left(\sqrt{\frac{1}{n32} + \frac{1}{n02}}\right)} \end{pmatrix}$. Then $\mathbf{\Lambda}'X + \Delta = Z_2 = \begin{pmatrix} Z_{1,2} \\ Z_{2,2} \\ Z_{3,2} \end{pmatrix}$ is a three-dimensional vector of normal random variables, with a correlation matrix $\mathbf{\Sigma}$, representing the test statistics of post-interim participants with the same effect size as the interim participants.

5. Generate 5,000 random vectors of $Z_2$.

For each random vector $Z_2$, calculate:

$Z_i = \sqrt{\tau_2}\, Z_{i,1} + \sqrt{\tau_2}\, Z_{i,2}$, $i = 1,2,3$, which is the (combination) Z-statistics for the three

pairwise comparisons at the end of the study. Note that in each of the 5,000 samples, $Z_{i,1}, i = 1,2,3$, is fixed and equals to the interim $Z_{i,1}$ statistic.

In each of the 5,000 simulations, if:

$$(\text{abs}(Z_1) > C) \text{ or } (\text{abs}(Z_2) > C) \text{ or } (\text{abs}(Z_3) > C) \text{ then reject } H_0,$$

where $C$ is the critical value from the two-sided Dunnett test for 3 pairwise comparisons calculated using PROBMC of SAS:

$C$ = PROBMC("DUNNETT2", ., 0.95, . ,3,lam1,lam2,lam3),
where lami is as defined in Section 4.11. For equal sample sizes $C \approx 2.349$.

6. The proportion of times that $H_0$ is rejected is the estimated conditional power.

Note that the application of Dunnett's test to a combination test relies on large sample property of the t-distribution namely the asymptotic distribution is normal. This allows to approximate the correlation matrix among the combination test statistics using the correlation matrix associated with the t-statistics.