Statistical Analysis Plan

A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Proof of Concept Study of the Efficacy and Safety of Fremanezumab for Treatment of Patients with Fibromyalgia

Study Number TV48125-PN-20028

NCT03965091

SAP Approval Date: 8 February 2022

Statistical Analysis Plan

Study TV48125-PN-20028 with Protocol Amendment 05

A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Proof of Concept Study of the Efficacy and Safety of Fremanezumab for Treatment of Patients with Fibromyalgia

Phase 2

IND number: 141519; NDA number: Not Applicable; EudraCT number: Not Applicable

Approval Date: 8th February 2022

Sponsor

Teva Branded Pharmaceutical Products R&D, Inc.
145 Brandywine Parkway
West Chester, Pennsylvania 19380
United States of America

Prepared by: [redacted], [redacted]
Teva Global Statistics

STATISTICAL ANALYSIS PLAN APPROVAL

Study No.: **TV48125-PN-20028**

Study Title: **A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Proof of Concept Study of the Efficacy and Safety of Fremanezumab for Treatment of Patients with Fibromyalgia**

Statistical Analysis Plan for:

- ☐ **Interim Analysis**
- ☒ **Final Analysis**
- ☐ **Integrated Summary of Efficacy**
- ☐ **Integrated Summary of Safety**

Amendment: **NA**

Author: [redacted]
[redacted] **Global Statistics**

DocuSigned by:
[redacted]
08-Feb-2022 | 18:06 GMT
Signer Name: [redacted]
Signing Reason: I approve this document
Approver: [redacted] 2022 | 18:06 GMT **Date**
[redacted]
Global Statistics

DocuSigned by:
[redacted]
08-Feb-2022 | 19:00 GMT
Signer Name: [redacted]
Signing Reason: I approve this document
Approver: [redacted] 2 | 19:00 GMT **Date**
[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
ADA	anti-drug antibody
AE	adverse event
AESI	adverse events of special interest
AR	autoregressive
ARH	heterogeneous autoregressive correlation
BAPS	Baseline Assessment Period Start
CGRP	calcitonin gene-related peptide
CI	confidence interval
CRF	Case report form
CS	compound symmetry
CSC	clinical study center
CSH	heterogeneous compound symmetry
CSR	Clinical Study Report
C-SSRS	Columbia-Suicide Severity Rating Scale
CV	coefficient of variation
ECG	Electrocardiogram/Electrocardiography
EOS	End of study
EOT	End of treatment
ET	Early termination
EQ VAS	EQ Visual Analogue Scale
EQ-5D-5L	EuroQol-5D-5L
EudraCT	European Clinical Trials
FM	Fibromyalgia
FIQR	Fibromyalgia Impact Questionnaire Revised
GEE	generalized estimating equation
HADS	Hospital Anxiety and Depression Scale
IMD	Inventory of Medical Diagnoses
IMP	the investigational medicinal product
ITT	Intent-to-Treat
LOE	lack of efficacy
LS	Least square
MedDRA	Medical Dictionary for Regulatory Activities

Abbreviation	Term
MI	multiple imputation
mITT	Modified Intent-to-Treat
MMRM	Mixed model for repeated measures
PD	Protocol deviations
PGIC	Patient Global Impression of Change
PI-NRS	Pain Intensity-Numerical Rating Scale
POMS-2A Short	Profile of Mood States 2 – Adult Short Form
PP	Per-Protocol
PROMIS	Patient-Reported Outcomes Measurement Information System
RTSM	Randomization and Trial Supply Management
SAP	Statistical Analysis Plan
SD	Standard Deviation
SE	Standard Error
SF	Short form
GPSP	Global Patient Safety and Pharmacovigilance
SI	standard international
SOC	system organ class
SS	symptom severity
ULN	upper limit of normal
WHO	World Health Organization

INTRODUCTION

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for Teva Branded Pharmaceutical Products R&D, Inc. study TV48125-PN-20028, (A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Proof of Concept Study of the Efficacy and Safety of Fremanezumab for Treatment of Patients with Fibromyalgia), and was written in accordance with SOP GBP_RD_702 (Statistical Analysis Plan).

The reader of this SAP is encouraged to read the study protocol for details on the conduct of this study, the operational aspects of clinical assessments, and the timing for completing the participation of a patient in this study.

The SAP is intended to be in agreement with the protocol, especially with regards to the primary and all secondary endpoints and their respective analyses. However, the SAP may contain more details regarding these particular points of interest, or other types of analyses (e.g. other endpoints). When differences exist in descriptions or explanations provided in the study protocol and this SAP, the SAP prevails; the differences will be explained in the Clinical Study Report (CSR).

1. STUDY OBJECTIVES AND ENDPOINTS

1.1. Primary and Secondary Study Objectives and Endpoints

The primary and secondary study objectives and endpoints are as follows:

Objectives	Endpoints
The **primary objective** of the study is to estimate the treatment effect of fremanezumab administered subcutaneously in reducing pain in adult patients with Fibromyalgia (FM).	The primary efficacy endpoint is the change from baseline to week 12 in the weekly average of the daily average PI-NRS score over the past 24 hours.
A **secondary objective** is to evaluate the effect of fremanezumab on other efficacy measures, including pain, quality of life, sleep, fatigue, improvement in health, physical functioning, and mood.	The secondary efficacy endpoints are as follows: • change from baseline to week 12 in the individual components of the Fibromyalgia Impact Questionnaire Revised (FIQR): symptom subscore, impact subscore, and functional subscore score • responder rate of the Patient Global Impression of Change (PGIC) rating (percentage of patients much improved or very much improved) at week 12 • the percentage of patients who experience a ≥30% reduction in the weekly average of the daily average PI-NRS score at week 12 • the percentage of patients who experience a ≥50% reduction in the weekly average of the daily average PI-NRS score at week 12 • change from baseline to week 12 in the weekly average of the daily worst PI-NRS score over the past 24 hours • change from baseline to week 12 in the Patient-Reported Outcomes Measurement Information System (PROMIS) Sleep Disturbance Short Form (SF) 8a score • change from baseline to week 12 in the PROMIS Physical Function SF12a score • change from baseline to week 12 in the PROMIS Fatigue SF8a score

Objectives	Endpoints
	• number (%) of patients who did not complete the treatment (Kaplan-Meier Survival Analysis) due to lack of efficacy
A **secondary objective** of the study is to evaluate the safety and tolerability of fremanezumab administered subcutaneously in adult patients with FM.	The safety and tolerability endpoints are as follows: • occurrence of adverse events • change from randomization to week 12 in the clinical laboratory tests (serum chemistry, hematology, and urinalysis) • change from baseline to week 12 in vital signs (pulse, respiratory rate, systolic and diastolic blood pressure, and oral body temperature measurements at each visit) • clinically significant changes in physical examination, including body weight • abnormal standard 12-lead electrocardiogram (ECG) findings • tolerability at the injection site including but not limited to pain, erythema, induration, and ecchymosis • occurrence of hypersensitivity/anaphylaxis reactions using standardized criteria (Appendix C of the study protocol) • suicidal ideation and behavior as measured by the Columbia-Suicide Severity Rating Scale (C-SSRS) • number (%) of patients who did not complete the treatment (Kaplan-Meier Survival Analysis) due to adverse events

1.2. [redacted]

[redacted]

1.3. Immunogenicity Objective and Endpoint

The immunogenicity objective and endpoint are as follows:

Objectives	Endpoints
To evaluate the immunogenicity of fremanezumab and the impact of anti-drug antibodies (ADAs) on clinical outcomes	• incidence and characteristics of ADAs (eg, titers, kinetics, and neutralizing activities)

1.4. Estimand for Primary Efficacy Endpoint

The combination of treatment policy and hypothetical strategy will be used for the primary efficacy endpoint to account for the intercurrent events. The four attributes of the estimand are defined as follows:

A. Population: The patients with fibromyalgia met all inclusion criteria and met none of the exclusion criteria.

The modified intent-to-treat (mITT) analysis set defined in Section 3.2 will be the primary analysis set for efficacy. In the data analyses, the treatment will be assigned based on the treatment to which patients were randomized, regardless of which treatment they actually received.

B. Variable: Change from baseline to week 12 in the weekly average of the daily average PI-NRS score over the past 24 hours.

C. Intercurrent events: Treatment discontinuation due to an adverse event, due to lack of efficacy, or due to other reasons.

For those who discontinue the treatment before week 12, no further efficacy data will be collected. The use of rescue medication for fibromyalgia pain will not be considered as an intercurrent event.

D. Population-level summary: The primary efficacy endpoint, the change from baseline to week 12 in the PI-NRS scores, will be analyzed using a Mixed Model Repeated Measures (MMRM) model with change from baseline in the weekly average of the daily average PI-NRS score over the past 24 hours at weeks 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12 as the dependent variable, sex, age group at FM onset, week, treatment, and treatment by week interaction as fixed factors, and baseline average of the daily average PI-NRS score as a covariate. An adjustment for missing data due to adverse event or due to lack of efficacy on the PI-NRS scores will be on average return to baseline values, while due to all other reasons will be imputed assuming missing at random. With missing data imputed, each of the 500 imputed data sets will be analyzed using the MMRM model. The results from the 500 analyses will be combined using Rubin's formulae. The least square (LS) mean estimates for the mean change from baseline to each time point, as well as the difference of the estimates between the fremanezumab and placebo group, with the corresponding standard error (SE), p-value and associated 95% confidence interval (CI) will be provided.

Sensitivity Analysis: The primary analysis will be repeated; however, the daily average pain scores measured on the days where the patient took rescue medication will be replaced with his or her PI-NRS score at the time of rescue medication use before the weekly average PI-NRS scores are derived. On days where a patient uses rescue medication more than once, the mean PI-NRS score at the time of rescue mediation use for the day will be used as the pain score in place of the average pain intensity score for that day.

2. STUDY DESIGN

2.1. General Design

This is a 21-week, multicenter, randomized, double-blind, placebo-controlled, parallel-group study to evaluate the efficacy and safety of fremanezumab in adult patients with FM. Fremanezumab 225 mg, fremanezumab 675 mg, or placebo will be administered sc for 4 monthly doses. The study will consist of a 17- to 35-day screening period (Visit 1, phone contact, a Baseline Assessment Period Start [BAPS] Visit [Visit 2]), and a 16-week double-blind treatment period (Visits 3, 4, 5, 6, and 7). Patients enrolled in the study will be male or female patients with FM with an average of the daily average PI-NRS score of ≥4 and ≤9 during the 14-day baseline assessment period. Patients will be required to wash out of all prohibited concomitant medication(s) prior to the BAPS Visit (Visit 2).

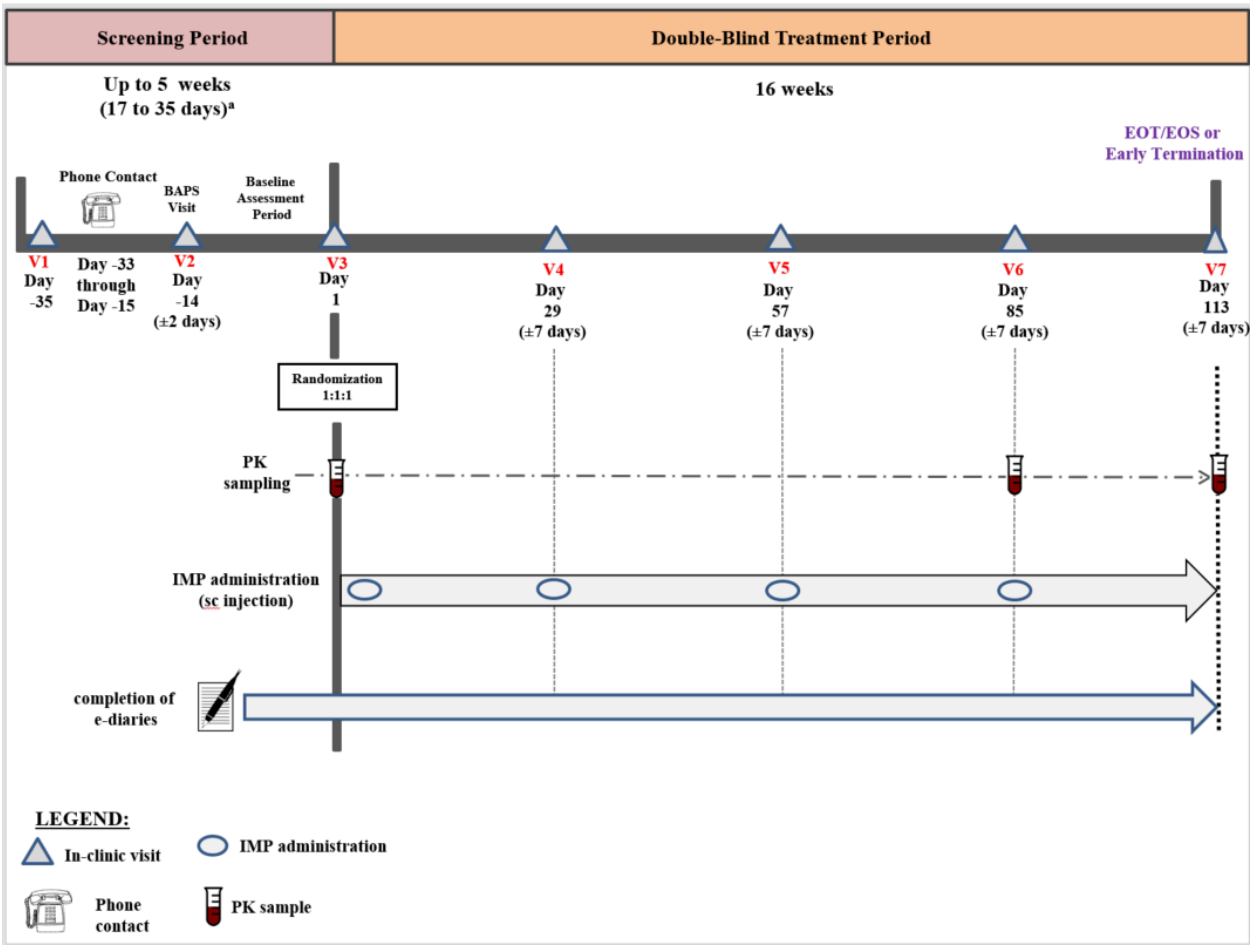
Therefore, the total duration of patient participation in the study is planned to be 21 weeks, consisting of the screening period of up to 5 weeks (ranging from 17 to 35 days) and the double-blind treatment period of 16 weeks. The end-of-study (EOS) is defined as completion of the last visit of the last patient.

Note, there was the 14-week follow-up period prior to Study TV48125-PN-20028 protocol amendment 04.

The study schematic diagram is presented in Figure 1.

Study procedures and assessments with their timing are summarized in Table 1 of the study protocol.

Figure 1: Overall Study Schematic Diagram



[a] The screening period is up to 5 weeks and can range from 17 to 35 days. The duration of the washout period will be 7 to 19 days; suggested washout periods for commonly used analgesics are listed in the study protocol.
Note: For Visits 4 and 5 (when at-home dosing is available), home visits are permitted for patients who are not able to go to the investigational center or if the investigational center staff are not able to see patients at the investigational center.
BAPS = Baseline Assessment Period Start; COVID-19 = Coronavirus Disease 2019; e-diary = electronic diary; EOS = end-of-study; EOT = end-of-treatment; PK = pharmacokinetic; sc = subcutaneous; V = visit.

The follow-up period (14 weeks) in the protocol was removed in Study TV48125-PN-20028 protocol amendment 04. The data collected in the follow-up period will be listed.

2.2. Randomization and Blinding

This is a randomized, double-blind, placebo-controlled study. All the investigational medicinal products (IMPs) will be identical and indistinguishable. Patients, investigators, and all clinical study center (CSC) staff will remain blinded to treatment assignment during the study. Eligible patients will be randomly assigned via a qualified Randomization and Trial Supply Management (RTSM) system in a 1:1:1 ratio to fremanezumab at 225 mg sc or fremanezumab 675 mg sc, or placebo sc. Randomization will be stratified by sex (male/female) and age at FM onset (<40 years old and ≥40 years old).

Patients will be randomly assigned to treatment groups by means of a computer-generated randomization list. The generation of the randomization list and management of the RTSM system will be done by a qualified service provider. The specifications for randomization will be under the responsibility and oversight of Teva Global Statistics.

The sponsor's clinical personnel (and delegates) involved in the study will be blinded to the identity of the IMPs until the database is locked for analysis and the IMP assignment is released. However, if a prioritized sample analysis is needed, bioanalytical and clinical pharmacology personnel may be unblinded.

In the event of an emergency, it will be possible to determine to which treatment group and dose a patient has been allocated by accessing the RTSM system. All investigational centers will be provided with details of how to access the system for code breaking at the start of the study. The Medical Monitor or equivalent should be notified, if possible, prior to unblinding or following unblinding. Any unblinding of the IMP performed by the investigator must be recorded in the source documents.

2.3. Data Monitoring Committee

There will be no Independent Data Monitoring Committee for this study.

2.4. Sample Size and Power Considerations

The sample size of 80 subjects per treatment group was selected based on expert consensus agreement at a scientific advisory board meeting; this sample size is expected to provide sufficient precision to estimate the treatment effect of fremanezumab versus placebo. Moreover, assuming a standard deviation (SD) of 2.2 for change from baseline in the weekly average of the daily average PI-NRS score, 80 patients per treatment group will provide at least 90% probability to observe the half length of the 95% confidence interval (CI) of the treatment difference between fremanezumab 225 mg sc or fremanezumab 675 mg sc and placebo less than 0.74. In total, 240 patients will be enrolled in this study in a 1:1:1 randomization ratio.

2.5. Sequence of Planned Analyses

2.5.1. Planned Interim Analyses

There were two interim analyses for futility performed in this study. The details of the interim analyses were specified in seperate interim statistical analysis plans. To ensure study integrity, the interim analyses were performed by an independent third party; the study team were kept blinded. The decision was made to end this study for futility following the results of the second interim analysis.

2.5.2. Final Analyses and Reporting

This study has been early terminated due to futility. All analyses identified in this SAP will be performed after study closure. This SAP and any corresponding amendments will be approved before database lock, in accordance to SOP GBP_RD_702 (Statistical Analysis Plan).

The randomization codes will not be unblinded until this SAP has been approved, the database is locked, and the analysis populations are determined.

[redacted]

3. ANALYSIS SETS

3.1. Intent-to-Treat Analysis Set

The intent-to-treat (ITT) analysis set will include all randomized patients.

In the ITT analysis set, treatment will be assigned based on the treatment to which patients were randomized, regardless of which treatment they actually received.

Demographics and baseline characteristics will use the ITT analysis set.

3.2. Modified Intent-to-Treat Analysis Set

The modified intent-to-treat (mITT) analysis set is a subset of the ITT analysis set including only patients who received at least 1 dose of study drug and at least 1 post-baseline entry of daily average pain intensity using the PI-NRS.

In the mITT analysis set, treatment will be assigned based on the treatment to which patients were randomized, regardless of which treatment they actually received.

All efficacy analyses will use the mITT analysis set unless noted otherwise.

3.3. Safety Analysis Set

The safety analysis set will include all randomized patients who received at least 1 dose of study drug. In the safety analysis set, treatment will be assigned based on the treatment patients actually received, regardless of the treatment to which they were randomized.

All safety analyses will use the safety analysis set unless noted otherwise.

3.4. Per-Protocol Analysis Set

The per-protocol (PP) analysis set is a subset of the mITT analysis set including only patients without major protocol violations which will be determined by data review and documented before study unblinding. The PP analysis set will serve as supportive for the efficacy analyses.

In the PP analysis set, treatment will be assigned based on the treatment patients actually received, regardless of the treatment to which they were randomized.

The primary efficacy endpoint and selected secondary efficacy endpoints will be analyzed using PP analysis set.

4. GENERAL ISSUES FOR DATA ANALYSIS

4.1. General

Descriptive statistics for continuous variables include n, mean, SD, standard error (SE), median, minimum, and maximum. In addition, for fremanezumab concentration, percentage coefficient of variation (%CV) and geometric mean will also be calculated. Descriptive statistics for categorical variables include patient counts and percentages, missing category will be displayed as appropriate.

Summaries of potentially clinically significant abnormal values will include all postbaseline values (including scheduled, unscheduled, and early termination visits).

For by-visit analyses, if there are multiple assessments at a scheduled postbaseline visit then the last non-missing scheduled assessment at that visit will be used for the summary. If a visit has scheduled and unscheduled assessments, the scheduled assessment will be used for the analysis. If a visit only has unscheduled assessments, the unscheduled assessments will be used for analyses; if there are more than one unscheduled visits, the assessment from the last unscheduled visit will be used for analyses.

4.2. Specification of Baseline Values

The baseline values are defined as the last non-missing assessments before the first dose of IMP unless otherwise specified.

Baseline value for the average of the daily average PI-NRS pain score is defined as the average of the daily average PI-NRS score over the last 14 days before the first dose of IMP drug (day -14 to day -1) and will be calculated using formula [1].

Baseline value for the average of the daily worst PI-NRS pain score is defined as the average of the daily worst PI-NRS score over the last 14 days before the first dose of IMP (day -14 to day -1) and will be calculated using formula [1].

The baseline values for the following variables are measured on the day of the first study drug (day 1) before the first dose of study drug.

- individual components of the FIQR: symptom subscore, impact subscore, and functional subscore score
- PROMIS Physical Function SF12a score
- PROMIS Fatigue SF8a score
- HADS score
- EQ-5D-5L
- POMS-2A Short Form score

The baseline value for IMD assessment rating is measured at the screening visit.

PROMIS Sleep Disturbance SF 8a score are collected on e-diary weekly during BAPS. The baseline PROMIS Sleep Disturbance SF 8a scores will be the mean of the weekly scores recorded during last 14 days before the first dose of IMP drug (day -14 to day -1).

4.3. Handling Withdrawals and Missing Data

Missing data for the primary efficacy endpoint, efficacy endpoint for sensitivity analysis, and selected secondary efficacy endpoints will be imputed with the multiple imputation (MI) method described in Section 6.2.2. All other endpoints will be not be imputed for missing data unless otherwise specified.

Dates that have incomplete information (the month and year or just the year is available) will be estimated for the purpose of calculating variables that are dependent on time if necessary. Day will be estimated as the first day (01) of the month (if month and year of partial date are available) or middle (July 1) of the year (if only year is available), unless otherwise noted. The imputations for partial dates are only for calculation purpose. Original date variables will not be modified. Listings will list dates as collected.

4.4. Study Days and Visits

Study visits are detailed in Table 1.

Table 1: Study Visits

Study period	Screening Period (Up to 5 weeks [17 to 35 days])			Double-Blind Treatment Period (16 weeks)				
	Screening Visit	Phone Contact (Washout)	Baseline Assessment Period	Randomization (Dose 1)	Dose 2	Dose 3	Dose 4	
Visit	V1	Phone Contact	V2 (BAPS Visit)	V3	V4	V5	V6	V7 (EOT/ EOS or ET)
Day and allowed time windows	Week-5 Day -35	Day -33 through Day -15	Day -14 (±2 days)	Day 1	Day 29 ±7 days	Day 57 ±7 days	Day 85 ±7 days	Day 113 ±7 days

BAPS=Baseline Assessment Period Start; EOT=End of treatment; ET=Early termination; EOS=End of Study.

There was a follow-up period (14 weeks) in the Study TV48125-PN-20028 protocol prior to amendment 04. The data collected in the follow-up period will be listed.

Study days will be numbered relative to the first day of IMP administration. The start of treatment (Day 1 at Visit 3) is defined as the date on which a patient takes the first dose of IMP, as recorded on the study drug administration CRF. Days will be numbered relative to treatment start (ie, ..., –2, –1, 1, 2, ...; with day 1 being the first day of IMP administration and day –1 being the day before the first day of IMP administration).

'Last Assessment' will be derived for analysis purposes and is defined as the last observed postbaseline data, excluding the follow-up visit. For by-visit summaries, data from the ET visit will be included as one of the visits indicated in the database.

For data collected daily on e-diary, analysis weeks (week 1, week 2, …, etc.) will be derived for the purpose of efficacy endpoint analyses. Details are provided in Section 6.1.2.

Data collected after week 12 will be included in descriptive summaries but will not be included in any inferential analyses.

5. STUDY POPULATION

5.1. General

Patients enrolled in the study will be male and female patients, ≥18 to 75 years old, inclusive, with FM and an average of the daily average PI-NRS score of ≥4 and ≤9 during the last 14 days of the 14-day baseline assessment period.

For continuous variables, descriptive statistics (number [n], mean, SD, SE, median, minimum, and maximum) will be provided. For categorical variables, patient counts and percentages will be provided. Categories for missing data will be presented, if necessary.

The ITT analysis set will be used for all study population summaries unless otherwise specified. Summaries will be presented by treatment group and for all patients.

5.2. Patient Disposition

Descriptive statistics will be used to summarize data from the following groups: patients screened; patients screened but not randomized and reason not randomized; patients who are randomized; patients randomized but not treated (and reason); patients in the ITT, mITT, safety, and PP analysis sets; patients who complete the treatment and the study; and patients who discontinue from the treatment and the study. Data from patients who discontinue from treatment and the study will also be summarized by reason for discontinuation using descriptive statistics.

5.3. Demographics and Baseline Characteristics

Patient demographic characteristics (age, sex, ethnicity, and race) and baseline characteristics will be examined to assess the comparability of the treatment groups and will be summarized using descriptive statistics. No inferential analyses will be performed. Demographic and baseline characteristics will be summarized using ITT analysis set.

Baseline characteristics include

- age at time of FM onset or diagnosis of FM
- month since FM onset
- month since FM diagnosis
- alcohol history
- psychiatric history
- ECG findings at screening

Month will be calculated as (date of informed consent - the date of the event + 1)/30.42. Rules for handling partial dates are in Section 4.3.

5.4. Medical History

All medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The incidence of medical history abnormalities will be summarized using descriptive statistics by system organ class (SOC) and preferred term. Patients are counted only

once in each preferred term and SOC category. Summaries will be presented by treatment group and for all patients.

5.5. Prior Therapy and Medication

Any prior therapy or medication will be coded using the World Health Organization (WHO) Drug Dictionary. The incidence of prior therapies and medications, past fibromyalgia medication will be summarized using descriptive statistics by therapeutic class and preferred term. Patients are counted only once in each therapeutic class category, and only once in each preferred term category. Prior therapies and medications will include all medications taken and therapies administered before the first day of study drug administration.

Past fibromyalgia medication will also be summarized separately.

5.6. Childbearing Potential and Methods of Contraception

For female patients, information related to childbearing potential and menopause will be collected at the screening visit. Data will be listed.

For female and male patients, methods of contraception will be collected at the screening visit. Data will be listed.

5.7. Study Protocol Deviations

All protocol deviations (PD) will be collected and entered in an excel spreadsheet. All PDs will be reviewed and approved by the study team prior to database lock. In addition, the important PDs will be also determined by study team before database lock. The important PDs will be summarized for overall and for each category using descriptive statistics and all PDs will be listed in a data listing.

6. EFFICACY ANALYSIS

6.1. General

The mITT analysis set will be the main analysis set for all efficacy analyses. The PP analysis set will serve as supportive analysis set for analyses of the primary efficacy endpoint and following selected secondary efficacy endpoints:

- responder rate of PGIC rating (percentage of patients much improved or very much improved) at week 12
- change from baseline to week 12 in the weekly average of the daily worst PI-NRS score over the past 24 hours
- the percentage of patients who experience a ≥30% reduction in the weekly average of the daily average PI-NRS score at week 12
- the percentage of patients who experience a ≥50% reduction in the weekly average of the daily average PI-NRS score at week 12

For efficacy summaries and analyses using the last visit, this last visit should not include the follow-up visit which was scheduled in Study TV48125-PN-20028 prior to protocol amendment 04. In addition, data collected after week 12 will be included in descriptive summaries but will not be included in any inferential analyses.

6.1.1. Pain Intensity-Numerical Rating Scale and Rescue Medication Use

The PI-NRS is an 11-point pain intensity numerical rating scale where 0=no pain and 10=worst possible pain. The PI-NRS will be used to rate average FM pain intensity over the past week at screening via the site tablet at the screening visit. From Visit 2 to Visit 7 the PI-NRS will be recorded in the patient's e-diary every evening (Evening Report) between 1800 and 2200 hours for the following:

- The average FM pain intensity over the past 24 hours
- The worst FM pain intensity over the past 24 hours

If a patient cannot complete or forgets to complete the Evening Report, a Make-up Evening Report will become available for completion between 0300h and 1300h the next day. In case both the Evening Report and the Make-up Evening Report are completed, PI-NRS collected on the Evening Report will be used for analyses.

The Rescue Medication tab on the e-dairy will be available to patients 24 hours per day. Patients will record the following in the e-diary:

- Time of rescue medication used for a specific breakthrough pain episode
- Amount (tablets) of rescue medication used for a specific breakthrough pain episode
- At-the-moment pain intensity just prior to the use of rescue medication

If a patient cannot record or forgets to record the rescue medication taken, the Evening Report will be used to record any missed rescue medication taken. A patient can take 1 to 2 caplets with dose up to 1000 mg, a maximum of 6 caplets(3000 mg) in a 24-hour period.

6.1.2. Baseline and Analysis Weeks for Data Collected Daily

Baseline is defined as the last 14 days before the first dose of IMP (day -14 to day -1). Day 1 will be defined as the day of first dose of IMP (Visit 3).

For data collected daily (eg. PI-NRS scores), analysis weeks will be derived using algorithm in Table 2. Analyses will be based on the derived analysis weeks.

Table 2: Analysis Weeks

4-Week Interval	Visit	Dose (n)	Target Injection or Visit Day	Actual Injection or Visit Day	Analysis Week	Analysis Week Interval
1[a]	3	1	1	1	1	[1, v-22]
					2	[v-21, v-15]
					3	[v-14, v-8]
					4	[v-7, v-1]
2[a]	4	2	29	v (2nd dosing day)	5	[v, x-22]
					6	[x-21, x-15]
					7	[x-14, x-8]
					8	[x-7, x-1]
3[a]	5	3	57	x (3rd dosing day)	9	[x, y-22]
					10	[y-21, y-15]
					11	[y-14, y-8]
					12	[y-7, y-1]
4[b] (last dose)	6	4	85	y (4th dosing day)	13	[y, y+6]
					14	[y+7, y+13]
					15	[y+14, y+20]
					16	[y+21,y+27]
5	5	EOT	113		-	-

[a] In 4-week intervals before the last dose, the 4th week will be 7 days before the date of the injection, and counting backward by 7-day increments to define remaining weeks. The 1st week of these 4-week intervals will include all remaining days in the interval.

[b] The weeks after the last dose, an analysis week will be defined as a 7-day interval counting onward.

6.1.3. Data Derivation for weekly average of PI-NRS

The weekly average of an efficacy variable in an analysis week will be calculated as follow:

$$\frac{\sum non-missing\ efficacy\ variable\ in\ an\ analysis\ week}{Number\ of\ days\ with\ non-missing\ efficacy\ variable\ in\ the\ analysis\ week} \quad [1]$$

Where the efficacy variable will be

- Daily average PI-NRS pain score over the past 24 hours
- Daily worst PI-NRS pain scores over the past 24 hours

If all of the data are missing within an analysis week, the average of the efficacy variable within the analysis week will be missing.

6.2. Primary Efficacy Endpoint and Analysis

6.2.1. Definition

The primary efficacy endpoint is the change from baseline to week 12 in the weekly average of the daily average PI-NRS score over the past 24 hours.

Baseline is defined as the last 14 days before the first dose of IMP (Visit 3). The analysis weeks for post-baseline will be derived using rule described in Table 2.

The baseline and post baseline weekly average of the daily average PI-NRS pain scores over the past 24 hours will be derived using formula [1], respectively.

6.2.2. Primary Efficacy Analysis

The primary efficacy endpoint will be analyzed using a MMRM model with change from baseline in the average of the daily average PI-NRS score over the past 24 hours at weeks 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12 as the dependent variable, sex, age group at FM onset, week, treatment, and treatment by week interaction as fixed factors, and baseline average of the daily average PI-NRS score as a covariate. The heterogeneous autoregressive correlation structure [ARH (1)] for repeated observations within patients will be used, and the denominator degrees of freedom will be estimated using Kenward-Roger's approximation. An adjustment for missing data due to lack of efficacy or adverse events assumes the PI-NRS scores will be on average return to baseline values. Missing data (due to early treatment discontinuation or missing e-dairy recording in an entire analysis week) will be imputed 500 times to generate 500 complete data sets using the SAS procedure PROC MI following the 3 steps below.

Notes, if ARH(1) does not converge, then simpler covariance structures with less parameters will be used in the following order until the model converges: autoregressive (1) (AR[1]); heterogeneous compound symmetry (CSH); and compound symmetry (CS).

Step 1: The monotone missing pattern will be induced by the Markov Chain Monte Carlo method in the PROC MI procedure using seed number 4751523.

The average of the daily average PI-NRS score up to and including week 12 will be imputed by using following SAS code to impute the arbitrary missing data and to make the missing data pattern monotonic:

```
PROC MI DATA=name  NIMPUTE= 500 SEED=4751523 OUT= MI_MONOTONE;
     BY TRT01PN AGEATFMONST SEX;
     MCMC IMPUTE= MONOTONE;
     VAR BASE WEEK1 … WEEK12;
RUN;
```

Where TRT01PN denotes the planned treatment group; AGEATFMONST denotes the age group at FM onset; BASE denotes the baseline average of the daily average PI-NRS score; WEEK1, . . ., WEEK12 denote the weekly average of the daily average PI-NRS score in weeks. In bold are the SAS key words.

Step 2: Once the arbitrary missing data are imputed, the following SAS code will be used to impute the monotone missing data.

The remaining missing data at subsequent weeks will be imputed using the regression method for the monotone pattern with seed number 4751523 and adjustment for covariates including baseline value and all values at preceding weeks.

```
PROC MI DATA= MI_MONOTONE NIMPUTE= 1 SEED= 4751523
OUT= MI_COMPLETE;
      CLASS TRT01PN AGEATFMONSET SEX;
      MONOTONE REG;
     VAR TRT01PN AGEATFMONSET SEX BASE WEEK1 … WEEK12;
RUN;
```

MI_COMPLETE will be transposed to a new data set (WEEKPI0) with each week as a row.

Step 3: Use by-week mean change from baseline to adjust those who discontinued treatment due to lack of efficacy (LOE) or adverse event (AE) before week 12.

The initially missing and now imputed data for patients who discontinued from the study drug due to LOE or AE will be center adjusted at the mean baseline value. That is, the final imputed score will be equal to the imputed score under missing at random minus the observed mean change from baseline score at the post baseline time point.

```
DATA WEEKPI;
      SET WEEKPI0;
      CHANGEi = AVALi - BASE;
      IF DTREAS IN ('LACK OF EFFICACY', 'ADVERSE EVENT') AND IMP= 1
      THEN DO;
          CHANGEi = CHANGEi - MWEEK_CHG;
      END;
RUN;
```

Where AVALi denotes the value of the post baseline weekly average of the daily average PI-NRS after missing data imputation; CHANGEi denotes the change from baseline after missing data imputation; MWEEK_CHG denotes the change from baseline calculated without missing data imputation; DTREAS denotes the reason for treatment discontinuation.

Step 4: Each of the 500 imputed data sets described above steps will be analyzed using the MMRM. SAS code is:

```
ODS OUTPUT LSMEANS= LS DIFFS = DIFF (WHERE = (week = _week));
PROC MIXED DATA= WEEKPI (WHERE= (WEEK>= 1));
    BY _IMPUTATION_;
    CLASS SUBJID TRT01PN WEEK AGEATFMONSET  SEX;
     MODEL CHANGEi= BASE TRT01PN AGEATFMONSET SEX WEEK
                          TRT01PN*WEEK/DDFM=KR;
     REPEATED WEEK/SUBJECT= SUBJID TYPE= ARH(1);
    LSMEANS TRT01PN*WEEK / DIFF CL;
RUN;

ODS OUTPUT CLOSE;
```

Where _IMPUTATION_ denotes the numbering of imputations; SUBJID denotes the subject ID; WEEK denotes the analysis week (12 levels: weeks 1 to 12).

Step 5: The results from the 500 analyses from Step 4 will be combined using Rubin's formulae (PROC MIANALYZE). The least square (LS) mean estimates for the mean change from baseline to each time point, as well as the difference of the estimates between the fremanezumab and placebo group, with the corresponding SE, p-value and associated 95% CI will be provided.

(a) For LS means

```
PROC SORT DATA= LS;
    BY WEEK TRT01PN _IMPUTATION_;
RUN;

ODS OUTPUT PARAMETERESTIMATES= LSM_O;
PROC MIANALYZE DATA= LS;
    BY WEEK TRT01PN;
    MODELEFFECTS ESTIMATE;
    STDERR STDERR;
RUN;
ODS OUTPUT CLOSE;
```

Where ESTIMATE and STDERR are LS mean and standard error from Step 4.

(b) For treatment difference, treatment difference CI and p-values

```
PROC SORT DATA= DIFF;
    BY WEEK TRT01PN _TRT01PN _IMPUTATION_;
RUN;

ODS OUTPUT PARAMETERESTIMATES= DIFF_O;

PROC MIANALYZE DATA= DIFF;
     BY WEEK TRT01PN _TRT01PN;
     MODELEFFECTS ESTIMATE;
     STDERR STDERR;
RUN;
ODS OUTPUT CLOSE;
```

The observed data (ie, without imputation for missing data) will be summarized using descriptive statistics by analysis weeks (baseline, weeks 1 to 16).

6.2.3. Sensitivity Analysis

Sensitivity analyses will be performed to address the assumptions in the primary model. The primary analysis will be repeated; however, the daily average PI-NRS scores measured on the days where the patient took rescue medication will be replaced as follows:

- If a patient took rescue medication once on a day, the PI-NRS score immediately before using rescue medication will be used as the daily average PI-NRS score for that day.

- If a patient took rescue medication more than once on a day, the mean value of PI-NRS scores immediately before using rescue mediation uses for the day will be used as the daily average PI-NRS score for that day.

Note, if the PI-NRS score immediately before using rescue medication is not recorded, the daily average PI-NRS scores will not be replaced.

After the replacement, the baseline and post baseline weekly average of the daily average PI-NRS pain scores over the past 24 hours will be derived using formula [1], respectively. The data will be analyzed in the manner analogues to the primary endpoint as described in Section 6.2.2.

6.2.4. Sub-Group Analyses

Sub-group analyses will be performed on the primary and selected secondary efficacy endpoints for sex and age group at FM onset (<40 vs ≥40 years) based on observed values only using descriptive statistics. The selected secondary endpoints are:

- Change from baseline to week 12 in the individual components of the Fibromyalgia Impact Questionnaire Revised (FIQR): symptom subscore, impact subscore, and functional subscore score
- Patient's Global Impression of Change (PGIC) by visit
- Change from baseline in the weekly average of the daily worst pain PI-NRS score by analysis week

Sub-group analysis will be based on mITT and/or PP analysis sets.

In addition, the primary endpoint will also be summarized by the following subgroups using descriptive statistics:

- PI-NRS BL ≥5/<5,
- PI-NRS BL ≥6/<6;
- PI-NRS BL ≥7/<7;
- Enrollment prior to/post Study TV48125-PN-20028 protocol amendment 05;
- Enrollment prior to/post start of Covid-19 pandemic using the date of 26 Mar 2020 (site contact date) to represent the start of the pandemic;
- NSAIDs/no NSAIDs use at any time during the study (with NSAIDs to be defined by clinical review)

6.3. Secondary Efficacy Endpoints and Analysis

The secondary efficacy endpoints are listed in Section 1.1. Analysis will be based on mITT and PP analysis sets.

Missing data for the following secondary endpoints will be imputed using the MI method detailed in Section 6.2.2:

- the percentage of patients who experience a ≥30% reduction in the weekly average of the daily average PI-NRS score at week 12
- the percentage of patients who experience a ≥50% reduction in the weekly average of the daily average PI-NRS score at week 12
- change from baseline to week 12 in the weekly average of the daily worst PI-NRS score over the past 24 hours

There will be no imputation for missing data for all other secondary efficacy endpoints unless noted otherwise.

6.3.1. Change From Baseline to Week 12 in the Individual Components of the Fibromyalgia Impact Questionnaire Revised: Symptom Subscore, Impact Subscore, and Functional Subscore Score

The FIQR is a commonly used instrument in the evaluation of FM patients. It contains 21 questions in 3 domains: function (9 questions), overall impact (2 questions), and symptoms (10 questions). Questions are graded on a 0 to 10 numeric scale with 10 being the worst. All questions are framed in the context of the **last 7 days**. The sub-total scores (subscores) and total scores will be calculated by ICON using the scoring guidance [The Revised Fibromyalgia Impact Questionnaire (FIQR): validation and psychometric properties] and will be provided in the database. The sub-total score for each domain is the summation of scores in the domain. The function sub-total score ranges from 0 to 90; overall impact sub-total score ranges from 0 to 20; the symptoms sub-total score ranges from 0 to 100. According to the scoring guidance, the FIQR total score is calculated as (the function sub-total score divided by 3) + the overall impact sub-total score + (the symptoms sub-total score divided by 2). Thus FIQR total score ranges from 0 to 100.

The FIQR will be assessed at BAPS visit and visits 3 to 7 (weeks 1, 4, 8, 12, and 16).

The change from baseline in FIQR symptom sub-score, impact sub-score, functional sub-score, and total score will be analyzed, respectively, using a MMRM method with change from baseline in the FIQR score at weeks 4, 8, and 12 as the dependent variable, sex, age group at FM onset, week, treatment, and treatment by week interaction as fixed factors, and the baseline FIQR score as a covariate. ARH(1) for repeated observations within patients will be used, and the denominator degrees of freedom will be estimated using Kenward-Roger's approximation. Missing total score will be considered as missing at random.

Notes, if ARH(1) does not converge, then simpler covariance structures with less parameters will be used in the following order until the model converges: AR[1]; CSH; and CS.

The LS mean estimates for the mean change from baseline as well as the difference of the estimates between the fremanezumab and placebo group, with the corresponding SE, p-value and associated 95% CI will be provided. The results from all visits will be presented.

SAS codes are:

ODS OUTPUT LSMEANS= LS DIFFS = DIFF (WHERE = (AVISIT = _AVISIT));
PROC MIXED DATA= *name*;

CLASS SUBJID TRT01PN AVISIT AGEATFMONSET SEX;

```
    MODEL CHANGE= BASE TRT01PN AGEATFMONSET SEX AVISIT
                  TRT01PN*AVISIT/DDFM=KR;
    REPEATED AVISIT/SUBJECT= SUBJID TYPE= ARH(1);
   LSMEANS TRT01PN*AVISIT / DIFF CL;
RUN;

ODS OUTPUT CLOSE;
```

Where AVISIT denotes the study visit/week; BASE denotes the baseline value for the FIQR scores.

In addition, the FIOR scores will be summarized using descriptive statistics for all visits.

6.3.2. The Patient Global Impression of Change Rating at Week 12

The PGIC scale evaluates all aspects of patients' health and assesses if there has been an improvement or decline in clinical status since the start of the study. Patients will record responses to the PGIC scale visits 6 and 7 (weeks 12 and 16). Improvement is recorded on a 7-point scale, with 1 indicating very much improved, 4 indicating no change, and 7 indicating very much worse:

- 1 = Very Much Improved
- 2 = Much Improved
- 3 = Minimally Improved
- 4 = No Change
- 5 = Minimally Worse
- 6 = Much Worse
- 7 = Very Much Worse

For analysis purpose, a dichotomous scale of 'Yes' or 'No' will be derived. Scale "much improved" or "very much improved" will be considered improved (Yes), and all other non-missing responses will be considered not improved (No). Missing category will be presented as appropriate.

The response variable (yes or no) at week 12 will be analyzed using a logistic regression model with the treatment, age group at FM onset, and sex as explanatory factors. Missing category will not be included.

SAS code is:

```
ods output diffs = dif LSMeans = ls;
PROC GENMOD DATA =name;
    CLASS TRT01PN STRATAR SEX;
    MODEL AVALCAT1 = TRT01PN STRATAR SEX / DIST=BIN TYPE3;
  LSMEANS trt01pn/exp DIFF CL;
RUN;
ods output close;
```

```
data ls;
  set ls;
    ** p = estimated response rate;
       p = logistic(estimate);
       p  l = logistic(Lower) ;
       p_u = logistic(Upper);
run;
```

Where AVALCAT1 denotes the responses with 'yes' for improved and 'no' for not improved.

In addition, the PGIC data will be summarized as a categorical variable using descriptive statistics for all visits, including missing category as appropriate.

6.3.3. Percentage of Patients Who Experience a ≥30% or ≥50% Reduction in the Weekly Average of the Daily Average PI-NRS Score at Week 12

The weekly average of the daily average PI-NRS score with missing data imputation from the 500 imputed data sets derived as described in steps 1 to 3 of Section 6.2.2 will be used to calculate the pain intensity reduction (%) from baseline in the weekly average of the daily average PI-NRS score using formula [2].

$$\frac{baseline\ value - postbaseline\ value}{baseline\ value} \times 100\% \quad [2]$$

The percentage of patients who experience ≥30% reduction and ≥50% reduction in the weekly average of the daily average PI-NRS score at week 12 will be analyzed using a generalized estimating equation (GEE) method where the dependent variable is ≥30% reduction (Yes/No) or ≥50% reduction (Yes/No), respectively, [redacted]

SAS code for GEE model:

```
ods output LSMeans = ls;
PROC GENMOD DATA =name;
     BY _IMPUTATION_;
      CLASS SUBJID TRT01PN WEEK STRATAR SEX;
      MODEL AVALCAT = BASE TRT01PN WEEK STRATAR SEX
                          TRT01PN*WEEK/ DIST=BIN TYPE3;
      REPEATED SUBJECT = SUBJID / TYPE = AR(1);
   LSMEANS trt01pn*week/exp DIFF CL;
RUN;
ods output close;
```

Where AVALCAT denotes the responses with 'yes' for ≥30% or ≥50% or 'no' for <30% or <50% pain intensity reduction, respectively.

The 500 sets of results from above analyses will be combined using Rubin's formulae (PROC MIANALYZE).

```
PROC SORT DATA= LS;
```

```
    BY WEEK TRT01PN _IMPUTATION_;
RUN;

ODS OUTPUT PARAMETERESTIMATES= LSM_O;
PROC MIANALYZE DATA= LS;
    BY WEEK TRT01PN;
    MODELEFFECTS ESTIMATE;
    STDERR STDERR;
RUN;

DATA LSM_O;
  SET LSM  O;
  EXPESTIMATE = EXP(ESTIMATE);
  P = 100*EXPESTIMATE/(1 + EXPESTIMATE);
  P  LL = 100*EXP(LCLMEAN)/(1 + EXP(LCLMEAN));
  P  UL = 100*EXP(UCLMEAN)/(1 + EXP(UCLMEAN));
RUN;
```

In addition, the mean percentage of reduction and its associated 95% CI will be reported by week.

A plot of percentage of improvement versus percentage of patients meeting the improvement level at week 12 will be presented based on observed data and missing data at week 12 will be deemed as pain intensity increased from baseline. The x-axis representing percentage of improvement will be from >= 0 to 100% by an increment of 10%, ie., the x-axis will be at cutoff points ≥ 0%, ≥10%, ≥20%, …, ≥90% and 100% while y-axis will be percent of patients meet the cutoff points. The percentage of improvement will be calculated using formula [2].

6.3.4. Change From Baseline to Week 12 in the Weekly Average of the Daily Worst PI-NRS Score Over the Past 24 Hours

The weekly average of the daily worst PI-NRS score over the past 24 hours will be analyzed in a manner analogous to the primary endpoint as described in Section 6.2.2. The dependent variable is the change from baseline in the weekly average of the daily worst PI-NRS scores. The covariate will be baseline average of the daily worst PI-NRS scores.

6.3.5. Change From Baseline to Week 12 in the Patient-Reported Outcomes Measurement Information System Sleep Disturbance Short Form 8a Score

The PROMIS sleep disturbance Short Form 8a (SF8a) scale contains 8 items and assesses sleep disturbance over the past 7 days. One item assessing overall sleep quality use a scale of *very poor, poor, fair, good, or very good.* Other 7 items use a scale of *not at all, a little bit, somewhat, quite a bit, or very much*. Each item of the PROMIS sleep disturbance SF8a is on 5-point scales (1 to 5), with 1 for the lowest sleep disturbance and 5 for the highest sleep disturbance. The total score ranges from 8 to 40 (highest sleep disturbance). Collection of the PROMIS sleep disturbance SF8a will start the day of the BAPS Visit (Visit 2) and will be administered weekly via the e-diary. The total raw score ranging from 8 to 40 (highest sleep disturbance) will be calculated as the summation of 8 non-missing scores. If any of 8 items has a missing score, the total raw score will be missing. The calculated total raw score will be converted into t-score

using the scoring table (PROMIS-Sleep Disturbance Scoring Manual). All questions must be answered in order to produce a valid score using the scoring table. The analysis will be based on t-score.

The change from baseline to week 12 in t-score of the PROMIS sleep disturbance SF8a will be analyzed using MMRM method in manner analogous to the analysis as described in Section 6.3.1. The dependent variable is the change from baseline in the t-score of the PROMIS sleep disturbance SF8a at analysis weeks 1 to 12 (12 levels). BASE is the t-score of the PROMIS sleep disturbance SF8a at baseline. Missing will be considered as missing at random. The results from all visits will be presented using descriptive statistics.

Summary statistics will include data from all analysis weeks.

6.3.6. Change from Baseline to Week 12 in the Patient-Reported Outcomes Measurement Information System Physical Function SF12a Score

The PROMIS physical function scale measures self-reported capability rather than actual performance of physical activities. This includes the functioning of one's upper extremities (dexterity), lower extremities (walking or mobility), and central regions (neck, back), as well as instrumental activities of daily living, such as running errands over the past 7 days. A single physical function capability score is obtained from a SF. The PROMIS physical function SF12a scale contains 12 items with 5-point scales (1 to 5) for each item with a total score ranging from 12 to 60. A lower score indicates more limited physical disability.

The PROMIS physical function SF12a will be assessed at BAPS Visit (visit 2) and visits 3 to 7 (weeks 1, 4, 8, 12, and 16).

The total raw score will be calculated as the summation of 12 non-missing scores for patients who can walk 25 feet or summation of 6 non-missing scores for patients who cannot walk for 25 feet. The calculated total raw score will be converted into scale score using the scoring tables (PROMIS-Physical Function Scoring Manual). Different scoring tables will be used for patients who can walk and patients who cannot walk. All questions must be answered in order to produce a valid score using the scoring table. The analysis will be based on the scale score.

The change from baseline to week 12 in the scale score of PROMIS physical function SF12a will be analyzed using MMRM method in manner analogous to the analysis as described in Section 6.3.1. The dependent variable is the change from baseline in the scale score of PROMIS physical function SF12a at weeks 4, 8, and 12 (3 levels). BASE is the scale score of PROMIS physical function SF12a at baseline. Missing will be considered as missing at random. The results from all visits will be presented using descriptive statistics.

Summary statistics will include data from all visits.

6.3.7. Change From Baseline To Week 12 In The Patient-Reported Outcomes Measurement Information Fatigue SF8a Score

The PROMIS Fatigue SF8a contains 8 items with 5-point scales (1 to 5) for each item over the past 7 days. The total score ranges from 8 to 40 (highest level of fatigue). The PROMIS fatigue SF8a will be assessed at BAPS Visit (visit 2) and visits 3 to 7 (weeks 1, 4, 8, 12, and 16).

The total raw score will be calculated as the summation of 8 non-missing scores. The calculated total raw score will be converted into t-score using the scoring table (PROMIS-Fatigue Scoring Manual). All questions must be answered in order to produce a valid t-score using the scoring table. The analysis will be based on t-score.

The change from baseline to week 12 in the t-score of the PROMIS Fatigue SF8a will be analyzed using MMRM method in manner analogous to the analysis as described in Section 6.3.1. The dependent variable is the change from baseline in the t-score of the PROMIS Fatigue SF8a at weeks 4, 8, and 12 (3 levels). BASE will be the t-score of PROMIS fatigue SF8a from the baseline. Missing will be considered as missing at random. The results from all visits will be presented using descriptive statistics.

Summary statistics will include data from all visits.

6.3.8. Number (%) of Patients Who Did Not Complete the Treatment Due to Lack of Efficacy

Time to treatment discontinuation will be assessed for all patients and will be measured from the date of the first dose of study medication to the date of treatment completion or treatment discontinuation before visit 7. Time to treatment discontinuation is defined as the date of EOT - the date of 1st dosing of IMP + 1. Patients who completed the treatment will be censored at the date of EOT. Patients who did not complete the treatment due to reasons other than lack of efficacy will be censored at the date of ET. Patients who are lost to follow-up will be censored at the date of their last visit.

Time to treatment discontinuation will be analyzed using Kaplan-Meier survival analysis method. Number (%) of patients who do not complete the treatment due to lack of efficacy (the event) and patients censored will be presented. Kaplan-Meier estimates (median time to treatment discontinuation due to lack of efficacy, survival probability at weeks 4, 8, 12 and 16) and survival plot will also be presented.

6.4. [redacted]

6.4.1. [redacted]

[redacted]

6.4.2. [redacted]

[redacted]

6.4.3. [redacted]

[redacted]

[redacted]

6.4.4. [redacted]

6.4.5. [redacted]

[redacted]

7. MULTIPLE COMPARISONS AND MULTIPLICITY

No multiplicity adjustment is planned for this study.

8. SAFETY ANALYSIS

8.1. General

The safety analysis set will be used for all safety analyses. Summaries will be presented by treatment group as actually received unless otherwise stated.

In the Study TV48125-PN-20028 protocol amendment 04, the data collections had been changed:

- Some data are not collected as they were prior to protocol amendment 04 at some visits; for example, clinical laboratory data are no longer collected at visits 4 and 5.
- Follow-up visit was removed

This section reflects data collections in the study including data collected prior to Study TV48125-PN-20028 protocol amendment 04. For by-visit summaries, all data collected in the study will be summarized.

8.2. Duration of Exposure to Study Drug

Duration of exposure to study drug (days) for the individual patient is the number of days patient received drug (last day of study drug – first day of study drug + 28).

Number (%) of patients receiving 1 dose, 2 doses, 3 doses, and 4 doses and duration of treatment (days) will be summarized by treatment group using descriptive statistics.

8.3. Adverse Events

An adverse event (AE) is any untoward medical occurrence in a patient administered a pharmaceutical product which does not necessarily have to have a causal relationship with this treatment.

For recording of AE, the study period is defined for each patient as the time period from signature of the informed consent form (ICF) to the end of the follow-up period or the early termination visit for patients who discontinue from the study for any reason. Any AE occurring on or after the first dose of study drug is considered a treatment-emergent AE (TEAE). All AE summaries will be conducted for TEAEs.

The following are considered protocol-defined adverse events of special interest (AESI) to be sent to the sponsor's Global Patient Safety and Pharmacovigilance (GPSP) Department for evaluation:

- ophthalmic-related adverse events of at least moderate severity, as determined by the investigator
- severe hypersensitivity reactions

All AEs will be coded using MedDRA. The incidence of AEs will be summarized using descriptive statistics by SOC and preferred term (PT). Each patient will be counted only once in each PT or SOC category for the analyses of safety.

Summaries will be presented for all AEs (overall and by severity), AEs determined by the investigator to be related to test IMP (ie, reasonable possibility) (defined as related or with missing relationship) (overall, overall by severity, serious AEs, and AESI), serious AEs, AESI, AEs causing discontinuation from treatment, AEs causing early termination from the study, injection site AEs, and AEs leading to death.

Listings of AEs, AESI, serious AEs, injection site AEs, and AEs leading to early termination will be presented separately.

8.4. Deaths

If any patient dies during the study all relevant information will be discussed in the patient's narratives included in CSR.

8.5. Clinical Laboratory Tests

Clinical laboratory tests (serum chemistry, hematology, coagulation, and urinalysis) will be performed at the time points detailed in Table 1 in the study protocol using a central laboratory. Clinical laboratory tests will be performed using a central laboratory. Specific laboratory tests to be performed are provided in Table 3.

Table 3: Clinical Laboratory Tests

Serum Chemistry	Hematology and Coagulation	Urinalysis
Calcium Phosphate Sodium Potassium Chloride Creatinine Glucose Blood urea nitrogen (BUN) Urate Alanine aminotransferase (ALT) Aspartate aminotransferase (AST) Lactate dehydrogenase (LDH) Gamma-glutamyl transpeptidase (GGT) Alkaline phosphatase Bicarbonate Carbon dioxide Protein Albumin Bilirubin Direct bilirubin Hepatitis B	Hemoglobin Hematocrit Erythrocytes Platelets Leukocytes • Neutrophils • Lymphocytes • Eosinophils • Monocytes • Basophils Lymphocytes atypical Prothrombin International Normalized Ratio (INR)	Protein Glucose Ketones Hemoglobin pH Specific gravity Microscopic tests • Bacteria • Erythrocytes • Leucocytes • Crystals • Casts

Laboratory test results will be presented in standard international (SI) units.

Summary statistics for chemistry, hematology, coagulation, and urinalysis laboratory tests will be presented at baseline, week 4, week 8, week 12, week 16, and last assessment. Laboratory values and changes from baseline to week 4, week 8, week 12, week 16, and last assessment will be summarized using descriptive statistics.

Shifts (below, within, and above the normal range) from baseline to week 4, week 8, week 12, week 16, and last assessment will be summarized using patient counts.

Summaries of potentially clinically significant abnormal values will include all postbaseline values (including scheduled, unscheduled, and withdrawal visits). The incidence of potentially clinically significant abnormal values will be summarized for laboratory variables using descriptive statistics with the criteria specified in Table 4.

Table 4: Criteria for Potentially Clinically Significant Laboratory Values

Test		Criterion value
Serum chemistry		
Alanine aminotransferase (ALT)		≥3x ULN
Aspartate aminotransferase (AST)		≥3x ULN
Alkaline phosphatase (ALP)		≥3x ULN
Gamma-glutamyl transpeptidase (GGT)		≥3x ULN
Lactate dehydrogenase (LDH)		≥3x ULN
Blood urea nitrogen (BUN)		≥10.71 mmol/L
Creatinine		≥177 µmol/L
Uric acid	Men	≥625 µmol/L
	Women	≥506 µmol/L
Bilirubin (total)		≥34.2 µmol/L
Hematology		
Hematocrit	Men	<0.37 L/L
	Women	<0.32 L/L
Hemoglobin	Men	≤115 g/L
	Women	≤95 g/L
White blood cell (WBC) counts		≤3 x 10^9/L ≥20 x 10^9/L
Eosinophils		≥10%
Absolute neutrophil counts (ANC)		≤1 x 10^9/L
Platelet counts		≤75 x 10^9/L ≥700 x 10^9/L
Urinalysis		
Blood (HGB)		≥2 unit increase from baseline
Glucose		≥2 unit increase from baseline

Test	Criterion value
Ketones	≥2 unit increase from baseline
Total protein	≥2 unit increase from baseline

ULN=upper limit of normal range.

Listings for potentially clinically significant abnormal laboratory data will be presented.

8.5.1. Laboratory Values Meeting Hy's Law Criteria

All occurrences of possible drug-induced liver injury that meet Hy's law criteria, defined as all of the below, must be reported by the investigator to the sponsor as a serious adverse event:

- ALT or AST increase of >3× the ULN
- total bilirubin increase of >2× the ULN
- absence of initial findings of cholestasis (ie, no substantial increase of alkaline phosphatase)

An adverse event that does not meet any of the criteria for seriousness listed above will be regarded as a non-serious adverse event.

The serious adverse events will be summarized as indicated in Section 8.3.

8.5.2. Other Clinical Laboratory Tests

At the screening visit, patients will be tested for human immunodeficiency virus (HIV), hepatitis B surface antigen (HBsAg), hepatitis C antibody, and thyroid stimulating hormone. Data will be listed.

Beta-human chorionic gonadotropin tests in serum will be performed for all women of childbearing potential at the screening visit. Data will be listed.

Follicle-stimulating hormone (FSH) test (for postmenopausal women only) will be performed at screening and baseline visit for data collected up to and including Study TV48125-PN-20028 protocol amendment 04, but at screening only for data collected following Study TV48125-PN-20028 protocol amendment 05 The data will be summarized using descriptive statistics.

A urine drug screen will be performed at the time points specified in Table 1 in the study protocol. Data will be listed.

8.6. Physical Examinations

Physical examinations, including height (to be measured at the screening visit only) and weight, will be performed at the time points specified in Table 1 in the study protocol. Any physical examination finding that is judged by the investigator as a clinically significant change (worsening) compared with a baseline value will be considered an adverse event, recorded on the eCRF, and monitored as described in the study protocol.

Descriptive statistics for weight and height will be provided. Abnormal physical examination findings will be summarized by visits.

8.7. Vital Signs

Vital signs (blood pressure [systolic/diastolic], respiratory rate, oral body temperature, and pulse) will be measured at the time points detailed in Table 1 in the study protocol.

Summary statistics for vital signs (blood pressure [systolic/diastolic], respiratory rate, oral body temperature, and pulse) will be presented at baseline, week 4, week 8, week 12, week 16 (end of treatment), week 30 (end of study), and last assessment. Vital signs values and changes from baseline to each visit will be summarized using descriptive statistics.

Summaries of potentially clinically significant abnormal values will include all postbaseline values (including scheduled, unscheduled, and withdrawal visits). The incidence of potentially clinically significant abnormal values will be summarized using descriptive statistics with the criteria specified in Table 5. Systolic and diastolic blood pressure will be included in all summaries.

Table 5 specifies the criteria for identifying vital signs as potentially clinically significant abnormal values. Note that in order to qualify as potentially clinically significant abnormal, a value needs to meet both criteria below: ie, have a value beyond the criterion value and a change of at least the magnitude specified in the change relative to baseline column.

Table 5: Criteria for Potentially Clinically Significant Vital Signs

Vital Sign	Criterion value	Change relative to baseline
Pulse	≥120 bpm	Increase of ≥15
	≤50 bpm	Decrease of ≥15
Systolic blood pressure	≥180 mm Hg	Increase of ≥20
	≤90 mm Hg	Decrease of ≥20
Diastolic blood pressure	≥105 mm Hg	Increase of ≥15
	≤50 mm Hg	Decrease of ≥15
Respiratory rate	<10 breaths/min	
Body temperature	≥38.3°C	Change of ≥1.1°C

bpm=beats per minute

Listings for potentially clinically significant abnormal laboratory data will be presented.

8.8. Electrocardiography

A 12-lead ECG will be recorded at the time points detailed in Table 1 of the study protocol.

All ECG results outside of the reference ranges will be judged by the investigator as belonging to one of the following categories:

- abnormal and not clinically significant
- abnormal and clinically significant

Any ECG finding that is judged by the investigator as clinically significant (except at the screening visit) will be considered an adverse event and recorded on the source documentation and in the eCRF.

Shifts (normal and abnormal) from baseline to overall result interpretation, week 12, week 16 (end of treatment), and last assessment will be summarized using patient counts. For overall result interpretation the worst postbaseline finding for the patient (the abnormal finding if there are both normal and abnormal findings) will be used in the summaries. Summary statistics for ECG variables will be presented. Actual values and changes from baseline to week 12, week 16 (end of treatment), and last assessment will be summarized using descriptive statistics.

8.9. Concomitant Medications or Therapies

Concomitant therapies and medications, including medications that are taken on an as needed basis and occasional therapies, will be monitored during the study. Details of prohibited medications may be found in the study protocol. All concomitant medications will be coded using the WHO Drug Dictionary.

The incidence of concomitant therapies and medications will be summarized using descriptive statistics by therapeutic class category and preferred term. Patients are counted only once in each therapeutic class, and only once in each preferred term category. Concomitant therapies and medications will include all medications taken after the 1st dose of study drug administration up to the end of study as defined in the study protocol.

Similarly the prohibited medications will be summarized separately.

8.10. Columbia-Suicide Severity Rating Scale

The C-SSRS will be used to assess the patient's suicidal ideation (severity and intensity) and behavior. The C-SSRS Baseline/Screening version will be administered by certified site staff at Visit 1, and the C-SSRS Since Last Visit version will be completed by certified site staff at all other time points, as described in Table 1 of the study protocol. Any significant findings on the C-SSRS require evaluation by the site investigator.

Patients having positive findings will be listed.

9. TOLERABILITY VARIABLES AND ANALYSIS

Spontaneous reports of injection site reactions will be recorded as adverse events according to the following severity assessment criteria:

- Spontaneous reports of injection site erythema, induration, and ecchymosis will be assessed and recorded by site personnel and categorized according to the following measurements: 5 to ≤50 mm (mild), >50 to ≤100 mm (moderate), and >100 mm (severe).
- Spontaneous reports of local pain after the injection will be recorded as mild, moderate, or severe according to patient's self-report.
- Appropriate treatment may be provided if necessary, in which case it must be recorded as concomitant medication.

Local tolerability findings (eg, erythema, induration, ecchymosis, and occurrence of injection site pain) will be listed and summarized descriptively.

Injection site adverse events will be summarized as indicated in Section 8.3.

10. PHARMACOKINETIC ANALYSIS

There are no prespecified pharmacokinetic endpoints.

Fremanezumab plasma concentration will be summarized using descriptive statistics by active treatment groups and each planned sampling time point (samples from patients who received placebo will not be analyzed). The summary will be based on the safety analysis set. The plasma concentration will be listed by active treatment groups, scheduled visits, and time points.

For summary displays of plasma concentration data, values below the limit of quantitation (BLQ) will be treated as 0. If the mean concentration at a time point is less than the lower limit of quantitation (LLOQ), then "BLQ" will be reported in place of summary statistics. Values below LLOQ will be imputed as ½*LLOQ for calculation of geometric mean and geometric CV. The LLOQ of fremanezumab is 250 ng/mL.

11. PHARMACODYNAMIC ANALYSIS

Details of the pharmacodynamic analysis will be provided in a separate SAP.

12. PHARMACOKINETIC/PHARMACODYNAMIC ANALYSIS

Details of the pharmacokinetic/pharmacodynamic analysis will be provided in a separate SAP.

13. BIOMARKER ANALYSIS

Details of the biomarker analysis will be provided in a separate SAP.

14. IMMUNOGENICITY ANALYSIS

Details of the immunogenicity analysis will be provided in a separate SAP.

15. ANCILLARY STUDIES ANALYSIS

No ancillary analyses are planned for this study.

16. PLANNED INTERIM ANALYSIS

There were two interim analyses for futility performed in this study. The details of the interim analyses were specified in seperate interim statistical analysis plans. To ensure study integrity, the interim analyses were performed by an independent third party; the study team were kept blinded. The decision was made to end this study for futility following the results of the second interim analysis.

17. STATISTICAL SOFTWARE

All data listings, summaries, and statistical analyses will be generated using SAS® version 9.4 or later.

18. CHANGES TO ANALYSES SPECIFIED IN THE STUDY PROTOCOL

NA.

19. REFERENCES

Arthritis Research & Therapy 2009, 11:R120 (doi:10.1186/ar2783)

The Patient-Reported Outcomes Measurement Information System. Fatigue Scoring Manual. Available at http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Fatigue_Scoring_Manual.pdf. Retrieved 15 May 2018.

The Patient-Reported Outcomes Measurement Information System. Physical Function Scoring Manual. Available at http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Physical_Function_Scoring_Manual.pdf. Retrieved 15 May 2018.

The Patient-Reported Outcomes Measurement Information System. Sleep Disturbance Scoring Manual. Available at http://www.healthmeasures.net/images/promis/manuals/PROMIS_Sleep_Disturbance_Scoring_Manual.pdf. Retrieved 15 May 2018.