HELIXMITH

Statistical Analysis Plan

Protocol Title:	**An Adaptive Phase 3, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study to Assess the Safety and Efficacy of Engensis in Participants with Painful Diabetic Peripheral Neuropathy**
Protocol Number:	VMDN-003-2
Protocol Version:	6.0
Approval Date:	13 Dec 2023
Compound:	Engensis
Short Title:	Phase 3 Trial to Assess Safety and Efficacy of Engensis in Painful Diabetic Peripheral Neuropathy
Acronym:	REGAiN-1A
Sponsor Name:	**Helixmith Co., Ltd.**

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

DISCLOSURE STATEMENT

This study will be conducted in compliance with the protocol, US Code of Federal Regulations applicable to clinical studies, principles of ICH Good Clinical Practice, the Declaration of Helsinki, and all applicable regulatory requirements. This protocol is the confidential information of Helixmith and is intended solely for the guidance of the clinical investigation. This protocol may not be disclosed to parties not associated with the clinical investigation or used for any purpose without the prior written consent of Helixmith.

[redacted]

[redacted]

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

AE	adverse event
AESI	adverse event of special interest
AIC	Akaike information criterion
ADPS	Average Daily Pain Score
BPI-DPN	Brief Pain Inventory for Diabetic Peripheral Neuropathy
CDRC	Clinical Data Review Committee
CRO	Clinical Research Organization
CS	clinically significant
CSR	clinical study report
DPN	diabetic peripheral neuropathy
DSMB	Data Safety Monitoring Board
ECG	electrocardiogram
HEENT	head, eyes, ears, nose, and throat
HGF	hepatocyte growth factor
IDMC	Independent Data Monitoring Committee
IM	intramuscular
IRB	Institutional Review Board
ISR	injection site reaction
ITT	intent-to-treat
mITT	modified intent-to-treat
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Affairs
MOTH	Mean of the Other Group
MNSI	Michigan Neuropathy Screening Instrument
NCS	not clinically significant
PGIC	Patients' Global Impression of Change
PP	per protocol
SOC	System Organ Class
TEAE	treatment-emergent adverse event
WHODrug	World Health Organization Drug Dictionary

DEFINITIONS

Adverse Event	An adverse event (AE) is the development of an untoward medical occurrence or the deterioration of a pre-existing medical condition following or during exposure to an investigational product, whether or not it is considered causally related to the product. Changes in a chronic condition or disease that are consistent with natural disease progression are NOT considered AEs.
Serious AE	Any untoward medical occurrence which results in death; is a life-threatening experience; requires hospitalization (admission to hospital with a stay > 24 hours) or prolongation of an existing hospitalization which is not specifically required by the protocol or is elective; results in permanent impairment of a body function or permanent damage to a body structure; or requires medical or surgical intervention to preclude permanent impairment of a body function or permanent damage to a body structure.
Treatment Emergent Adverse Events and Treatment Emergent Serious Adverse Events	Adverse events that occur after dosing and pre-existing medical conditions that worsen following exposure to an investigational product.

1. INTRODUCTION

This document contains a detailed description of the statistical methods[1] to be implemented during the analyses of data collected within the scope of Helixmith Co., Ltd. Protocol VMDN-003-2 (An Adaptive Phase 3, Double-Blind, Randomized, Placebo-controlled, Multicenter Study to Assess the Safety and Efficacy of Engensis in Participants with Painful Diabetic Peripheral Neuropathy). The purpose of this plan is to provide specific guidelines from which the analyses will proceed. Any deviations from these guidelines will be documented in the clinical study report (CSR).

2. OBJECTIVES AND ENDPOINTS

Objectives	Endpoints
Primary	
• To evaluate the efficacy of IM administration of Engensis on pain in participants with painful DPN in the feet and lower legs, as compared to Placebo	• Change in the means of the Average Daily Pain Scores (ADPSs) from the Brief Pain Inventory for Diabetic Peripheral Neuropathy (BPI-DPN) from the 7 days prior to the Day 0 Visit to the 7 days prior to the Day 180 Visit for Engensis compared to Placebo in the intent-to-treat (ITT) population
Secondary Efficacy	
• To evaluate the efficacy of IM administration of Engensis on the worst pain in Participants with painful DPN in the feet and lower legs as compared to Placebo	• Change in the means of the Worst Pain Scores from the BPI-DPN from the 7 days prior to the Day 0 Visit to the 7 days prior to the Day 180 Visit for Engensis compared to Placebo
• To evaluate the efficacy of IM administration of Engensis on reducing pain in Participants with painful DPN in the feet and lower legs as compared to Placebo	• Proportion of Responders (≥ 50% reduction in the means of ADPSs from the BPI-DPN) from the 7 days prior to the Day 0 Visit to the 7 days prior to the Day 180 Visit for Engensis compared to Placebo

Objectives	Endpoints
Secondary Safety	
• To evaluate the safety of IM administration of Engensis in Participants with painful DPN in the feet and lower legs as compared to Placebo	• Incidence of treatment-emergent adverse events (TEAEs) and treatment-emergent serious adverse events (TESAEs) for Engensis compared to Placebo • Incidence of injection site reactions for Engensis compared to Placebo • Incidence of clinically significant laboratory values for Engensis compared to Placebo
• To evaluate the possibility of cellular and/or humoral responses to Engensis compared to Placebo in Participants with painful DPN in the feet and lower legs	• Change from Baseline in the cytokine profile through post-dose on Day 104 for Engensis compared to Placebo • Presence of anti-hepatocyte growth factor (HGF) antibodies following Engensis administration compared to Placebo
Exploratory Efficacy	
• To evaluate the efficacy of IM administration of Engensis on Quality of Life and Patient Reported Outcomes in Participants with painful DPN in the feet and lower legs as compared to Placebo	• Patient Global Impression of Change (PGIC) on Day 90 and on Day 180 for Engensis compared to Placebo
• To evaluate the durability of the analgesic response to IM administration of Engensis in Participants with painful DPN in the feet and lower legs as compared to Placebo	• Proportion of Responders (≥ 50% reduction in the means of the ADPSs) from the 7 days prior to the Day 180 Visit who were Responders on Day 104 (final injection visit) for Engensis compared to Placebo
• To determine whether IM administration of Engensis has a positive effect (reverses/reduces pain, improves neurological function and quality of life) on painful DPN as compared to Placebo	• Change in the Bedside Sensory Testing (BST) from Baseline to Day 90 and to Day 180 for Engensis compared to Placebo • Proportion of Responders (≥20, 30, 40, 60, and 70% reduction in the means of the ADPSs from the BPI-DPN) from the 7 days prior to the Day 0 Visit to the 7 days prior to the Day 90 Visit and the 7 days prior to the Day 180 Visit for Engensis compared to Placebo

Objectives	Endpoints
	• Change in the means of the ADPSs from the BPI-DPN from the 7 days prior to the Day 0 Visit to the 7 days prior to the Day 90 Visit for Engensis compared to Placebo • Changes in the severity scores (Average Pain, Worst Pain, Least Pain, and Pain Right Now) from the full BPI-DPN from the 7 days prior to the Day 0 Visit to the 7 days prior to the Day 90 Visit and the 7 days prior to the Day 180 Visit for Engensis compared to Placebo • Changes in Michigan Neuropathy Screening Instrument (MNSI) assessments from Baseline to Day 90 and to Day 180 for Engensis compared to Placebo • Change in the 36-Item Short Form Health Survey (SF-36) from Baseline to Day 90 and to Day 180 for Engensis compared to Placebo • Change in the EuroQol Health Utilities Index (EQ-5D) from Baseline to Day 90 and to Day 180 for Engensis compared to Placebo

3. STUDY DESIGN

This is a Phase 3, double-blind, randomized, placebo-controlled, multicenter, 6-month study designed to assess the safety and efficacy of repeat bilateral intramuscular (IM) injections of Engensis in the calves of participants with painful diabetic peripheral neuropathy (DPN). The study hypothesis is that Engensis will significantly reduce participant-reported pain.

First Dose: Participants will receive Engensis or Placebo by IM injections in both calves on Day 0 and Day 14 as shown in Table 1.

4. STUDY ENDPOINTS

4.1. Primary Efficacy Endpoint

[redacted]

$at\ D_i$

[redacted]

4.2. Secondary Efficacy Endpoints

[redacted]

- Proportion of Responders (≥ 20, 30, 40, 60 and 70% reduction in the means of the ADPSs from the BPI-DPN) from the 7 days prior to the Day 0 Visit to the 7 days prior to the Day 90 Visit and the 7 days prior to the Day 180 Visit for Engensis compared to Placebo
- Changes in the means of the Average Daily Pain Scores (ADPSs) from the BPI-DPN from the 7 days prior to the Day 0 Visit to the 7 days prior to the Day 90 Visit for Engensis compared to Placebo
  - This will be calculated the same way described in Section 4.1.
- Changes in the severity scores (Average Pain, Worst Pain, Least Pain and Pain Right Now) from the full BPI-DPN from the 7 days prior to the Day 0 Visit to the 7 days prior to the Day 90 Visit and to the 7 days prior to the Day 180 Visit for Engensis compared to Placebo
  - The severity score consists of four scores (Average Pain, Worst Pain, Least Pain and Pain Right Now), and the overall average of the scores can be calculated. The change in each score will be calculated and analyzed separately.
- Changes in Michigan Neuropathy Screening Instrument (MNSI) from Baseline to Day 90 and to Day 180 for Engensis compared to Placebo
  - The MNSI comprises a Participant questionnaire (15 questions) with two domains: Physical Assessment and History. The change will be calculated for each domain separately.
- Changes in Quality of Life in the 36-item Short Form Health Survey (SF-36) from Baseline to Day 90 and to Day 180 for Engensis compared to Placebo
  - The SF-36 survey has 8 domains and 2 component scores; the 8 domains include Physical Functioning, Social Functioning, Role–Physical, Role–Emotional, Bodily Pain, General Health, Vitality, and Mental Health, and the 2 component scores include a Physical Component score and Mental Component score. The change will be calculated for each domain and component score separately.
- Changes in Quality of Life in the EuroQol Health Utilities Index (EQ-5D) from Baseline to Day 90 and Day 180 for Engensis compared to Placebo
  - EQ-5D has 2 main parts. Part 1 is the descriptive system and Part 2 is the visual analog scale of self-rated health. Part 1 consists of 5 questions assessing: Mobility, Self-Care, Usual Activities, Anxiety/Depression and Pain/Discomfort. The overall health state value index combining the 5 questions of Part 1 in EQ-5D will be calculated. The overall scores and change in each Part will be calculated and summarized.

4.4. Safety Outcomes

Safety analyses in this study will evaluate the safety profile of Engensis as compared with Placebo. No formal statistical testing will be conducted for the safety analyses. All participants in the safety subset will be included in these analyses. Participants will be grouped by treatment received. All summaries will be derived based on available data. No imputation will be performed for missing values. All safety analyses will be made on the Safety Population.

- Change from Baseline (Day 0) in the cytokine profile through post-dose to Day 104 for Engensis compared to Placebo
- Presence of anti-hepatocyte growth factor (HGF) antibodies following Engensis administration compared to Placebo
  - HGF antibodies will be collected on Day 0, 60, 90, 150 and 180, and the presence will be summarized according to each time point.
- Adverse events
- Injection site reaction assessment
- Vital signs
  - Blood pressure
  - Weight
  - Heart rate
  - Respiration rate
  - Oxygen saturation
  - Temperature
- HbA1c
- Laboratory values (Serum chemistry and hematology)
- Retinal fundoscopy

4.5. Other Clinical Parameters

- Total acetaminophen (rescue medication) used during the study

4.6. Planned Covariates

As defined in the analysis of the primary and secondary efficacy endpoints, the following can be included in the model as covariates after testing their significance:

- Baseline pain score
- Baseline HbA1c ($<$ and $\geq$ median)
- Gender (male and female)
- Age ($<$65 years and $\geq$ 65 years)

5. SAMPLE SIZE CALCULATION

The sample size for this clinical investigation was based on the primary efficacy endpoint. Based on the VMDN-002 and VMDN-003b findings, the standard deviation of the pain score change from baseline is assumed to be 2.18. Table 3 provides the summary statistics for the corresponding efficacy endpoints based on change in the Average Daily Pain Score for the low- dose and placebo groups from the ITT population and efficacy population in the VMDN-002 trial as well as the mean and standard deviation of change from baseline in the ITT population in the VMDN-003b trial.

Table 3: Mean and Standard Deviation of Pain Score Changes from Baseline in VMDN-002 and VMDN-003b Trials

Trial (Population)	Placebo	Engensis	Mean Difference	Placebo STD	VM202 STD	Pooled STD
VMDN-002 (ITT population)	-1.63	-2.58	0.95	1.76	2.19	2.08
VMDN-002 (Efficacy population)	-1.59	-2.78	1.19	1.89	2.24	2.18
VMDN-003b (ITT population)	-1.49	-2.52	1.03	1.77	2.23	2.13

Abbreviations: ITT, intent-to-treat; STD, standard deviation.

The estimates presented in Table 4 consider a range of scenarios. Following a careful review, the maximum sample size is 250 participants. The target sample size is 152 participants and is based on the following assumptions:

- Mean difference between active and control of 1 point
- Common standard deviation of 2.18 units
- Randomization ratio: 1:1
- 2-sided test with a type 1 error rate of 5%

Table 4: Sample Size Scenarios

Scenario	Primary Endpoint Mean Difference (Engensis minus Placebo)	Nominal Power	Actual Power	Total Sample Size
1	-1	80%	0.802	152
2	-1	85%	0.853	174
3	-1	90%	0.900	202
4	-1	95%	0.951	250
5	-1.2	80%	0.802	106
6	-1.2	85%	0.854	122
7	-1.2	90%	0.903	142
8	-1.2	95%	0.951	174

When 50% of the minimum target sample size (76 subjects) has been enrolled and completed or withdrawn prematurely, a re-estimation of the sample size based on the conditional power will be conducted based on the method described by Mehta and Pocock.[2]

The maximum sample size of 250 will provide 95% power for the treatment difference of 1.0 point when other assumptions are held the same. Simultaneously, the maximum sample size of 250 is considered an appropriate sample size to claim statistical significance for the clinically meaningful difference with an effect size of $\geq$0.3.

Independently of the interim assessment, an effect size of 0.3670 (Mean Difference/SD=0.8/2.18) translates into an overall power of 82.4%. While this is on the cusp of what is considered a small to moderate effect, this can also be described as an observed difference (active minus control) of −0.8 units with 250 subjects, based on the original planning estimates.

An effect size of 0.3440 (0.75/2.18) would still be on the cusp of what is considered a small to moderate effect; this translates into an observed difference (active minus control) of -0.75 units with 250 subjects, based on the original planning estimates.

An effect size of 0.3211 (0.7/2.18) would still be considered a small yet meaningful clinical effect; this translates into an observed difference (active minus control) of -0.7 units with 250 subjects, based on the original planning estimates.

Clinically, a small effect size of 0.3 would still offer benefit in the form of pain relief to the subjects; this effect size needs to be considered relative to the risk of taking the drug. Therefore, maximum sample size of 250 subjects would not allow for a statistical result that is significant ($p<0.05$) with an effect size of <0.3; this would not be considered clinically meaningful.

6. ANALYSIS POPULATIONS

Reference Section 9.2 Protocol VMDN-003-2 for the complete definitions of the populations for analysis. The sections below align with the definitions in the protocol.

6.1. Intent-to-Treat (ITT) Population

The intent-to-treat (ITT) population includes all participants who are randomized. The ITT population will be the primary population used for the efficacy analyses.

All baseline characteristics will be summarized based on the ITT population. The primary analyses of the primary and secondary efficacy endpoints will be based on the ITT population.

Participants in the ITT population will be analyzed according to the randomized treatment assignment, regardless of the actual treatment administered.

6.2. Safety Population

The safety population includes all participants who are randomized and receive at least one study drug injection. The safety population will be used for the safety analyses.

Participants will be analyzed by the actual treatment administered, not according to their randomization assignment. Participants treated with any Engensis dose will be grouped in the Engensis group; participants never treated with any Engensis will be grouped in the Placebo group.

6.3. Modified Intent-to-Treat (mITT) Population

The modified ITT (mITT) population includes all participants that meet the following:

- Underwent (any) injections
- Correctly completed for at least 5 out of 7 days the BPI-DPN eDiary at baseline and the 6-month follow-up

Participants will be grouped based on the randomly assigned treatments, not the actual treatment received. The mITT population will be used in the sensitivity analyses for the primary and secondary efficacy endpoints.

6.4. Per Protocol (PP) Populations

The Per Protocol population is a subset of mITT. It includes all mITT participants who meet all of the following criteria:

- The participant meets major protocol eligibility criteria determined by Clinical Data Review Committee (CDRC) prior to database lock and breaking the randomization codes.
- The participant received all injections based on the randomized treatments and had a Day 180 Visit, including a 7-day ADPS.

The CDRC, which is masked to the treatment information of each participant, may determine additional criteria, if any, before unblinding the randomization code.

The PP population will be used in the sensitivity analyses for the primary and secondary efficacy endpoints.

6.5. Subgroup Analysis Subsets

The subgroup analyses will be exploratory in nature and will be conducted in the ITT population. The primary and secondary efficacy endpoints will be evaluated based on the categories of the covariates described in Section 4.6 as well as the stratification factor. These subgroups that will be re-examined and may be recategorized or eliminated due to small sample size (if there are <10% of participants within each subgroup) before unblinding for analysis. For example, if <10% of overall participants are ≥65 years, then analyses for this subgroup will not be performed. The treatment by subgroup interaction will be examined and tested as described in Section 8.4.5.

7. GENERAL PRINCIPLES OF DATA HANDLING

Data screening will be conducted in a blinded fashion periodically during the conduct of the study. The objective of the data screening is to assess the quantity, quality, and statistical characteristics of the data relative to the requirements of the planned analyses. The designated Contract Research Organization (CRO) will be responsible for data cleaning and dictionary coding of AEs, medical history, and medications. Any questionable values or situations will be reported to the CDRC for final review and confirmation.

7.1. Baseline Definition

Unless specified otherwise, the baseline value for each variable is the value recorded at the last visit on or before start of dosing.

For the pain scores from the BPI-DPN eDiary, the baseline value is the Average Daily Pain Score prior to the Day 0 visit.

Note that, for eligibility, the mean of Average Daily Pain Scores of the BPI-DPN eDiary should be ≥ 4 during the 7 days prior to Day 0. For the Average Daily Pain Score calculation, at least five (5) days need to have the available scores.

7.2. Visit Windows

Data at each scheduled follow-up visit will be analyzed according to the nominal visit identified on the data record.

In case of multiple different visits with the same nominal visit designation, the visit with the visit date closest to the target days of each protocol-specified visit schedule will be used for the efficacy analyses. For visits with the same distance to the target days, the later nominal visit record will be used. Data from all visits will be provided in the data listings.

If a subject fails to complete a study visit but provides necessary eDiary data during the time period expected for that visit (i.e. based on study day relative to the date of the first dose in VMDN-003-2), those eDiary data will be used to derive applicable endpoints. Day 90 endpoint would be derived by the eDiaries recorded 7 days up to 90 days from the date of first dose. Also, Day 104 endpoint would be derived by the eDiaries recorded 7 days up to 104 days from the date of first dose. Similarly, Day 180 endpoint would be derived by the eDiaries recorded 7 days up to 180 days from the date of first dose. Otherwise, if subject completes study visit, then date of visit will be used to derive applicable endpoints.

7.3. Unmasking of Randomization Codes

Following database lock, the randomization code will be unmasked to the project team. The randomization code will be unmasked to the Data Safety Monitoring Board (DSMB; Section 7.5) members and the CRO that will prepare the unmasking summary tables for the DSMB meetings. However, to prevent bias, the unmasking detailed data summaries will not be shared with the sponsor management team, the CDRC, or the team that is monitoring the clinical data collection. The participant and study personnel, including core lab, principal investigator, co-investigators, study coordinators, study monitors and study director will remain blinded to individual data and group results until all data has been entered into the database and the database is locked.

7.4. Multiplicity Adjustment

To control the family-wise type 1 error rate at 0.05 for multiplicity across the primary and secondary

endpoints, a hierarchical testing strategy[3] will be used. The hierarchical testing strategy starts with the primary endpoint followed by sequential testing of the secondary endpoints. The statistical significance will be evaluated in the following order: If a statistical significance at the 0.05 level (2-sided) is shown for the first ranking order, then the next endpoint in the immediate subsequent order will be evaluated; evaluation of subsequent endpoints will continue in the same manner. If no statistical significance is shown at 0.05 level at any endpoint, then the endpoint and all subsequent endpoints will not be considered statistically significant, regardless of their p-values.

The hierarchical testing order follows:

1. Change in the means of the Worst Pain Scores from the BPI-DPN from the 7 days prior to the Day 0 Visit to the 7 days prior to the Day 180 Visit for Engensis compared to Placebo
2. Proportion of Responders (≥ 50% reduction in the means of ADPSs from the BPI-DPN) from the 7 days prior to the Day 0 Visit to the 7 days prior to the Day 180 Visit for Engensis compared to Placebo

The p-values for other secondary and exploratory endpoints will be used for informative or supportive purposes only. Therefore, there will be no further adjustment for multiple tests.

7.5. Data Safety Monitoring Board

In addition to the interim assessment of the sample size by the Independent Data Monitoring Committee (IDMC), a separate independent DSMB will periodically review a limited set of unblinded tables and/or listings, including all reported AEs. The objectives of the DSMB meetings are to review the safety outcomes of the study and provide guidance to the study sponsor regarding the safety of the Engensis. The data analyses for the DSMB meetings will be directly provided to the DSMB members and no data will be released to the study sponsor and blinded designees. There will be no adjustment for multiple testing due to the DSMB data review. The DSMB may be asked to review and provide guidance regarding protocol deviations that may affect the determination of the PP populations. Further details of DSMB responsibilities are included in the DSMB Charter.

7.6. Handling of Missing and Incomplete Data

Participants may have missing specific data points for a variety of reasons. In general, data may be missing due to a participant's early withdrawal from study, a missed visit, or a clinical parameter not measured at a particular point in time. The general procedures outlined below describe how missing data will be addressed in the analyses.

7.6.1. Missing Pain Scores for the Average Scores from BPI-DPN

The means of the Average Daily Pain Scores (Question 5 from the BPI-DPN) for a visit will be considered as missing if fewer than 5 of 7 days of BPI-DPN eDiary entries are provided. Any missing mean pain score of ADPS will be imputed using the methods described below.

- Multiple Imputation (MI) method:

  Step 1: Intermittent missing values will first be imputed by treatment group using the Markov Chain Monte Carlo (MCMC) method, resulting in data with a monotone pattern.

  Step 2: Multiple imputation by treatment group using linear mixed model regression with factors of baseline pain score, baseline HbA1c, gender, and age will be applied to the data

obtained from the MCMC step. The imputed score will be rounded to the first decimal point.

The entire procedure will be repeated 10 times, and 10 complete datasets will be used for the analysis.

- Mean of the other group (MOTH):[4]

1. For a participant with missing average pain score/the worst pain score in the full BPI-DPN instrument from the 7 days prior to the Day 0 to Day 90, and Day 180 visits, identify the participant's following baseline characteristics:
   - Study treatment group
   - Baseline Average Daily Pain Score (<median or ≥median)
   - HbA1c (<median or ≥median)
   - Gender (male or female)
   - Age (<65 years and ≥65 years)
2. The missing mean pain score from the 7 days prior to the Day 0, Day 90, and Day 180 visits will be imputed by using the mean pain score obtained at the same time point of those participants in the other treatment group who match the participant's baseline characteristics. For example, for the missing pain scores of the Engensis participants, the mean pain scores of the placebo participants within the same covariate groups will be used to impute the missing pain scores.
3. The baseline characteristics will be re-examined for appropriateness and may be re-categorized (due to small sample size) before unblinding the study.
4. The imputed score will be rounded to the first decimal point. The imputed scores will be included in the continuous Repeated Measures Model (Section 8.4.1) analysis.

7.6.2. Addressing Missing Data for the Secondary and Exploratory Efficacy

Missing individual item scores of the full BPI-DPN will be imputed for the calculation of the average score within a domain using the average score across the non-missing items within the domain at a subject's visit provided that the proportion of missing items scores within that domain is less than 25%; otherwise, the domain score at that visit is missing.

The same method as the full BPI-DPN will be used when addressing the missing data for SF-36 and EQ-5D.

Individual item scores on the MNSI (both physical assessment and history) questionnaire will not be imputed. The total MNSI scores will be imputed using a "worst-case" imputation approach. Missing individual items will be imputed as the category corresponding to the largest number of points in the total score calculation. This equates to a response of "Yes" to questions 1-3, 5-6, 8-9, 11-12, and 14-15 and "No" to questions 7 and 13 on the history assessment. On the physical assessment, it equates to a response of "No" on the "Appearance of Feet" question, a response of "Present" on the "Ulceration" question, and a response of "Absent" on the "Ankle Reflexes", "Vibration perception at great toe", and "Monofilament" questions.

Missing PGIC and BST scores and the use of rescue medication will not be imputed.

Missing values from the cytokine profiles and anti-hepatocyte growth factor (HGF) antibody results will not be imputed as they are a part of the safety assessment.

7.6.3. Missing Dates

If a start or stop date for an adverse event or a concomitant medication use is completely missing, it will not be imputed. If it is partially missing, imputed dates in accordance with Table 5 will be used to derive the duration of the adverse event or the duration of the medication use. Missing years will not be estimated under any conditions. Missing dates of medical history will not be imputed.

Table 5: Imputation Rules for Partial Adverse Event or Concomitant Medication Start and Stop Dates

	Missing	Imputation	Exception
Start Date	Day	01	Default to Study Day 0 (day of first injection procedure) if an event starts in the same year and month as Study Day 0
	Day/Month	01JAN	Default to Study Day 0 if an event starts in the same year as Day 0
Start Date			If the start date is completely missing and stop date is either after the date of the first dose of study drug or completely missing, then the start date will be estimated to be equal to the date of the first dose of study drug. Otherwise, the start date will be estimated to be the first day of the same year as the stop date.
Stop Date	Day	Last day of the month	Default to the End of Study Date if the imputed event stop date is after the End of Study Date or before start day of the event
	Day/Month	31DEC	

8. STATISTICAL METHODS

8.1. General Principles of Data Analyses

The primary analysis for this study will be performed and summarized after all randomized participants have had an opportunity to complete their 6-month follow-up visit.

The Safety Analysis will be based on the Safety population. The primary analyses of the efficacy endpoints will be based on the ITT population. Additional sensitivity analyses for the primary and secondary efficacy endpoints will be performed to further assess the effects of the treatment (Section 8.4.2 and 8.4.4).

The statistical analyses will be reported using summary tables, figures and listings. Continuous variables will be summarized with means, standard deviations, medians, minimums, maximums, 25th percentiles, 75th percentiles, and number of non-missing observations for each treatment group.

Categorical variables will be summarized by counts and by the percentage of participants in corresponding categories.

All inferential statistical analyses will be performed with a two-sided confidence level of 95% or a two-sided significance level of 0.05 unless otherwise noted.

All analyses and tabulations will be performed using SAS Version 9.4 or higher on a Server platform.

8.2. Participant Enrollment and Disposition

The reasons for participant enrolled but not randomized (including screen failures) will be summarized by the specific inclusion/exclusion not met for screen failures and any other reasons provided. Participant disposition will be summarized for all randomized participants. The summary including the number and percentage (based on total number of participants randomized) of participants in each of the following categories will be prepared:

- Available at each of the protocol-specified visits based on the ITT population
- Completing 6-month blinded assessment based on the ITT population
- Early Termination based on the ITT population
- Safety population, ITT, mITT, and PP populations

Major protocol deviations for participants not in the PP populations will be listed. Major protocol deviations will be summarized for the ITT population.

8.3. Demographics and Baseline Characteristics

The following outcomes will be summarized by the standard methods for continuous and categorical variables described in Section 8.1.

The demographics include the following parameters:

- Age at informed consent
- Sex
- Race
- Ethnicity

The baseline characteristics include the following:

- Baseline Average Daily Pain Score from the 7 days prior to the Day 0 Visit obtained from the BPI-DPN eDiary
- Vital signs: blood pressure, weight, BMI, heart rate, respiration rate, temperature
- Diabetes type
- 12-lead electrocardiogram (ECG): Normal, Abnormal Not Clinically Significant (NCS), Abnormal Clinically Significant (CS)

These parameters will be summarized by treatment group for the ITT and included in data listings. The demographic and selected baseline characteristics (diabetes type and specific medical history items of interest) will also be summarized for each level of the stratification variable.

8.3.1. Prior and Concomitant Medications

Prior medications are those medications taken within 30 days of the first injection of study drug. Concomitant medications are those medications taken after the initial dose of study drug. A medication with a missing start date and a stop date that is either missing or on or after the treatment start date will be considered as concomitant. All prior and concomitant medications will be assigned WHO drug names using the WHO Drug Dictionary (March 2018 version or later). Prior and concomitant medications will be summarized separately for each treatment group by preferred names. These summaries will present the number and percentage of participants using each medication.

8.3.2. Medical History

Abnormalities in participants' medical and surgical histories will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The number and percentage of participants with abnormalities in medical and surgical histories in each system organ class (SOC) and preferred term will be summarized by treatment group for the ITT population.

In addition, medical history of interest will be summarized separately. Particular medical histories of interest, based upon MedDRA SOC and preferred terms, will be determined by the CDRC in a blinded fashion prior to analyses.

8.4. Efficacy Endpoints Analyses

[redacted]

[redacted]

[redacted]

[redacted]

8.5. Safety Analyses

Safety analyses in this study will evaluate the safety profile of Engensis as compared with control. No formal statistical testing will be conducted for the safety analyses. The following sections summarize the descriptive analysis methods for these safety endpoints. All participants in the Safety population will be included in these analyses. Participants will be grouped by treatment administered. All summaries will be derived based on available data. No imputation will be performed for missing values.

8.5.1. Study Drug Exposure

Study drug exposure (number of injections, total volume and total milligrams administered per calf) will be summarized by treatment group by mg of Engensis for Day 0, Day 14, Day 90, Day 104, and overall using descriptive statistics for continuous variables.

8.5.2. Injection Site Reaction Assessments

The number and percentage of participants with an injection site AE will be summarized descriptively overall and by type (injection site reaction, ulceration, allergic reaction / hypersensitivity) by treatment group and study visit. The number and percentage of participants with a given type of injection site AE will be summarized by grade and treatment group for the pre- and post-injection assessments on Days 0, 14, 90, and 104; only participants who receive an injection at a given one of these visits will be counted in the post-injection assessment results for that visit. Participants without an injection site AE of a particular type will be assigned a grade of 0 for these summaries.

8.5.3. Adverse Events

All adverse event summaries will be restricted to Treatment Emergent Adverse Events (TEAE), which are defined as AEs that occur after dosing and pre-existing medical conditions that worsen following exposure to an investigational product. An AE with a missing start date and a stop date that is either missing or on or after the treatment start date will be considered as a TEAE. For summary purposes, verbatim terms reported by the study centers will be mapped to MedDRA (v21.0 or later) system organ classes (SOC) and preferred terms by the CRO and approved by the CDRC. It should be noted that only AEs that occurred after the first injection will be collected during the study.

The adverse event listings will be displayed by treatment group. The number of participants experiencing a particular event, the percentage of participants experiencing the event, and the total number of events will be presented. The following summaries will be created:

- TEAE by SOC and preferred term
- TEAE by SOC, preferred term and protocol version
- TEAE by SOC, preferred term and maximum severity. At the across-SOC and preferred term levels of participant summarization, a participant is classified according to the highest severity if the participant reported one or more events; severity within an SOC is not summarized. AEs with missing severity will be considered severe for this summary.

- TEAE by SOC, preferred term and closest relationship to study treatment (Related/Not Related). At each level of participant summarization, a participant is classified according to the closest relationship if the participant reported one or more events. AEs with a missing relationship will be considered related for this summary; events classified as "possibly," "probably," or "definitely" will be considered "related."
- Serious TEAEs by SOC and preferred term
- TEAEs leading to study discontinuation by SOC and preferred term
- Adverse events of special interest (AESI) by preferred term.
- Injection Site Reactions (ISRs) included under AESI will include designations of relatedness and severity

AESI summary tables will also be presented separately for each level of the randomization stratification factor.

8.5.4. Cytokine Profile

The absolute values and the change from baseline value will be summarized using descriptive statistics for each treatment group.

8.5.5. Anti-Hepatocyte Growth Factor (HGF) Antibodies

The presence of anti-HGF antibodies will be summarized by its frequency and percentage for each treatment group.

8.5.6. Vital Signs

Vital signs and change from baseline will be summarized descriptively at each visit by treatment group. Additionally, the number of subjects with the following vital signs of special interest will be summarized descriptively by treatment group:

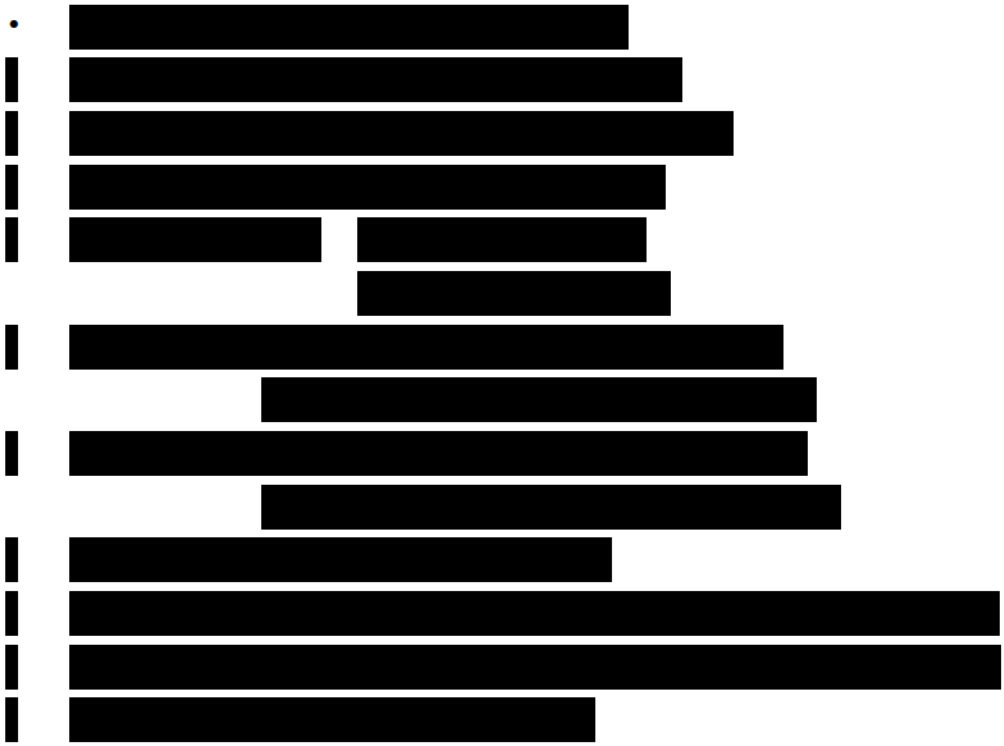
- Systolic BP levels ≥180 mmHg
- Diastolic BP levels ≥100 mmHg
- Heart rate ≥180 beats/min
- Heart rate ≤60 beats/min

8.5.7. HbA1c, Serum Chemistry, and Hematology

Central laboratory results will be summarized using International System (SI) of units. The laboratory results will be summarized by treatment and visit. The parameters within each panel are outlined in Section 8.4.6.1 of the trial protocol, and lab panels for hematology and serum chemistry will be included.

Shift tables (i.e., low, normal, high at baseline versus low, normal, high at follow-up in a 3-by-3 contingency table) will be provided to assess changes in laboratory values from baseline to follow-up result at each scheduled follow-up visit. The counts and percentage of participants with each of the 9 possible "shift" outcomes will be calculated by treatment group. Individual laboratory data from scheduled and unscheduled visits will be listed.

Additionally, the number of subjects with the following lab parameters of special interest will be summarized descriptively by treatment group:



8.5.8. Retinal Fundoscopy

Retinal fundoscopy findings in each eye (presence or absence of proliferative retinopathy, other finding) at screening (Baseline) and the Day 180 visit and any changes from the baseline at the follow-up visits will be summarized descriptively by treatment group.

8.5.9. Electrocardiogram (ECG)

The collected ECG values will be listed.

8.6. Other Clinical Parameters

8.6.1. Total Acetaminophen (Rescue Medication) Dose

The number and percentage of participants taking acetaminophen between baseline and Day 90, Day 90 and Day 180 and baseline and Day 180 will be calculated by the treatment group. The number and percentage of participants taking acetaminophen between baseline and Day 90, Day 90 and Day 180 and baseline and Day 180 will be summarized by the treatment group. The mean, standard deviation, minimum, maximum, median, and 1st and 3rd quartiles of days to the first start date of the acetaminophen will be summarized by the treatment group using observed data; participants not taking acetaminophen will not be included in these descriptive statistics. The total dose of acetaminophen of each participant between baseline and Day 90, Day 90 and Day 180 and baseline and Day 180 will be summarized by the treatment group. These analyses will be based on the ITT and PP Dosing population.

8.7. Interim Assessment

The interim assessment will be conducted by the IDMC; the details of the procedure are described in the IDMC Charter. Given that much of the programming to derive the estimates will come from this SAP, a summary of the process is outlined below.

After 76 subjects either discontinue or complete the study, representing 50% of the original sample size, the interim analysis will be performed. The interim analysis will be used to re-estimate the sample size based on the following decision rule; the trial may be halted for futility based on the interim analysis results.

- If the conditional power is in the unfavorable zone, <39.6%, the study will be halted; no additional participants will be enrolled.
- If the conditional power is in the promising zone, 39.6% to 80%, the sample size will be increased to a final sample size (N_2, ≤250), where N_2 satisfies the condition that the conditional power is 80%.
- If the conditional power is in the favorable zone, >80%, the study will continue with the original sample size of 152.

The interim assessment will follow the method described in the published paper by Mehta and Pocock,[2] which states that if "the sample size is only increased when the interim conditional power falls within a certain 'promising zone' then the conventional test will indeed protect the type-I error." The minimum of conditional power is not so small as to result in an inflated overall type-1 error using a conventional final analysis. Therefore, implementing the interim analysis for the re-estimation of sample size should protect inflation of the type 1 error rate.

8.7.1. Monitoring of the Primary Efficacy Endpoint

The monitoring question that will be addressed by the IDMC during the clinical investigation is as follows:

- Does the conditional power from the interim evaluation of the primary efficacy endpoint equal or exceed 39.6%, allowing a re-estimation of the target sample size?

8.7.2. Site Surveillance

Not all of the data supplied to the IDMC will be fully monitored at the time of the interim assessment. Given that the data extract for the IDMC will occur during the course of the study, it may not be feasible to have all of the necessary study subjects' data monitored, and the database will not be locked. The IDMC will be charged with identification of any investigational sites where the interim data are either incomplete or inconsistent and should not be used in the interim calculation of response. Incomplete or inconsistent data may indicate that the study subject selection by the investigational site requires further review. Possible action, as a result of the recommendation from the IDMC, will include suspension of enrollment for an individual site until the study subject case records can be fully monitored. The objective of the intra-site surveillance is to ensure that failure to complete the study is not a result of inappropriate study subject selection, or failure to adhere to the requirements outlined in the protocol.

Independent of the decisions made by the IDMC, the final analysis will be performed using all of the final and monitored data.

The research questions that will be addressed by the IDMC, relative to site surveillance, covers

efficacy:

- Based on the pattern of missing or incomplete data, coupled with the attrition rate, does the IDMC recommend continuing enrollment of study subjects at each of the active investigational sites?
- Are there significant differences within each of the treatment arms across the active sites that may affect the poolability of the results across sites?

The IDMC will make one of following non-binding recommendations to the sponsor after conducting the interim assessment:

- Continue enrolling subjects to the pre-specified target sample size
- Continue enrolling subjects to a sample size that exceeds the pre-specified target sample size but less than or equal to the maximum sample size of 250
- Stop enrolling subjects based on futility

An alpha level adjustment will not be necessary for the conditional power procedure if the conditional power for the primary endpoint is $\geq 39.6\%$.

9. STATISTICAL ANALYSIS PLAN AMENDMENT HISTORY

[redacted]

10. REFERENCES

1. Center for Biologics Evaluation and Research, Center for Drug Evaluation and Research, Food and Drug Administration. *Guidance for Industry: E9 Statistical Principles for Clinical Trials*. US Department of Health and Human Services. September 1998.
2. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Stat Med*. 2011;30(28):3267-84.
3. Dmitrienko A, Millen BA, Brechenmacher T, Paux G. Development of gatekeeping strategies in confirmatory clinical trials. *Biom J*. 2011;53(6):875-93.
4. Unnebrink K, Windeler J. Intention-to-treat methods for dealing with missing values in clinical trials of progressive deteriorating diseases. *Stat Med*. 2001;20:3931-46.
5. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med*. 1992;11(13):1685-701.