# Statistical Analysis Plan

| | |
|---|---|
| **Protocol Title:** | A Phase 3, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study to Assess the Safety and Efficacy of VM202 in Participants with Painful Diabetic Peripheral Neuropathy |
| **SAP Version:** | VMDN-003-SAP; Version E |
| **Protocol Number:** | VMDN-003; Version H |
| **NCT Number**: | NCT02427464 |
| **Document Date**: | 05 September 2019 |

# Statistical Analysis Plan

## A PHASE III, DOUBLE-BLIND, RANDOMIZED, PLACEBO-CONTROLLED, MULTICENTER STUDY TO ASSESS THE SAFETY AND EFFICACY OF VM202 IN SUBJECTS WITH PAINFUL DIABETIC PERIPHERAL NEUROPATHY

## SAP Version: VMDN003-SAP/E

Corresponding Protocol Version: Protocol VMDN-003/H

**September 5, 2019**

**Sponsor:** **Helixmith Co., Ltd.**

[redacted]

**Prepared by:** **Helixmith Co., Ltd.**

[redacted]

**Revision:** **E**

## STATISTICAL ANALYSIS PLAN (SAP) VERSION CONTROL

| Version Number | Notes/Summary of Changes | SAP Writer | Date |
|---|---|---|---|
| E | **VMDN003-SAP/E**; corrections to definitions of the ITT, mITT, and PP populations **[Consistent with Protocol VMDN-003/H (September 4, 2019)]** | [redacted] | 05-Sep-2019 |
| D | **VMDN003-SAP/D** [Updated name and address of Sponsor (Helixmith); refined definitions of ITT, mITT, PP, and Safety populations; added ISR table; added NCV sub-population, Day 270 added to MMRM and GLMM models, added MNSI scoring algorithm, updated PP Medication Population definition, text added to specify assignment of subjects to protocol version; medications now summarized by preferred name instead of preferred term, updated multiple imputation code. Multiple imputation will be used for the 9 month missing data for the primary efficacy endpoint. Method = quad, short for quadrature, has been added to Proc Mixed procedure in Section 8.4.] **[Consistent with Protocol VMDN-003/G (July 30, 2019)]** | [redacted] | 30-Jul-2019 |
| C | **VMDN003-SAP/C** [Updated description of imputation methods, monofilament test, nerve conduction test] | [redacted] | 21-Mar-2019 |
| B | **VMDN003-SAP/B [Consistent with Protocol VMDN-003 / E (Jan. 29, 2018)]** | [redacted] | 01-May-2018 |
| A | **VMDN003-SAP/A [Original SAP]** | [redacted] | 18-Mar-2015 |

Abbreviations: [redacted] Helixmith statistical consultant
[redacted] Helixmith statistical consultant
[redacted] Helixmith statistical consultant
[redacted] Helixmith Senior VP, [redacted]

## SIGNATURE PAGE

We, the undersigned, confirm that we have read and are in agreement with the contents of this document.

[redacted]
Statistical Consultant

[redacted] 05-09-2019

*Signature* *Date (DD-MM-YYYY)*

[redacted]
Biostatistician
[redacted]

*Signature* *Date (DD-MM-YYYY)*

[redacted]
Senior Vice President
[redacted]
Helixmith Co., Ltd.

*Signature* *Date (DD-MM-YYYY)*

## SIGNATURE PAGE

We, the undersigned, confirm that we have read and are in agreement with the contents of this document.

[redacted]
Statistical Consultant

*Signature* *Date (DD-MM-YYYY)*

[redacted]
Biostatistician
[redacted]

*Signature* 05-Sept.-2019 *Date (DD-MM-YYYY)*

[redacted]
Senior Vice President
[redacted]
Helixmith Co., Ltd.

*Signature* 05-Sep-2019 *Date (DD-MM-YYYY)*

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AE | Adverse Event |
| AESI | Adverse Event of Special Interest |
| BPI-DPN | Brief Pain Inventory for Diabetic Peripheral Neuropathy |
| BPNS | Brief Peripheral Neuropathy Screening |
| 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 |
| HBV | Hepatitis B Virus |
| HCV | Hepatitis C Virus |
| HEENT | head, eyes, ears, nose, and throat |
| HGF | Hepatocyte Growth Factor |
| HIV | Human Immunodeficiency Virus |
| HTLV | Anti-Human T-Cell Lymphotropic Virus |
| IM | Intra-Muscular |
| 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 AE |
| VAS | Visual Analog Scale |
| WHODrug | WHO 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. |
| Baseline | The last non-missing value prior to first dose of study drug. |
| 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 | 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 to be implemented during the analyses of data collected within the scope of Helixmith Co., Ltd. Protocol VMDN-003 [A Phase III, Double-blind, Randomized, Placebo-controlled, Multicenter Study to Assess the Safety and Efficacy of VM202 in Subjects with Painful Diabetic Peripheral Neuropathy]. The purpose of this plan is to provide specific guidelines from which the analysis will proceed. Any deviations from these guidelines will be documented in the clinical study report (CSR).

## 2. OBJECTIVES

- To evaluate the safety of intramuscular (IM) administration of VM202 in subjects with painful diabetic peripheral neuropathy (DPN) in lower extremities.

- To evaluate the efficacy of IM administration of VM202 in subjects with painful DPN in the lower extremities, when compared to placebo, on pain.

## 3. STUDY DESIGN

This is a phase III, double-blind, randomized, placebo-controlled, multicenter, 9-month study designed to assess the safety and efficacy of bilateral IM injections of VM202 in subjects with painful DPN. Subjects with painful DPN will be screened for study eligibility after giving informed consent.

Patients (from up to 30 sites) who meet the eligibility criteria will be randomized in a 2:1 ratio to one of two treatment groups: VM202 (16 mg) or placebo (VM202 vehicle), respectively. The randomization will be stratified by current use of gabapentin and/or pregabalin so that the enrolled eligible subjects will be randomized using different randomization schedules (with a ratio of 2:1 to VM202 and placebo) based on their use of gabapentin and/or pregabalin. A single treatment with VM202 is delivered as an equally divided dose administered two weeks apart. Subjects will receive VM202 or placebo by intramuscular injections in both legs (in the calf) on Day 0 and Day 14. Subjects will receive a second treatment on Day 90 and Day 104. Injections will be administered as follows:

First Treatment: Days 0, 14

| Treatment Group | Dose VM202 (mg) / Visit / Leg | | Final Dose VM202 / Leg (mg) | Final Dose VM202 / Subject / Treatment (mg) |
|---|---|---|---|---|
| | Day 0 | Day 14 | | |
| VM202 | 4 | 4 | 8 | 16 |
| Placebo | 0 | 0 | 0 | 0 |

0 = injections of VM202 vehicle

Second Treatment: Days 90, 104

| Treatment Group | Dose VM202 (mg) / Visit / Leg | | Final Dose VM202 / Leg (mg) | Final Dose VM202 / Subject / Treatment (mg) |
|---|---|---|---|---|
| | Day 90 | Day 104 | | |
| VM202 | 4 | 4 | 8 | 16 |
| Placebo | 0 | 0 | 0 | 0 |

0 = injections of VM202 vehicle

Patients randomly assigned to the VM202 group will receive the following intramuscular injections in each calf:

- Day 0: 16 injections of 0.5 mL of VM202/calf
- Day 14: 16 injections of 0.5 mL of VM202/calf
- Day 90: 16 injections of 0.5 mL of VM202/calf
- Day 104: 16 injections of 0.5 mL of VM202/calf

Patients randomly assigned to the placebo control group will receive the following intramuscular injections in each calf:

- Day 0: 16 injections of 0.5 mL of VM202 vehicle/calf
- Day 14: 16 injections of 0.5 mL of VM202 vehicle/calf
- Day 90: 16 injections of 0.5 mL of VM202 vehicle/calf
- Day 104: 16 injections of 0.5 mL of VM202 vehicle/calf

The schedule of study visits and the clinical parameters that will be measured at the visits are summarized in Table 1 below.

## TABLE 1. SCHEDULE OF EVALUATIONS AND VISITS

| PROCEDURE | Screening / Baseline (-90 – 0 D) | First Treatment: D0, D14 | | | | Day 21 ± 3 D | 2M Day 60 ± 3 D | Second Treatment: D90, D104 | | | | Day 111 7 ± 3 Days after 4th Injection |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1st Injection Day 0 | | 2nd Injection Day 14 ± 1 D | | | | 3rd Injection Day 90 ± 7 D | | 4th Injection 13-15 Days after D 90 | | |
| | | Pre-dose | Post-dose* | Pre-dose | Post-dose* | | | Pre-dose | Post-dose* | Pre-dose | Post-dose* | |
| **Visit Number** | 1 | 2 | | 3 | | 4 | 5 | 6 | | 7 | | 8 |
| **Baseline Evaluation** | | | | | | | | | | | | |
| Informed Consent | ✓ | | | | | | | | | | | |
| Medical History | ✓ | ✓ | | | | | | | | | | |
| Physical Exam | ✓ | | | | | | | | | | | |
| Symptoms of BPNS | ✓ | | | | | | | | | | | |
| Cancer Screening† | ✓ | | | | | | | | | | | |
| Viral Screening – HIV, HTLV, HBV, HCV | ✓ | | | | | | | | | | | |
| ECG | ✓ | | | | | | | | | | | |
| Urine Pregnancy Test | ✓ | | | | | | | | | | | |
| Retinal Fundoscopy | ✓ | | | | | | | | | | | |
| **Safety and Efficacy Parameters** | | | | | | | | | | | | |
| MNSI | ✓ | | | | | | | | | | | |
| VAS | ✓ | ✓ | | | | | | ✓ | | | | |
| Daily Pain and Sleep Interference Diary | ✓ | | | | | | | ✓ | | | | |
| HbA1c | ✓ | ✓ | | | | | | ✓ | | | | |
| Serum Chemistry and Hematology | ✓ | ✓ | | | | | | ✓ | | | | |
| Vital Signs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Concomitant Medications | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | | ✓ | | ✓ |
| BPI-DPN | | ✓ | | | | | | ✓ | | | | |
| Semmes-Weinstein Filament Test | | ✓ | | | | | | ✓ | | | | |
| PGIC | | | | | | | | ✓ | | | | |
| Study Injections | | ✓ | | ✓ | | | | ✓ | | ✓ | | |
| Copies of VM202 in whole blood | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Serum HGF | | ✓ | | ✓ | | ✓ | ✓ | ✓ | | ✓ | | ✓ |
| Nerve Conduction (only at select sites) | | ✓ | | | | | | | | | | |
| Tylenol Usage | | | | ✓ | | ✓ | ✓ | ✓ | | ✓ | | ✓ |
| **Treatment** | | | | | | | | | | | | |
| Injection Site Reaction Assessment | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Adverse Events | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

† Cancer screening: chest X-ray or chest CT scan if subject has a previous history of tobacco use within 3 months; pap smear and mammogram within past 12 months (females only); for subjects ≥ 50 years old, fecal occult blood test.

* 2 hours after injection (± 1 hour)

## TABLE 1 SCHEDULE OF EVALUATIONS AND VISITS (CONTINUED)

| PROCEDURE | 5M Day 150 ± 7 D | 6M Day 180 ± 7 D | 9M Day 270 ± 14 D | Early Withdrawal |
|---|---|---|---|---|
| **Visit Number** | **9** | **10** | **11** | |
| **Safety and Efficacy Parameters** | | | | |
| Vital Signs | ✓ | ✓ | ✓ | ✓ |
| Concomitant Medications | ✓ | ✓ | ✓ | ✓ |
| Retinal Fundoscopy | | ✓ | ✓ | ✓ |
| VAS | | ✓ | ✓ | |
| MNSI | | ✓ | ✓ | |
| Daily Pain and Sleep Interference Diary | | ✓ | ✓ | |
| BPI-DPN | | ✓ | ✓ | |
| PGIC | | ✓ | ✓ | |
| Semmes-Weinstein Filament Test | | ✓ | ✓ | |
| Nerve Conduction (only at select sites) | | ✓ | ✓ | |
| HbA1c | | ✓ | ✓ | |
| Serum Chemistry and Hematology | | ✓ | ✓ | ✓ |
| Study Injections | | | | |
| Copies of VM202 in whole blood | ✓ | ✓ | ✓ | ✓ |
| Serum HGF | ✓ | ✓ | ✓ | ✓ |
| Tylenol Usage | ✓ | ✓ | ✓ | ✓ |
| **Treatment** | | | | |
| Injection Site Reaction Assessment | ✓ | | | ✓[1] |
| Adverse Events | ✓ | ✓ | ✓ | ✓ |

1 If withdrawal occurs before Day 150 Visit

The first 62 subjects were randomized under protocol revision C. Protocol revision D was submitted on November 10th, 2016. The protocol revision from C to D did not change any efficacy assessment but clarified that the nerve conduction testing safety assessment would be conducted on a single leg and not bilaterally.

Approximately 278 additional subjects were randomized under protocol version D prior to implementation of version E. Protocol revision E, submitted on January 29th, 2018, did not change any of the efficacy assessments; notable changes included:

- Change in the wording of exclusion criteria 17, 18, and 20 to define the period during which prohibited medications may not be taken: 'for the first 6 months of the study' changed to 'until Day 180 visit of the study.' This ensures that primary and secondary endpoints will be captured without interference from prohibited medications.
- Addition of nerve conduction testing at Day 270.

Descriptive summaries of the primary efficacy endpoints and overall adverse event rates will be produced for the subjects enrolled under different versions of the protocol.

# 4. STUDY ENDPOINTS

## 4.1. Primary Efficacy Endpoints

There are two primary efficacy endpoints that will be evaluated in sequential order. The primary efficacy endpoints are as follows:

1. The change in the average 24-hour pain score from baseline to the 3-month follow-up [Day 90] obtained from the Daily Pain and Sleep Interference Diary.

2. The outcome of at least a 50% reduction (i.e., $\geq 50\%$) in the average 24-hour pain score from baseline to the 3-month follow-up obtained from the Daily Pain and Sleep Interference Diary.

The statistical hypotheses for the first primary efficacy endpoint are:

$H_0$: $\mu_t = \mu_p$ versus $H_a$: $\mu_t \neq \mu_p$, (I)

where $\mu_t$ and $\mu_p$ are the mean pain change from baseline to the 3-month follow-up for the VM202 and Placebo groups, respectively. A negative mean value indicates a reduction in the pain score, and a positive mean value indicates an increase in the pain score.

If the null hypothesis for the first primary efficacy endpoint above is rejected, then the formal statistical test will be performed for the second primary efficacy endpoint. The statistical hypotheses for the second primary efficacy endpoint are:

$H_0$: $p_t = p_p$ versus $H_a$: $p_t \neq p_p$, (II)

where $p_t$ and $p_p$ are the percentage of subjects with a change in the average 24-hour pain score from the Daily Pain and Sleep Interference Diary from baseline to the 3-month follow-up of a ≤ -50% (i.e., a reduction of ≥ 50%) for the VM202 and Placebo groups, respectively.

Since the formal statistical test for the second primary efficacy endpoint will not be performed if the null hypothesis of the first primary endpoint is not rejected, the significance level for both sets of the statistical hypotheses is not adjusted and kept at a two-sided 0.05.

### 4.2. Secondary Efficacy Endpoints

There are two key secondary endpoints that will be evaluated in sequential order. The secondary efficacy endpoints are

1. The change in the average 24-hour pain score from baseline to the 6-month follow-up [Day 180] (3 months after the Day 90 injection) obtained from the Daily Pain and Sleep Interference Diary.

2. The outcome of at least a 50% reduction (i.e., ≥ 50%) in the average 24-hour pain score from baseline to the 6-month follow-up [Day 180] (3 months after the Day 90 injection) obtained from the Daily Pain and Sleep Interference Diary.

The same method used to evaluate the two sequential hypotheses described in Section 4.1 will be applied to the secondary efficacy endpoints. Similarly, the hierarchical approach will be used for the evaluation of the outcomes from the secondary efficacy endpoints.

### 4.3. Exploratory Efficacy Endpoints

- The change in the average 24-hour pain score from baseline to the 9-month follow-up [Day 270] (6 months after the Day 90 injection) obtained from the Daily Pain and Sleep Interference Diary.

- The outcome of at least a 50% reduction (i.e., ≥ 50%) in the average 24-hour pain score from baseline to the 9-month follow-up [Day 270] (6 months after the Day 90 injection) obtained from the Daily Pain and Sleep Interference Diary.

- The outcome of at least a 20%, 30%, 40%, 60%, or 70% reduction in the average 24-hour pain score from baseline to the 3-month follow-up obtained from the Daily Pain and Sleep Interference Diary

- The outcome of at least a 20%, 30%, 40%, 60%, or 70% reduction in the average 24-hour pain score from baseline to the 6-month follow-up obtained from the Daily Pain and Sleep Interference Diary

- The outcome of at least a 20%, 30%, 40%, 60%, or 70% reduction in the average 24-hour pain score from baseline to the 9-month follow-up obtained from the Daily Pain and Sleep Interference Diary

- Change in VAS for Pain from baseline to the 3-month, 6-month, and 9-month follow-up; note that VAS is read by 2 different investigators at each assessment and the average of the two scores will be used for analysis

- Michigan Neuropathy Screening Instrument (MNSI) physical assessment and history changes from baseline to the 6-month and 9-month follow-up

- Change in average sleep interference score from baseline to the 3-month, 6-month, and 9-month follow-up obtained from the Daily Pain and Sleep Interference Diary

- BPI-DPN pain interference score changes from baseline to the 3-month, 6-month, and 9-month follow-up; note that an average score of BPI-DPI pain interference will be analyzed

- BPI-DPN pain severity score changes from baseline to the 3-month, 6-month, and 9-month follow-up; note that not only an overall average score of BPI-DPN pain severity but also each individual severity components (worst pain, least pain, average pain, pain now) will be analyzed

- Patient's Global Impression of Change (PGIC) at the 3-month, 6-month, and 9-month follow-up

- Semmes-Weinstein monofilament testing changes from baseline to the 3-month, 6-month, and 9-month follow-up.

### 4.4. Safety Outcomes

- Adverse events
- Injection site reaction assessment
- Vital signs
    - Blood pressure
    - Weight
    - Heart rate
    - Respiration Rate
    - Temperature
- HbA1c

- Serum Chemistry and Hematology
- Retinal fundoscopy.

### 4.5. Other Clinical Parameters

- Nerve conduction at the 6-month and 9-month follow-up (selected sites)
- Total Tylenol (rescue medication) used during the study.

### 4.6. Planned Covariates

As defined in the primary analysis of the primary and secondary efficacy endpoints, the baseline pain and the randomization stratification factor (i.e., baseline use of gabapentin and/or pregabalin) with be introduced as covariates within the models.

As a separate secondary examination of the primary and secondary efficacy endpoints, three additional covariates will be introduced into the model to derive the covariate-adjusted estimates. (ref. Section 8.4.6 for additional details):

- Baseline HbA1c (< and ≥ median)
- Gender (male and female)
- Age (<65 years and ≥ 65 years).

## 5. SAMPLE SIZE CALCULATION

Table 2 provides the summary statistics for the corresponding efficacy endpoints based on change in the average pain score for the low dose and placebo groups from the intent-to-treat population and efficacy groups in the Phase II study (Protocol VMDN-002, injections at Day 0 and Day 14 only).

**TABLE 2. Mean and Standard Deviation of Pain Score Changes from Baseline and the Percentage of Subjects with at Least a 50% Reduction; Low Dose versus Placebo Group in Phase II Study**

| | ITT | | EFFICACY | |
|---|---|---|---|---|
| | VM202 | PLACEBO | VM202 | PLACEBO |
| Mean (SD) | | | | |
| Month 3 | -2.89 (2.47) | -1.70 (1.72) | -3.03 (2.53) | -1.53 (1.76) |
| Month 6 | -2.58 (2.18) | -1.63 (1.75) | -2.78 (2.23) | -1.59 (1.89) |
| Percentage of Subjects with a ≥ 50% Reduction | | | | |
| Month 3 | 45.9% | 23.8% | 48.4% | 17.6% |
| Month 6 | 34.3% | 15.0% | 38.7% | 17.6% |

The sample size for the primary efficacy endpoints was calculated based on the following assumptions:

- Based on the Phase II findings, it is assumed that the standard deviation of the pain score change from baseline will be 3.0. The true percentage of subjects with a pain reduction of ≥ 50% is assumed to be 35% and 18% for the VM202 group and placebo group, respectively.

- The statistical power is 90% for a detectable mean difference of –1.0 (i.e., mean pain reduction for the VM202 group is higher than that for the Placebo group) for Statistical Hypotheses (I) in Section 4.1 using a t-test. The statistical power is also 90% for Statistical Hypotheses (II) in Section 4.1 provided the assumptions of $p_t = 35\%$ and $p_c = 18\%$ using a chi-square test.

- The two-sided significance level is 0.05. Since a formal statistical conclusion for the second primary efficacy endpoint will not be made unless the null hypothesis of the first primary efficacy endpoint is rejected, no adjustment on the significance level is performed.

- The randomization ratio for the VM202 group and the placebo group is 2:1.

Based on the assumptions above, the sample sizes calculated for the VM202 group and placebo group are 286 and 143 subjects, respectively, for Statistical Hypotheses (I); those for Statistical Hypotheses (II) are 212 VM202 subjects and 106 placebo subjects.

Therefore, 286 VM202 subjects and 143 placebo subjects are needed with data at the 3-month follow-up for the primary data analyses. With a dropout rate of 10%, at least 477 subjects should be randomized in order to have 318 VM202 subjects and 159 placebo subjects.

# 6. ANALYSIS POPULATIONS

## 6.1. Intent-to-Treat (ITT) Population

This population includes all subjects who are randomized. In the efficacy outputs, subjects in the ITT population will be analyzed according to the randomized treatment assignment, regardless of the actual treatment administered. In the safety outputs, subjects will be analyzed using the treatment they received in the VMDN-003 study, regardless of original treatment assigned.

The **intent-to-treat (ITT) population** will include all subjects who were randomized on the VMDN-003 study.[1] 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.

For the ITT population, entries in the Daily Pain and Sleep Interference Diary will be considered fully valid if made anytime during the 24 hour period specified in the date entry.

## 6.2. Safety Population

The safety population will contain all subjects who are randomized and receive at least one study drug injection. Subjects will be grouped according to the actual treatment administered, not according to their randomization assignment. Subjects treated with any

---

[1] Per FDA/ICH E9 Statistical Guidance, pg. 29:

> "In some situations, it may be reasonable to eliminate from the set of all randomized subjects any subject who took no trial medication. The intention-to-treat principle would be preserved despite the exclusion of these patients provided, for example, that the decision of whether or not to begin treatment could not be influenced by knowledge of the assigned treatment. In other situations it may be necessary to eliminate from the set of all randomized subjects any subject without data post randomization. No analysis should be considered complete unless the potential biases arising from these specific exclusions, or any others, are addressed."

Helixmith is aware of 7 subjects in VMDN-003 who were randomized in error (did not meet Inclusion / Exclusion criteria or who were randomized without PI oversight). These 7 subjects were subsequently withdrawn before receiving study medication and without knowledge of treatment assignment. Details are summarized in a Note-to-File, July 12, 2019, which was prepared according to Helixmith SOP QA-002.

VM202 dose will be grouped in the VM202 group; subjects never treated with any VM202 will be grouped in the placebo group.

Subjects that sign informed consent but screen fail will still be considered part of the safety population and will have their collected safety data summarized.

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

The mITT population includes all subjects randomized that meet the following criteria:

- Received at least one dose of study medication.
- Correctly completed the Daily Pain and Sleep Interference Diary at the 3-month follow-up, i.e., by completing a minimum of 5 of 7 days of diary entries within 14 days prior to the Day 90 visit.
- For the mITT population, entries in the Daily Pain and Sleep Interference Diary will be considered fully valid if made anytime during the 24 hour period specified in the date entry.
- Satisfies Inclusion/Exclusion criteria

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

### 6.4 Per Protocol (PP) Populations

Three PP populations are defined. All PP population should be absent of any major protocol deviations.

The Per Protocol Medication population is a subset of the mITT and corresponds to the mITT population defined in protocol version H. It includes all mITT subjects who meet the following criterion:

- Have not used the protocol specified prohibited concomitant medications, such as COX-2 inhibitor drug(s) or non-specific COX-1/COX-2 inhibiting drugs which may interfere with VM202 or pain medication usage which may intensify the effect of VM202/placebo on pain, for more than 14 cumulative days during the entire study. The use and effect of protocol specified prohibited concomitant medications will be determined by the Clinical Data Review Committee (CDRC) in a blinded fashion prior to analyses.

The Per Protocol Dosing population is a further subset of the mITT and corresponds to the PP population defined in protocol version H. It includes all mITT subjects who meet the following criterion:

- Subject received all injections based on the randomized treatments.
- Additional criteria, if any, established before unblinding the randomization code by the CDRC that is masked to the treatment information of each study subject.

The Per Protocol Diary population is a subset of the mITT and includes all mITT subjects whose Daily Pain and Sleep Interference Diary at the 3-month follow-up is completed for a minimum of 5 of 7 days within 14 days prior to the Day 90 visit.

All three PP populations will be used in the sensitivity analyses for the primary efficacy endpoints. The PP Medication and PP Dosing populations will also be used in sensitivity analyses of the 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 randomization stratification factor. Age will also be categorized as <65 years and ≥ 65 years. These subgroups will be re-examined and may be re-categorized or eliminated due to small sample size (if there are < 10% of subjects within each subgroup) before unblinding for analysis. For example, if < 10% of overall subjects are ≥ 65 years, then analyses for this subgroup will not be performed. Descriptive summaries for the primary and secondary efficacy endpoints will also be provided for each study center regardless of the sample size per center, and similarly for the different protocol versions.

Nerve conduction velocity (NCV) analysis will be performed on an NCV sub-population. The NCV sub-population will include all randomized subjects that have at least one post-baseline NCV assessment.

The treatment by subgroup interaction will be examined and tested as described in Section 8.4.5. Data Handling

## 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 and sleep interference scores from the Daily Pain and Sleep Interference Diary, the baseline value is the average 24-hour pain score and average sleep interference score from the diary collected prior to Day 0.

It should be noted that, for eligibility, the average 24-hour pain score of the Daily Pain and Sleep Interference Diary completed after medication wash-out, if applicable, should be $\geq 4$ with a standard deviation $\leq 2$ at Screening and within the 14 days prior to Day 0. For the average 24-hour pain score or the sleep interference 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 (Table 3) 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 the other visits (if any) will be provided in data listings only.

**TABLE 3. Target Day**

| VISIT | 1ST INJECTION | 2ND INJECTION | DAY 21 | MONTH 2 | 3RD INJECTION | 4TH INJECTION | DAY 111 | MONTH 5 | MONTH 6 | MONTH 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| TARGET DAY | 0 | 14 | 21 | 60 | 90 | 104 | 111 | 150 | 180 | 270 |

### 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 7.6) members and the team that will prepare the unmasking summary tables for the DSMB meetings. However, in order 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 subject 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

There is no adjustment of the type 1 error rate based on multiplicity. A step-down approach will be used that controls for the potential inflation of the type 1 error rate.

The analyses of the primary efficacy endpoints will be conducted hierarchically based on the order of the two primary efficacy endpoints. Similarly, the analyses of the two secondary efficacy endpoints will be conducted hierarchically based on the corresponding order for the two secondary efficacy endpoints. The secondary efficacy endpoints will be tested only when the null hypotheses of both the primary efficacy endpoints are rejected at a significance level of 0.05[2].

The significance level for each exploratory efficacy endpoint will be 0.05.

All statistical testing will be 2-sided.

### 7.5. Data Safety Monitoring Board

An 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 study sponsor regarding the safety of the VM202. The data analyses for the DSMB meetings will be directly

---

[2] Alex Dmitrienko and Ralph D'Agostino, Sr., "Traditional multiplicity adjustment methods in clinical trials." Statistics in Medicine, 2013; 32(29):5172-218.

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

Subjects may have missing specific data points for a variety of reasons. In general, data may be missing due to a subject'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.

### 7.7. Missing Average 24-hour Pain Scores at 3, 6, or 9 Months

The Daily Pain and Sleep Interference Diary is to be completed within 14 days of the 3-, 6- and 9-month visits. The average 24-hour pain score for a visit will be considered as missing if fewer than 5 of 7 days of Daily Pain and Sleep Interference Diary entries are provided. Sensitivity analyses for the mean change in the average 24-hour pain score will include the following imputation approaches for missing values at 3-months, 6-months, or 9-months. It is important to note that over-stratification may result when the total list of pre-specified factors are considered. Under this scenario, the least represented factor will be removed and the imputation will be re-run. This process will be followed until the imputed dataset is complete.

- Multiple imputation: Each missing pain score will be imputed ten times to generate ten imputed complete data sets based on the Markov Chain Monte Carlo (MCMC) method with baseline pain score, 3-month, 6-month, 9 month pain score, and categorical covariates of baseline use of gabapentin and/or pregabalin, baseline HbA1c, gender, and age. The imputed score will be rounded to the first decimal point. The results of the ten tests from the continuous Repeated Measures Model (Section 8.4.2) using these data will be combined.

- Mean of the other group (MOTH)[3] as follows:

    1. For a subject with a missing 3-month, 6-month, or 9-month average 24-hour pain score, identify the subject's following baseline characteristics:
        - Study treatment group
        - Baseline use of gabapentin and/or pregabalin.

---

[3] Unnebrink K and Windeler J, "Intention-to-treat methods for dealing with missing values in clinical trials of progressive deteriorating diseases," Statistics in Medicine, 2001;20: 3931-3946.

- Baseline average 24-hour pain score (< median or ≥ median)
- HbA1c (< median or ≥ median)
- Gender (male or female)
- Age (<65 years and ≥ 65 years).

2. The missing 3-month, 6-month, or 9-month average 24-hour pain score will be imputed by using the mean average 24-hour pain score obtained at the same time point of those subjects in the other treatment group who match the subject's baseline characteristics. For example, for the missing pain scores of the VM202 subjects, the mean pain scores of the placebo subjects 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 8.4.2) analysis.

Sensitivity analyses for the percentage of subjects with a reduction in the average 24-hour pain score of at least 50% will include the following imputation approaches for missing values at 3, 6, or 9 months.

- Imputed average 24-hour pain scores at 3, 6, or 9 months from the multiply-imputed pain score datasets described above. The results of the ten tests from the categorical Repeated Measures Model (Section 8.4.2) will be combined.

- Missing pain score will be imputed from the MOTH-imputed dataset described above and the categorical Repeated Measures Model (Section 8.4.2) will be used for the data analyses.

- Multiple imputation for the missing responder outcomes, 1 (Yes) or 0 (No), by fully conditional specification logistic regression method: each missing responder outcome will be imputed ten times to generate ten imputed complete data sets using a logistic regression model with baseline use of gabapentin and/or pregabalin, and categorical baseline HbA1c, gender, and age as covariates. The results of the ten tests from the categorical Repeated Measures Model (Section 8.4.2) will be combined.

Steps to generate the MMRM analysis are presented below; different options will be used depending on the exact model being run:

Run PROC MI using output dataset ("outmi") containing the original and imputed values, add "by imputation" to the SAS code and "solution" as an option on the MIXED model statement. The resulting file is available via ODS output: solutionf=mixparms;"

Run the following PROC MIANALYZE code:
PROC MIANALYZE parms(classfull)=mixparms;
class trtp avisitn basemed;
modeleffects intercept trtp avisitn trtp*avisitn basemed;
run;

### 7.7.1. Missing Values for BPI-DPN and MNSI

No imputation for missing individual item scores within the BPI-DPN (both pain severity and interference) and MNSI (both physical assessment and history) questionnaire will be performed. Missing individual item scores of the BPI-DPN may 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. A similar method will be used when deriving total MNSI scores 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.

### 7.7.2. 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 specified in Table 4 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 4. 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 |

| | | | |
|---|---|---|---|
| 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 subjects have had an opportunity to complete their 9-month follow-up visit.

The primary analyses of the safety endpoints 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 8.4.4).

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

Categorical variables will be summarized by counts and by the percentage of subjects 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. Subject Enrollment and Disposition

The reasons for subject 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. Subject disposition will be summarized for all randomized subjects. The summary including the number and percentage (based on total number of subjects randomized) of subjects in each of the following categories will be prepared:

- Available at each of the protocol-specified visits based on the ITT population
- Completing 9-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 subjects not in the Per-Protocol Populations will be listed.

### 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:

- Use of gabapentin and/or pregabalin
- Baseline average 24-hour pain score obtained from the Daily Pain and Sleep Interference Diary
- Baseline average sleep interference score obtained from the Daily Pain and Sleep Interference Diary
- Vital signs: blood pressure, weight, BMI, heart rate, respiration rate, temperature
- Physical examination: head, eyes, ears, nose, and throat (HEENT), heart, lungs, abdomen, extremities, lymph nodes, musculoskeletal, neurological, and gastrointestinal systems. Any abnormalities are categorized as clinically significant (CS) or not clinically significant (NCS).
- Diabetes type
- Medical history categorized by Medical Dictionary for Regulatory Affairs (MedDRA) System Organ Class (SOC) and preferred term (version 21.0 or later)
- Symptoms of BPNS: total score for right and left leg separately and combined.
- Cancer screening findings: Positive and Negative

- Viral screening findings: Positive and Negative
- 12-lead ECG: Normal, Abnormal NCS, Abnormal CS
- Urine pregnancy test: Positive, Negative, and Not Applicable.

These parameters will be summarized by treatment group for the ITT and included in data listings. Other collected baseline characteristics will be listed only. 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. 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

### 8.4.1. Primary and Secondary Efficacy Endpoints

Subjects will be asked to assess the level of pain they feel by selecting a score from 0 (No Pain) to 10 (Worst Possible Pain) in the Daily Pain and Sleep Interference Diary for 7 days at screening (the baseline for the pain score), 3-month, 6-month, and 9-month visits. The Daily Pain and Sleep Interference Diary must be completed within the 14 days prior to each specified visit. The average of the available pain scores from the Daily Pain and Sleep Interference Diary will be calculated for each subject at each visit and will be rounded to the first decimal point.

The primary efficacy endpoints are the change in average 24-hour pain score as determined by the Daily Pain and Sleep Interference Diary between baseline and the 3-month follow-up and outcome of a pain score change of ≤ -50% at 3 months. The change and the percent change in pain will be calculated for each subject as follows:

$$\text{Change} = \text{3-month Pain Score} - \text{Baseline Pain Score}$$

$$\%\ \text{Change} = \text{Change} \div \text{Baseline Pain Score} \times 100\%.$$

Since higher scores indicate worse pain, a negative value of change means an improvement, and a positive value of change means deterioration. Subjects with a percent change of ≤ -50% (i.e., reduction of at least 50%) will be classified as a responder at 3 months. The means of the change in the average 24-hour pain score at 3 months and the percentage of subjects with a change in the average 24-hour pain score of ≤ -50% will be compared between the treatment groups (VM202 and placebo) at 3 months.

The key secondary efficacy endpoints are the change in the average 24-hour pain score from baseline to the 6-month follow-up, and the outcome of reduction in average 24-hour pain score of at least 50% at 6 months. The means of the change in the average 24-hour pain score at 6 months and the percentage of subjects with a change in the average

24-hour pain score of ≤ -50% will be compared between the treatment groups (VM202 and placebo) at 6 months.

### 8.4.2. Primary Analysis of the Primary Efficacy Endpoints

The primary analysis for comparing the mean change in pain score at 3 months from baseline between the treatment groups will be based on the ITT using a linear mixed-effects model for repeated measures (hereinafter, the continuous Repeated Measures Model)[4]. The model will include treatment, visit (3-month, 6-month, and 9-month visits), treatment-by-visit interaction, and baseline use of gabapentin and/or pregabalin as the main fixed effects, and baseline average 24-hour pain score as a covariate using an unstructured variance-covariance matrix. The point estimate for the least-squares mean of the treatment difference (VM202 – Placebo) at 3 months and the corresponding 95% confidence interval and p-value will be summarized.

The SAS code to be used to conduct the analysis is presented below:

```
proc mixed data=work order=internal method = quad(Qpoint = x);
  class usubjid trtp avisitn basemed;
  model chg = trtp avisitn trtp*avisitn base basemed / ddfm=KR;
  repeated avisitn /subject = usubjid type = un;
  lsmeans trtp trtp*avisitn / pdiff cl;
run;
```

Where
USUBJID: subject ID
TRTP: treatment
AVISITN: variable representing visits (Primary model will include the 3, 6, and 9 month visits)
CHG: change in average pain score from baseline
BASE: baseline average pain score
BASEMED: a Y/N variable indicating if the subject took gabapentin/pregabalin at baseline.
Qpoint: Number of quadrature nodes to clarify the dimensionality

Other variance-covariance structures selected from among compound symmetry, Toeplitz, and autoregressive (1) options will be substituted based on the lowest AIC if convergence problems arise.

Similarly, a generalized linear mixed-effects model for repeated measures based on a logit link function (hereinafter, the categorical Repeated Measures Model) will be used for comparing the percentage of subjects with a percent change in average 24-hour pain score of ≤ - 50% at 3 months (responder rate) between the two study treatment groups.

---

[4] Vonesh EF and Chinchilli VM (1996), Linear and Nonlinear Models for the Analysis of Repeated Measurements, New York: Marcel-Dekker.

The model will include treatment, visit, treatment-by-visit interaction, baseline use of gabapentin and/or pregabalin, and baseline average 24-hour pain scores as covariate with an unstructured variance-covariance matrix. The point estimates for the least-squares mean of the treatment difference (VM202 – Placebo) at 3 months and the corresponding 95% confidence interval and p-value will be summarized.

The SAS code to be used to conduct the analysis is presented below:

```
PROC GLIMMIX DATA=work method=quad(Qpoint= );
  class usubjid trtp avisitn basemed;
  MODEL critvar(event='Y') = trtp base avisitn basemed trtp*avisitn / DIST=binary
LINK=LOGIT SOLUTION;
  random INTERCEPT / SUBJECT=usubjid type=un;
  lsmeans trtp *avisitn / diff cl;
RUN;
```

Where USUBJID: subject ID
TRTP: treatment
AVISITN: variable representing visits (Primary model will include the 3, 6, and 9 month visits)
CHG: change in average pain score from baseline
BASE: baseline average pain score
BASEMED: a Y/N variable indicating if the subject took gabapentin/pregabalin at baseline.
CRITVAR: a Y/N variable indicating if the subject had at least a 50% reduction in baseline pain score from baseline at that visit
Qpoint: Number of quadrature nodes to clarify the dimensionality

Other variance-covariance structures selected from among compound symmetry, Toeplitz, and autoregressive (1) options will be substituted based on the lowest AIC if convergence problems arise.

### 8.4.3. Primary Analysis of the Secondary Efficacy Endpoints

Analyses for the primary efficacy endpoints described in Section 8.4.2 will automatically produce the analysis results for the key secondary efficacy endpoints. The point estimate for the least-squares mean of the treatment difference (VM202 – Placebo) at 6 months and the corresponding 95% confidence interval and p-value will be summarized. To control the overall significance level, statistical inferences regarding the treatment effect on secondary efficacy endpoints will be made only if the treatment effect on both primary efficacy endpoints are statistically significant in their respective primary analysis.

### 8.4.4. Sensitivity Analysis of the Primary and Secondary Efficacy Endpoints

To further evaluate the robustness of the primary and secondary efficacy outcomes, the Repeated Measures Model analyses described in Section 8.4.2 will be performed on the imputed data sets described in Section 7.7.1.

Additionally, two sets of the Repeated Measures Model analyses will be performed without missing value imputation. For the first set, analysis will be conducted in the ITT, mITT, and PP populations using all available average 24-hour pain score data (i.e., observed cases). For the second set conducted in the ITT and mITT populations, any individual pain scores influenced by the protocol-prohibited concomitant medications, as described in Section 6.4, will also be excluded.

The point estimates for the least-squares mean of the treatment difference (VM202 – Placebo) at 3 and 6 months, and the corresponding 95% confidence intervals and p-values, will be summarized.

### 8.4.5. Subgroup Analysis of Primary and Secondary Efficacy Endpoints

The subgroups are described in Sections 4.6 and 6.5. For each subgroup of the ITT population, the change in average 24-hour pain score will be summarized by treatment group and visit using descriptive statistics including mean, median, standard deviation, minimum, and maximum. The count and percentage of subjects with a change in the average 24-hour pain score will also be summarized by treatment group and visit for each subgroup of the ITT population. Both types of descriptive summaries are also provided for the subgroups of baseline use of gabapentin and/or pregabalin, study center, and protocol version. Subjects will be assigned to a protocol version based on their randomization date and the date of IRB approval at their site.

Except for the subgroups of baseline use of gabapentin and/or pregabalin, study center, and protocol version, the Repeated Measures Model analyses described in Section 8.4.2 will be performed within each subgroup.

Additionally, the possible treatment-by-subgroup interaction will be tested for each subgroup variable as follows:

- A subgroup variable will be included in the primary analysis models described in Section 8.4.2 along with its interaction with the treatment effect. If the p-value of the interaction term is ≥ 0.05, the treatment-by-subgroup interaction is not significant.

- If the interaction effect is statistically significant (i.e., p-value < 0.05), then the Gail and Simon[5] test will be used to test for the qualitative interaction at a significance level of 0.05 and provided as an aid for interpretation.

---

[5] Gail MH and Simon R. Testing for qualitative interactions between treatment effects and patient subsets. Biometrics, 1985;41: 361-372.

The Repeated Measures Model analyses described in Section 8.4.2 already account for the subgroups of baseline use of gabapentin and/or pregabalin (i.e., Yes and No groups). For evaluating the interaction between treatment and baseline use of gabapentin and/or pregabalin, the interaction effect will be added to the primary analysis models. If the interaction effect is significant, then the Gail and Simon test will be used to test for the qualitative interaction at a significance level of 0.05.

### 8.4.6. Analysis of Covariates

The following covariate analyses will be conducted for the primary and secondary efficacy based on the ITT population to evaluate the treatment effect adjusted for the three covariates listed in Section 4.6:

- The primary analysis model (Section 8.4.2) adding an individual covariate of the three covariates will be used to obtain the 2-sided 95% confidence intervals for the covariate-adjusted estimate of the treatment effect and covariate effect along with their respective p-values.
- The primary analysis model (Section 8.4.2), adding all three covariates will be used to obtain the 2-sided 95% confidence intervals for the covariate-adjusted estimate of the treatment effect.

### 8.4.7. Exploratory Efficacy Endpoints

Each exploratory efficacy endpoint (described in Section 4.3, except for PGIC and Semmes-Weinstein monofilament testing) will be analyzed in a manner similar to the corresponding continuous or categorical Repeated Measures Models for the primary analyses described in Section 8.4.2 based on available data of ITT population.

For each such parameter, a repeated measures model analysis will be used accounting for all post-baseline visits where the parameter is collected. The model will include treatment, visit, baseline use of gabapentin and/or pregabalin, and treatment-by-visit interaction as fixed effects and baseline measurement of the parameter as a covariate. An unstructured variance-covariance matrix for the repeated measures model will be used unless convergence problems arise. The point estimate for the least-squares mean of the treatment difference (VM202 – Placebo) and the corresponding two-sided 95% confidence interval at each visit will be summarized.

The 7-category data for PGIC will be combined into 3-category data (1: very much improved, much improved, 0: minimally improved, no change, minimally worse, -1: much worse, very much worse). PGIC will be analyzed at each follow-up visit by Cochran-Mantel-Haenszel (CMH) test to account for the stratification variable and row means scores differ statistic (row variable: treatment group, column variable: 3-category data of PGIC) will be used.

A graph of the cumulative proportion of responders will be generated at each of the 3-, 6-, and 9-month follow-up visits to predict the likelihood of a response for each treatment group and level of the stratification variable (i.e., baseline use of gabapentin and/or pregabalin) over the indicated range of response cutoff-points (20% to 70%) for the reduction from baseline in the average 24-hour pain score.

Semmes-Weinstein monofilament testing will be conducted at 5 testing sites on each foot; each testing site will be graded based on Table 5. Shift tables of grade at baseline versus grade at follow-up (i.e., a 7-by-7 contingency table) will be provided to summarize changes for each testing site by treatment and stratification factor at each follow-up visit. A proportional-odds cumulative logit model with categorical terms for treatment, stratification factor, and baseline grade will be applied for each testing site at each follow-up visit based on the observed data in the ITT population. Multivariate cumulative logits to account for the within subject correlation between testing sites may be considered.

**TABLE 5. Grading Scale for Semmes-Weinstein Monofilament Testing**

| Force (gms) | Interpretation | Grade Points |
|---|---|---|
| 0.07 | Normal | 6 |
| 0.4 | Diminished Light Touch | 5 |
| 2.0 | Diminished Protective Sensation | 4 |
| 4.0 | Mild Loss of Protective Sensation | 3 |
| 10 | Moderate Loss of Protective Sensation | 2 |
| 300 | Deep Pressure Sensation Only | 1 |
| >300 | No Sensation | 0 |

### 8.5. Safety Analyses

Safety analyses in this study will evaluate the safety profile of VM202, 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 subjects in the Safety population will be included in these analyses. Subjects 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 and total volume administered per calf) will be summarized by treatment group 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 subjects 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 subjects 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 subjects who receive an injection at a given one of these visits will be counted in the post-injection assessment results for that visit. Subjects 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 subjects experiencing a particular event, the percentage of subjects 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 subject summarization, a subject is classified according to the highest severity if the subject 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 subject summarization, a subject is classified according to the closest relationship if the subject 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. Specific areas of special interest to be presented separately are for dizziness, somnolence, weight increase, peripheral edema, and falls as well as subsets of events within the Nervous System disorders, Investigations, and Injury, Poisoning and Procedural Complications system organ classes. The specific preferred terms within these subsets will be determined at the blinded data review.
- 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. Vital Signs

Vital signs and change from baseline will be summarized descriptively at each visit by treatment group.

### 8.5.5. HbA1c, Serum Chemistry and Hematology

Shift tables (i.e., normal, abnormal not-clinically significant, abnormal clinically significant at baseline versus normal, abnormal not-clinically significant, abnormal clinically significant 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. Determinations of clinical significance will be made by the individual study centers based on their laboratory normal ranges. The counts and percentage of subjects 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.

### 8.5.6. Prior and Concomitant Medications of Interest

Prior medications are those medications taken within 60 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 preferred drug names using WHODrug Global B3 (March 2018 version or later). Prior and concomitant medications of interest will be determined by the CDRC and will be summarized separately for each treatment group by preferred names. These summaries will present the number and percentage of subjects using each medication.

#### 8.5.7. Retinal Fundoscopy

Retinal fundoscopy findings in each eye (presence or absence of proliferative retinopathy, other finding) at screening (baseline), 6-month, and 9- month follow-up visits and any changes from the baseline at the follow-up visits will be summarized descriptively by treatment group.

### 8.6. Other Clinical Parameters

#### 8.6.1. Nerve Conduction

All nerve conduction data will be analyzed directly from adjudicated datasets provided by the Central Reading site at the Laboratory for Behavioral Neurophysiology, Albert Einstein College of Medicine in Bronx, New York.

All available data will be summarized descriptively for each parameter (sural nerve amplitude and conduction velocity, peroneal motor nerve amplitude and conduction velocity) by treatment group and scheduled visit based on the actual treatment.

A categorical Repeated Measures Model will be used for comparing the percentage of subjects with a ≥7% and ≥12% change in Nerve Conduction Velocity for both the peroneal nerve and the sural nerve at 9 months (responder rate) between the two study treatment groups. The model will include treatment, visit, treatment-by-visit interaction, baseline use of gabapentin and/or pregabalin, and baseline average 24-hour pain scores as covariate with an unstructured variance-covariance matrix. The point estimates for the least-squares mean of the treatment difference (VM202 – Placebo) at 9 months and the corresponding 95% confidence interval and p-value will be summarized.

#### 8.6.2. Total Tylenol (Rescue Medication) Dose

The number and percentage of subjects taking Tylenol during the 9-month follow-up will be calculated by the treatment group. The mean, standard deviation, minimum, maximum, median, and 1$^{st}$ and 3$^{rd}$ quartiles of days to the first start date of the Tylenol will be summarized by the treatment group using observed data; subjects not taking Tylenol will not be included in these descriptive statistics. The total dose of Tylenol of each subject during the 9-month follow-up will be summarized by the treatment group. These analyses will be based on the PP Dosing population.

### 8.6.3. Pharmacokinetics

HGF serum levels and the number of copies of VM202 in whole blood will be analyzed by an independent lab designated by study sponsor. HGF and VM202 data will be listed and summarized at each collection time point by treatment group. All values that are below the identified limit of quantitation will be set to 0 for the summary tables.

## 9. REFERENCES

1. FDA Guidance for Industry (1998). E9 Statistical Principles for Clinical Trials (ICH). U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Available at: www.fda.gov/media/71336/download. Accessed July 15, 2019.