PROTOCOL A4091059

A PHASE 3, RANDOMIZED, DOUBLE-BLIND, PLACEBO AND ACTIVE-CONTROLLED, MULTICENTER, PARALLEL-GROUP STUDY OF THE ANALGESIC EFFICACY AND SAFETY OF TANEZUMAB IN ADULT SUBJECTS WITH CHRONIC LOW BACK PAIN

Statistical Analysis Plan (SAP)

Version: 3.0

Author: PPD, PPD, PPD

Date: 18th Jan, 2019

TABLE OF CONTENTS

LIST OF FIGURES

APPENDICES

1. AMENDMENTS FROM PREVIOUS VERSION(S)

The major changes from the prior version of the SAP (v2.0, 15th May 2016) are listed below. The changes reflect updates from blinded data reviews and program decisions for alignment of analysis. Additionally, clarifications, removal of redundant/duplicated text, and correction of typos have been implemented.

- Section 2.1: Clarified that most safety results will be presented for the treatment period.
- Section 2.1: Moved the specification of the analysis of combining subjects who are on placebo for the first 16 weeks and receive tanezumab afterwards and patients who are on tanezumab since randomization to Section 8.2.
- Section 4: Updated the multiplicity control plan.
- Section 5.4: Removed descriptions of biomarker and NGF populations.
- Section 6.2: Added safety endpoints specified in the protocol for completeness. Updated definitions for Abnormal Peripheral Sensation and Sympathetic Nervous System. Moved the text related to the analysis to Section 8.2.
- Section 6.3.3: Moved the text related to the analysis to Section 8.2.
- Section 6.3.4: Added adjudication outcome endpoints and moved the text related to the analysis to Section 8.2.
- Section 6.5: Added subgroup analyses by region (EU vs Non-EU), by Baseline painDetect category (painDetect score <=12, 13 to 18, >=19) and by Baseline pain severity (LBPI <7 vs LBPI $\geq$7).
- Section 7: Clarified the details for multiple imputation and clarified the details for MMRM to keep consistent with changes in Section 8.2.
- Section 8.1: Moved the text related to the specific analyses to Section 8.2 so Section 8.1 remains a section for general statistical methods.
- Section 8.2: Specified the estimand strategy and description of planned on-treatment efficacy assessment period. Updated treatment display plan for the efficacy summaries and analyses.
- Section 8.2.1: Specified a sensitivity analysis excluding subjects from 2 sites with potential GCP compliance issues and excluding the subject who may have been enrolled twice in the study.
- Section 8.2.1: Removed the interaction analysis with Baseline score. Removed the unadjusted ANOVA analysis.

- Section 8.2.2: Updated to align with the beginning part of Section 8.2. Added the analysis of average low back pain intensity (aLBPI) score for individual Days 1 through 7.

- Section 8.2.2: Removed the MMRM analysis here as it is specified in Section 8.2.1.

- Section 8.2.2: Added analyses for PGA-LBP. Specified missing data approach (multiple imputation/BOCF/LOCF for Roland Morris Disability Questionnaire (RMDQ) and PGA-LBP, and multiple imputation for others).

- Section 8.2.2: Removed the text on the multiplicity control approach as it is stated in Section 4.

- Section 8.2.2: Moved the cumulative distribution of percent change from Baseline in RMDQ to "Response and Incidence Endpoints" section.

- Section 8.2.2: Updated missing data approach for response and incidence endpoints. Clarified the definition of "incidence of discontinuation due to lack of efficacy".

- Section 8.2.2: Removed the summary of the proportion of subjects who meet an aLBPI response definition at Week 16 (improvement from Baseline in aLBPI≥30% at Week 16, and ≥15% at any of Weeks 1-15).

- Section 8.2.2: Clarified the definition of event "discontinuation from treatment due to lack of efficacy".

- Section 8.2.2: Moved the analysis model for the number of days and amount of rescue medication endpoints here from an early section.

- Section 8.2.2: Removed the analysis for incidence of rescue medication use as it is included in an early section.

- Section 8.2.2: Moved the details for CMH test for PGA here from an early section. Removed LOCF and BOCF missing data approaches. Specified the WPAI analysis using ANCOVA model instead of CMH test

- Section 8.2.3: Specified general safety summary/analysis plan.

- Section 8.2.3.1: Moved the details of the NIS analysis here from an early section.

- Section 8.2.3.1: Specified the summaries for joint safety events. Removed time to event summaries for joint safety events.

- Section 8.2.3.1: Changed the Tier 2 AE cutoff from 5% to 3%.

- Section 8.2.3.1: Removed Tier 1 and 2 AE graphs.

- Section 8.2.3.1: Added a few non-standard summaries.
- Section 8.2.3.1: Removed AE patient-year summaries, AE plots, summaries of specific AE start day and duration; added NSAID use summaries.
- Section 8.2.3.1: Removed the summary of mean change in postural blood pressure.
- Section 8.2.3.1: Clarified neurologic data summaries.
- Section 8.2.3.2: Clarified PK data summaries.
- Section 8.2.3.3: Clarified NGF data summaries.
- Section 8.2.3.4: Removed as biomarker data are not generated.
- Section 8.2.4: Removed this section as its contents are included in early sections.
- Section 10: Clarified the windows for various data types.

The changes from the prior version of the SAP (v1.0, 14th Feb 2014) are listed below:

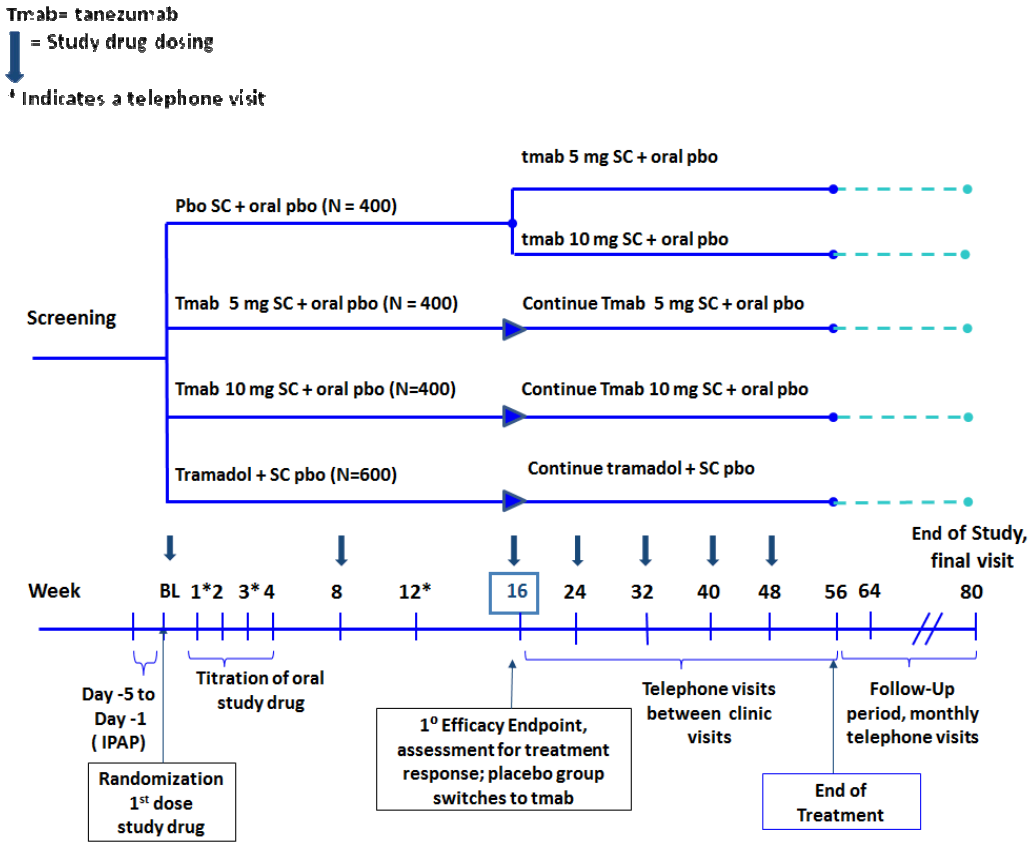
- Section 2.1: Added information regarding the study;
- Section 5.1, 5.4 and 5.5: Modified;
- Section 6.1: Added percent reduction and cumulative endpoints, Health Care Resource Utilization endpoint, Tertiary endpoints;
- Section 6.5: Modified and added more information;
- Section 8.2.2: Modified and added more information;
- Section 8.2.3.1 and 8.2.3.2: Added Safety Endpoints and Assessments;
- Section 8.2.4, 8.2.4.1, 8.2.4.2 and 8.2.4.3 : Modified and added more information regarding PK/PD and Biomarkers analyses;
- Section 8.2.5: Modified and added more information;
- Appendix 1.1. and Appendix 1.4: Visit windows modified, Rescue Medication days modified;
- Appendix 2.3: Schedule of visit modified using newest version of the final protocol;
- Appendix 3: Immunogenicity Data Reporting added.

2. INTRODUCTION

Note: in this document any text taken directly from the protocol is ***italicized***.

2.1. Study Design

The study design is summarized in the diagram below.



Subjects will be randomized at Baseline to one of four treatment groups: placebo SC + oral placebo, tanezumab 5mg SC + oral placebo, tanezumab 10mg SC + oral placebo, or to placebo SC + oral tramadol. These will be labeled as placebo, tanezumab 5mg, tanezumab 10mg and tramadol for the four treatment groups respectively. At Week 16, patients who received placebo will be switched in a blinded fashion to either tanezumab 5mg or tanezumab 10mg. Tramadol or oral placebo dosing will be once daily, and subject to patient dose titration in the range 100 to 300mg (with matching placebo titration). Tanezumab or placebo SC dosing will be every 8 weeks at Baseline and Weeks 8, 16, 24, 32, 40 and 48 (a maximum of 7 doses of SC tanezumab or placebo).

The end of treatment period is at Week 56, with the safety follow-up period up to Week 80. The primary time point for efficacy is Week 16. The period of interest for most safety results is the treatment period. Selected safety results will be provided separately for the safety follow-up period and for the combined overall study period comprising the treatment and safety follow-up periods. *A minimum sample size of approximately 400 subjects per*

treatment group are needed to provide at least 80% power to achieve statistical significance (at the 5% two-sided level) for both comparisons of tanezumab 10 mg and 5 mg versus placebo as well as the comparison of tanezumab 10 mg versus active comparator in the primary endpoint, Change from Baseline to Week 16 in the average LBPI score. Since placebo subjects reaching Week 16 response criteria will be switched to tanezumab treatment only, in order to balance subject exposure during the safety phase of the trial (post Week 16) the number of subjects randomized at Baseline to the active comparator group is increased to 600. The total sample size will be approximately 1800 subjects.

The randomization will not be stratified. To achieve the initial randomization, and re-randomization for placebo patients at Week 16, a blocked static randomization using the ratio 1:1:2:2:3 for placebo→tanezumab 5mg (at Week 16), placebo→tanezumab 10mg (at Week 16), tanezumab 5mg, tanezumab 10mg, and tramadol will be performed at the beginning of the trial. No active re-randomization will occur at Week 16.

2.2. Study Objectives

Primary Objective

- *Demonstrate superior analgesic efficacy of tanezumab 10 mg and 5 mg administered subcutaneously (SC) every 8 weeks compared to placebo at Week 16.*

Secondary Objectives

- *Evaluate the long-term safety of tanezumab 10 mg and 5 mg SC administered every 8 weeks (7 administrations);*
- *Estimate the long-term analgesic efficacy of tanezumab 10 mg and 5 mg SC administered every 8 weeks (7 administrations) up to Week 56;*
- *Compare the analgesic efficacy of tanezumab 10 mg SC administered every 8 weeks relative to an active comparator (oral tramadol PR) at Week 16.*

3. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING

There is no interim analysis for efficacy planned for this study. The final analysis will be performed after the database is released.

Safety data will be subject to regular and ongoing reporting and review throughout the study. The details of these interim analyses will be documented in a separate Statistical Analysis Plan. Review of the safety data will be by the tanezumab External Data Monitoring Committee (E-DMC).

Events relating to joint safety, including reported Osteonecrosis or events leading to Total Joint Replacement will be reviewed by a blinded expert adjudication panel. A stopping rule relating to a set of adjudicated outcomes has been defined, and is described below.

If the blinded Adjudication Committee identifies adjudicated events of rapidly progressive osteoarthritis type 2, subchondral insufficiency fractures, primary osteonecrosis, or pathological fracture, occurring at a rate that could trigger the protocol-based stopping criteria, an urgent, ad hoc assessment of the events will be made by the Data Monitoring Committee.

The protocol (or treatment group) stopping rule has three components; the difference in the number of subjects with an adjudicated joint safety event, the exposure-adjusted risk difference (RD) and the exposure adjusted risk ratio (RR) between each tanezumab treatment group and the tramadol PR treatment group. The exposure-adjusted RD will be calculated as the difference in the ratios of the number of subjects with an adjudicated joint safety event divided by exposure (patient-years) between each tanezumab group and the comparator group. The exposure-adjusted RR will be similarly calculated using the ratio of exposure adjusted event rates (number of subjects with an adjudicated joint safety event divided by exposure) for each tanezumab group relative to the comparator group. The exposure will be calculated as the combined treatment and follow-up periods. For subjects who started treatment on placebo and switched to tanezumab, the tanezumab exposure will not include the initial placebo time.

If the [redacted] *(ie, &* [redacted] *), and the RR is* [redacted] *and the difference in the number of subjects with adjudicated events joint safety events for any tanezumab treatment group versus the comparator treatment group, the protocol-based stopping rule will be triggered. If the protocol-based stopping rule is triggered, the DMC will formulate a recommendation whether it is safe to continue dosing in some or all treatment groups or whether the study should be terminated completely. This decision will be made by Pfizer in consultation with the Data Monitoring Committee.*

Separate sets of dosing suspension rules for specified Serious Adverse Events and events consistent with Hy's Law are described in Sections 9.6.1.1 and 9.6.1.2 of the protocol, respectively.

4. APPROPRIATE HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

The treatment comparisons being made in this study are tanezumab 5 mg and 10 mg versus placebo (for the primary study objective), and tanezumab 5 mg and 10 mg versus tramadol (for the secondary efficacy study objective). For these treatment comparisons, the null and alternative hypotheses are shown below (note $\mu_{TREATMENT}$ relates to the mean change from Baseline for the specified treatment group). All tests will be 2-sided.

Comparisons of tanezumab versus placebo will be made for data up to and including Week 16 and comparisons of tanezumab versus tramadol will be made for data up to and including Week 56.

Null Hypotheses	H0: $\mu_{\text{TANEZUMAB 5mg}} - \mu_{PLACEBO} = 0$
	H0: $\mu_{\text{TANEZUMAB 10mg}} - \mu_{PLACEBO} = 0$
	H0: $\mu_{\text{TANEZUMAB 5mg}} - \mu_{TRAMADOL} = 0$
	H0: $\mu_{\text{TANEZUMAB 10mg}} - \mu_{TRAMADOL} = 0$
Alternative Hypotheses	H1: $\mu_{\text{TANEZUMAB 5mg}} - \mu_{PLACEBO} \neq 0$
	H1: $\mu_{\text{TANEZUMAB 10mg}} - \mu_{PLACEBO} \neq 0$
	H1: $\mu_{\text{TANEZUMAB 5mg}} - \mu_{TRAMADOL} \neq 0$
	H1: $\mu_{\text{TANEZUMAB 10mg}} - \mu_{TRAMADOL} \neq 0$

The hypotheses for other types of analyses (eg, for the binary response endpoints) would be similar to those shown above.

4.2. Statistical Decision Rules

The primary efficacy endpoint and key secondary endpoints included in this study are:

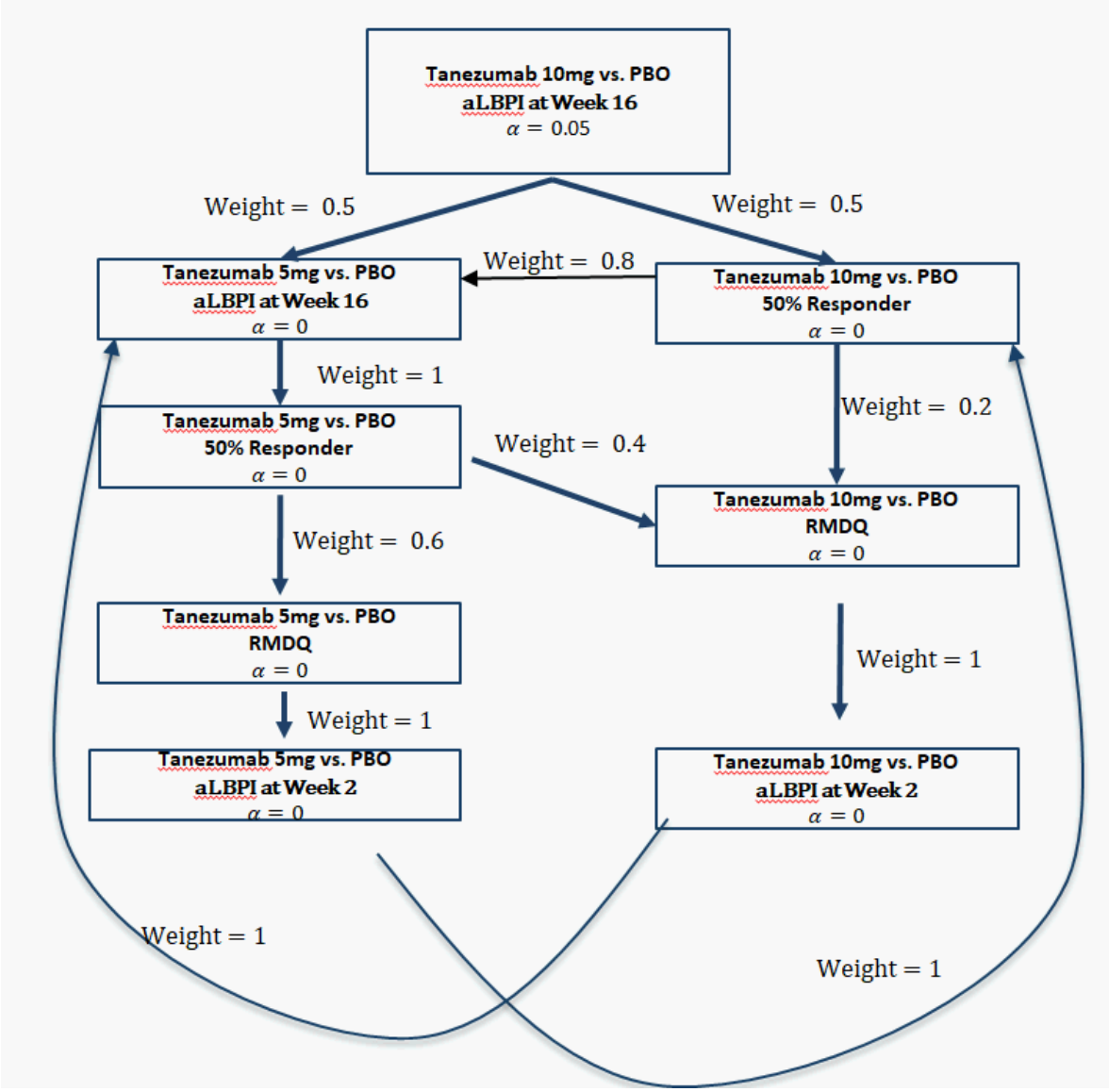
- Change from Baseline to Week 16 in the daily average LBPI score for tanezumab vs placebo (aLBPI at Week 16);
- Change from Baseline to Week 16 in the RMDQ for tanezumab vs placebo (RMDQ);
- Response as defined by a ≥50% reduction from Baseline in daily average LBPI score at Week 16 for tanezumab vs placebo (50% Responder);
- Change from Baseline to Week 2 in the average LBPI score for tanezumab vs placebo (aLBPI at Week 2).

The testing of the primary endpoint and key secondary endpoints will follow the graphical approach of gate keeping strategy proposed by Bretz et al (2011)[2], as depicted in the following figure. This will be implemented to control the family-wise type I error rate of 5% (two-sided), and this graphical approach is a closed testing procedure; hence, it strongly controls the family-wise error rate (Alosh et al. 2014).[1]

The graph starts with testing tanezumab 10 mg versus placebo for aLBPI at Week 16 at α=0.05, and if statistically significant (p≤0.05), α=0.05 is split with equal weight to testing tanezumab 10 mg versus placebo for 50% Responder at α=0.025, and testing tanezumab 5 mg versus placebo for aLBPI at Week 16 at α=0.025. If the test of tanezumab 10 mg versus placebo for 50% Responder is significant at α=0.025, additional α=0.02 (0.8*0.025) will be reallocated to the test of tanezumab 5 mg versus placebo for aLBPI at Week 16. Then tanezumab 5 mg versus placebo for aLBPI at Week 16 can be tested at 0.045. The testing process continues as long as at least one null hypothesis can be rejected at its assigned α-level. Each time a null hypothesis is rejected, the graph is updated to reflect the reallocation of α assigned to that hypothesis, which is considered "recycled" (Alosh et al.

2014).[1] This iterative process of updating the graph and reallocating α is repeated until all primary and key secondary endpoints have been tested or when no remaining hypotheses can be rejected at their corresponding α level.

Figure 1. Graphical Multiple Testing Procedure for Strong Control of Type I Error Rate



Control of the type I error rate accounting for multiplicity of contrasts will only apply to the primary endpoint and the key secondary endpoints. Regardless of the outcome of these contrast analyses, other efficacy endpoints will be tested. No adjustment for multiple comparisons will be made for these other efficacy endpoints, and for the safety endpoints. The α-level for each hypothesis test for the secondary and exploratory analyses will be 5%.

5. ANALYSIS SETS

5.1. Full Analysis Set

The intent to treat (ITT) analysis set is the primary analysis set for efficacy analyses. It consists of all randomized subjects who received at least one dose of SC study medication (either tanezumab or placebo SC). This analysis set is used in the presentations of all efficacy data, and is labeled as the 'ITT Analysis Set' or 'ITT Population'. Subjects will be assigned the treatment they were randomized to.

5.2. 'Per Protocol' Analysis Set

The per-protocol (PP) analysis set is the secondary efficacy analysis set. It is defined as all subjects in the ITT analysis set who have no major protocol deviators (which would potentially affect efficacy). The criteria for major protocol deviations are described below in Section 5.6. The identification of specific subjects for this analysis set will be done prior to unblinding. This analysis set is used in a specific sensitivity analysis of the primary efficacy endpoint.

5.3. Safety Analysis Set

The safety analysis set is defined as all subjects treated with tanezumab or placebo SC. This analysis set will be labeled as the 'Safety Analysis Set' or 'Safety Population' in the corresponding safety data analyses, summaries, and listings. Subjects will be assigned per the treatment they actually received.

5.4. Other Analysis Sets

TJR Subset Analysis Set: This analysis set includes all subjects who undergo total joint replacements of the hip, knee or shoulder during participation in the study.

5.5. Treatment Misallocations

If a subject was:

- Randomized but not treated with SC study medication, then that subject will be excluded from all efficacy and safety analyses.
- Treated but not randomized, then by definition that subject will be excluded from the efficacy analyses, but will be reported under the treatment they actually received for all safety analyses.
- Randomized but received incorrect treatment, then that subject will be reviewed on a case-by-case basis by the study team and a decision on potential changes related to the subject and on how to analyze the data will be made in a timely manner and prior to the database release, if possible. For any subject who is identified to receive incorrect treatment after the database release, that subject will be reported under the treatment he/she was randomized to for efficacy analyses, and under the treatment he/she actually received for safety analyses. For subjects who were randomized to tramadol but did not receive it, they will be included in the tanezumab 5mg group.

- Note, patients correctly moving from placebo to tanezumab 5mg or 10mg at Week 16 will be reported under the placebo group for data up to and including Week 16, and the appropriate tanezumab group for data after Week 16.

5.6. Protocol Deviations

The PP analysis set is the secondary efficacy analysis set. It is defined as all subjects in the ITT analysis set who have no major protocol deviations (which would potentially affect efficacy). The criteria for defining a major protocol deviation are described below in Section 5.6.1 and Section 5.6.2. The identification of specific subjects included and excluded (and reason for exclusion) for this analysis set will be made and documented prior to unblinding. Any other major deviation which is not pre-specified below, but results in a subject being excluded from the PP analysis set, will be specified in the protocol deviations document which is completed prior to unblinding.

The following protocol deviations are defined as 'major' and would exclude a subject from the PP analysis set (see Section 5.2). These deviation criteria can be split into those assessed prior to randomization relating to the protocol inclusion and exclusion criteria, and those assessed post randomization.

5.6.1. Major Deviations Assessed Prior to Randomization

- Inclusion criteria: #3-8.
- Exclusion criteria: #5 & 20.
- Randomization criteria: #1, 2, 3, 4.

5.6.2. Major Deviations Assessed Post-Randomization

- Rescue medication taken within 24 hours prior to the Week 16 visit.
- Prohibited medications that could affect pain and function assessments (protocol section 5.8.1) taken within 48 hours prior to Week 16 visit or within the minimum wash-out period specified by Appendix 4 of the protocol, for NSAID medications.
- Subjects who were <50% compliant with oral study medication between the baseline and the Week 16 visit.

In addition, unforeseen major protocol deviations may be added to this list. However the final definition of this criteria and the per-protocol population will be made prior to unblinding of this study.

6. ENDPOINTS AND COVARIATES

6.1. Efficacy Endpoint(s)

The primary efficacy endpoint is listed below.

- *Change from Baseline to Week 16 in the daily average Low Back Pain Intensity (LBPI) score as measured by an 11-point Numeric Rating Scale for tanezumab vs placebo.*

The secondary efficacy endpoints are listed below.

- *Change from Baseline to Week 16 in the Roland Morris Disability Questionnaire (RMDQ) for tanezumab vs placebo;*
- *Change from Baseline to Week 16 in the daily average LBPI score as measured by an 11-point Numeric Rating Scale for tanezumab vs tramadol PR.*
- *Change from Baseline to Weeks 2, 4, 8, 12, 24, 32, 40, 48, 56, and 64 in average LBPI score;*
- *Change from Baseline to Weeks 2, 4, 8, 16 (for tanezumab vs tramadol) 24, 32, 40,48, 56, 64 and 80 in RMDQ total score;*
- *Change from Baseline to Weeks 2, 4, 8, 16, 24, 32, 40, 48, 56, and 64 in Patient's Global Assessment of Low Back Pain;*
- *Cumulative distribution of percent change from Baseline in average LBPI score to Weeks 16, 24, and 56 (endpoint for summary only);*
- *Response as defined by a ≥30%, ≥50%, ≥70% and a ≥90% reduction from Baseline in daily average LBPI score derived from the subject diary at Weeks 2, 4, 8, 12, 16, 24, 32, 40, 48, and 56;*
- *Response as defined by a ≥30%, ≥50%, ≥70% and a ≥90% reduction from Baseline in the RMDQ score at Weeks 2, 4, 8, 16, 24, 32, 40, 48, and 56;*
- *Cumulative distribution of percent change from Baseline in RMDQ score to Weeks 16, 24, and 56 (endpoint for summary only);*
- *Change from Baseline to Weeks 2, 4, 8, 16, 24, 32, 40, 48, 56, and 64 in the Brief Pain Inventory-short form (BPI-sf) scores for Worst Pain, Average Pain, Pain Interference Index (composite function score), Pain Interference with General Activity, Pain Interference with Walking Ability, Pain Interference with Sleep, and Pain Interference with Normal Work;*
- *Chronic Low Back Pain Responder Index analysis [composite endpoint of average LBPI score, Patient's Global Assessment of Low Back Pain, and RMDQ total score at Weeks 2, 4, 8 16, 24, 32, 40, 48 and 56;*

- *Treatment Response: Improvement of ≥2 points in Patient's Global Assessment of Low Back Pain at Weeks 2, 4, 8, 16, 24, 32, 40, 48, and 564;*

- *Euro Quality of Life Health State Profile (EQ-5D-5LTM) dimensions and overall health utility score at Baseline, Weeks 8, 16, 24, 40, 56, and 64;*

- *Work Productivity and Activity Impairment Questionnaire: Low Back Pain (WPAI:LBP) change from Baseline to Weeks 16, 56, and 64 or Early Termination, in the percent work time missed due to chronic low back pain, percent impairment while working due to chronic low back pain, percent overall work impairment due to chronic low back pain, and percent activity impairment due to chronic low back pain;*

- *Incidence of and time to discontinuation due to lack of efficacy;*

- *Usage of rescue medication (incidence, and number of days of usage) during Weeks 2, 4, 8, 12, 16, 24, 32, 40, 48, 56 and Week 64;*

- *Usage of rescue medication (amount taken) during Weeks 2, 4, 8, 12 and 16;*

- *Health Care Resource Utilization at Baseline, Weeks 64, and 80.*

Among above secondary endpoints, the following are key secondary endpoints:

- Change from Baseline to Week 16 in the Roland Morris Disability Questionnaire (RMDQ) for tanezumab vs placebo;

- Response as defined by a ≥50% reduction from Baseline in daily average LBPI score derived from the subject diary at Week 16 for tanezumab vs placebo;

- Change from Baseline to Week 2 in average LBPI score for tanezumab vs placebo.

The secondary treatment satisfaction endpoints are listed below.

- *Treatment Satisfaction Measures: Treatment Satisfaction Questionnaire for Medication v.II (TSQM) score at Weeks 16 and 56;*

- *Patient Reported Treatment Impact Assessment-Modified (mPRTI) at Weeks 16 and 56.*

The tertiary endpoints are listed below.

- *Plasma tanezumab concentrations; (PK);*

- *Serum NGF assessments; (PD);*

- *Serum and urine osteoarthritis biomarker concentrations; (PD);*

- *NIH (National Institutes of Health) Pain Consortium Chronic Low Back Pain (CLBP) Minimum Dataset Baseline at Baseline, Weeks 16 and 56.*

For the definition of the Chronic Low Back Pain Responder Index at a certain week, a responder is defined as:

- A reduction of ≥30% in mean daily average LBPI from baseline to that week; and
- A decrease of ≥30% in Patient Global Assessment of low back pain from baseline to that week; and
- No worsening (increase) in Roland-Morris Disability Questionnaire total score from baseline to that week.

The BPI-sf Pain interference index is calculated as the mean of the seven BPI-sf Pain interference items (question 5a to g), being Pain interference with General Activity; Mood; Walking Ability; Normal work; Relations with other people; Sleep; Enjoyment of life.

6.2. Safety Endpoints

The following endpoints are included in the protocol (Section 2):

- *Standard safety assessments (safety laboratory testing [chemistry, and hematology], sitting vital signs, electrocardiogram [ECG 12-lead]);*
- *Orthostatic (supine/standing) blood pressure assessments;*
- *Survey of Autonomic Symptoms scores;*
- *Joint safety adjudication outcomes;*
- *Total joint replacements;Neurologic examination (Neuropathy Impairment Score [NIS]);*
- *Anti-drug antibody assessments (ADA);*
- *Physical examinations.*

The adverse events of Abnormal Peripheral Sensation (APS) are defined in the table below.

Allodynia	Neuritis
Axonal neuropathy	Neuropathy peripheral
Burning sensation	Paraesthesia
Carpal Tunnel Syndrome	Paraesthesia oral
Decreased Vibratory Sense	Peripheral sensorimotor neuropathy
Demyelinating polyneuropathy	Peripheral sensory neuropathy
Dysaesthesia	Polyneuropathy
Formication	Polyneuropathy chronic
Hyperaesthesia	Sensory disturbance
Hyperpathia	Sensory loss
Hypoaesthesia	Thermohypoaesthesia
Hypoaesthesia oral	Sciatica
Intercostal neuralgia	Tarsal Tunnel Syndrome
Neuralgia	

Adverse Events of Sympathetic Nervous System are defined in the table below.

Abdominal discomfort	Micturition urgency
Anal incontinence	Nausea
Anhidrosis	Nocturia
Blood pressure orthostatic decreased	Orthostatic hypotension
Bradycardia	Pollakiuria
Diarrhoea	Presyncope
Dizziness postural	Respiratory distress
Early satiety	Respiratory failure
Ejaculation delayed	Sinus bradycardia
Ejaculation disorder	Syncope
Ejaculation failure	Urinary hesitation
Heart rate decreased	Urinary incontinence
Hypertonic bladder	Vomiting
Hypohidrosis	

A smaller set of the above Adverse Events (to be called AEs of Decreased Sympathetic Function) may also be summarized. These are defined below.

Anhidrosis	Orthostatic hypotension
Bradycardia	Syncope
Hypohidrosis	

The lists given above may be updated depending on any additional adverse events observed in any tanezumab study. There are a number of summaries based on these groupings of adverse events.

The Neuropathy Impairment Score (NIS) is the sum of scores over all 37 items from both the Left and Right side. Items 1-24 are scored on a 0-4 scale (0, 1, 2, 3, 3.25, 3.5, 3.75, 4) and items 25-37 are scored on a 0-2 scale (0, 1, 2). The possible range of the NIS is 0-244.

A 3-tier approach will be used to summarize AEs. Under this approach, AEs are classified into 1 of 3 tiers. Different analyses will be performed for different tiers. A description of the three tiers and analyses are given in Section 8.2.3.

The definition of orthostatic hypotension to be summarized is:

- For patients with Baseline supine systolic Blood Pressure ≤150 mmHg:
  - Reduction in sBP (standing minus supine) ≥20; OR
  - Reduction in dBP (standing minus supine) ≥10.
- For patients with Baseline supine systolic Blood Pressure >150 mmHg:
  - Reduction in sBP (standing minus supine) ≥30; OR
  - Reduction in dBP (standing minus supine) ≥15.

The Survey of Autonomic Symptoms (SAS) is a 12 item (11 for females) questionnaire. From this the total number of symptoms (0-12 for males and 0-11 for females) will be calculated. Where a patient has a symptom, the impact of that symptom is then rated from 1 ('not at all') to 5 ('a lot'). The total impact score is calculated using this 1-5 scale, with 0 assigned where the patient does not have the particular symptom. The range for the total impact score is 0-60 for males and 0-55 for females.

6.3. Other Endpoints

6.3.1. PK Endpoints

- Plasma tanezumab concentrations.

6.3.2. PD Endpoints

The following assessments will be made:

- Serum NGF concentrations;
- Serum and urine osteoarthritis biomarker concentrations.

6.3.3. Outcomes Research Endpoints

Outcome research endpoints include the five dimensions (mobility; self-care; usual activities; pain/discomfort; anxiety/depression) and overall health utility score from the EuroQol 5 Dimensions (EQ-5D-5L), and the EQ-VAS. The overall health utility score is calculated using the EuroQol value sets, and is described in Appendix 1.4

The change from Baseline in the impairment scores of the Work Productivity and Activity Impairment Questionnaire for Low Back Pain (WPAI:LBP) to be summarized are listed below:

- Percent work time missed due to Low Back Pain.
- Percent impairment while working due to Low Back Pain.
- Percent overall work impairment due to Low Back Pain.
- Percent activity impairment due to Low Back Pain.

The calculation of these endpoints is described in Appendix 1.4.

The 11 questions of the TSQM are used to calculate the 4 endpoints of Effectiveness, Side Effects, Convenience and Global Satisfaction, each scored on a 0-100 scale with 100 being the best level of satisfaction. The calculation of these parameters is described in Appendix 1.4.

6.3.4. Total Joint Replacement and Surgical Endpoints

Adjudication outcomes to be summarized include rapidly progressive osteoarthritis (type-1 only), rapidly progressive osteoarthritis (type-2 only), rapidly progressive osteoarthritis (type-1 or type-2 combined), subchondral insufficiency fracture, primary osteonecrosis, and pathological fracture). Total joint replacements will also be summarized.

Events will be included in summaries if they occur up to the end of the safety follow-up period or 26 weeks (planned duration of the follow-up period + 2 weeks) after the end of the treatment period, whichever is later.

Total joint replacement events including surgery will be reported.

6.4. Covariates

For all models analyzing the continuous primary and secondary efficacy endpoints (except rescue medication) the corresponding Baseline value will be used as a covariate, together with the Average Low Back Pain Intensity (aLBPI, for endpoints other than the aLBPI). Study site will be fitted as a random effect in the ANCOVA models.

For the models analyzing the amount and number of days of rescue medication use the model will include term for Baseline aLBPI.

For response endpoints relating to aLBPI, the Baseline aLBPI score will be used as a covariate. For response endpoints relating to the Patients Global Assessment of Low Back Pain (PGA-LBP) the Baseline aLBPI and PGA will be used as covariates. For the Chronic Low Back Pain Responder Index (CLBP-RI) the Baseline aLBPI, PGA and RMDQ will be used as covariates.

Additional analyses of the primary endpoint will examine the treatment interactions with corresponding Study site and Country.

6.5. Subgroup analyses

Analysis of the data relating to Total Joint Replacement surgery will be performed using the TJR subset analysis set. As described in Section 8.2.3, the analysis for these tables will be described in a separate Statistical Analysis Plan, but reported within the tables for this study.

Separate tables will be produced for patients in Japan and European Union (EU). These tables will be defined prior to the unblinding of the study.

The analysis of the primary endpoint will be performed for (1) by region (European vs Non-European), and (2) by Baseline painDetect category (painDETECT score <=12, 13 to 18, >=19). For these analyses, the ANCOVA model with terms of baseline LBPI score and treatment will be used. The same analysis will also be performed by baseline pain severity (LBPI <7 vs LBPI $\geq$ 7).

The analysis of the key secondary endpoint, Change from Baseline to Week 16 in the Roland Morris Disability Questionnaire (RMDQ) for tanezumab vs placebo, will be performed by baseline pain severity (LBPI <7 vs LBPI $\geq$ 7) as well. For this analysis, the ANCOVA model with terms of baseline RMDQ, baseline LBPI, and treatment will be used.

7. HANDLING OF MISSING VALUES

The primary efficacy endpoint is the change from Baseline to Week 16 in the aLBPI. The primary analysis of the primary endpoint will use multiple imputation for missing data at Week 16 (where the method for imputation will be dependent on the reason for missing data) followed by the ANCOVA analysis with the model described below for the multiple imputed datasets. The imputation strategies are described in the following table. While the table describes the multiple imputation strategy specifically for the Week 16 time point, multiple imputation analysis at other time points will use the same strategy but with the appropriate time point, eg, 'Week 2,' substituted for 'Week 16' in the table below. Efficacy data missing from windows after the Week 56 window, eg, Week 80, will not be imputed for any summary or analysis unless otherwise indicated.

Type of Missing Data	Imputation Method
Missing data resulting from discontinuation of treatment due to Death, Adverse Events (AEs) or Insufficient Clinical Response or Patient's Meeting Protocol-Specified Pain Criteria for Discontinuation prior to or during the Week 16 visit reporting window*.	Multiple imputations will be created by sampling from a normal distribution based on the subject's baseline score and the standard deviation (over all treatment groups) of the observed efficacy data at Week 16 overall ITT subjects. This is a multiple imputation version of BOCF single imputation method.
Missing data for other reasons, ie, • Subject did not discontinue on or before Week 16 (includes discontinuation for any reason after the end of the Week 16	Multiple imputations will be created by sampling from a normal distribution based on the subject's last score and the standard deviation (over all treatment groups) of the

visit reporting window*), • Subject discontinued for a different reason prior to or during the Week 16 visit reporting window*.	observed efficacy data at Week 16 over all ITT subjects. For example if last observation for a subject is at Week 12, then the imputation sample for that subject is created using the subject's Week 12 observation and the standard deviation of the Week 16 observations for all subjects. Note: a subject's last observation may be the Baseline observation. This is a multiple imputation version of LOCF single imputation method.

* See Appendix 1.1 for a definition of the reporting windows.

The imputation of baseline-like data for subjects with missing data due to discontinuation due to Death, AE or LoE is intended to impute conservative efficacy values for those subjects who discontinue because of a reason that is considered to be a poor outcome for the subject, and so a poor outcome is imputed. For those subjects with missing data that is likely to not be related to treatment group, the intention is that missing data should be imputed based on a 'missing at random' assumption taking into account the subject's previous available data.

One hundred imputed samples will be used in this analysis. In order to pre-define the analysis (and not to allow the results to change if run again), the following seeds will be used in the creation of the multiple imputed data: aLBPI: 1001-1100; RMDQ: 2001-2100; PGA-LBP: 3001-3100; BPI-sf scores for Worst Pain, Average Pain, Pain Interference Index (composite function score), Pain Interference with General Activity, Pain Interference with Walking Ability, Pain Interference with Sleep, and Pain Interference with Normal Work: 4001-4100. Imputed Week 16 data for the PGA-LBP will be rounded to integer scores in the range 1 to 5. Imputed Week 16 data for the aLBPI that are <0 and >10 will be truncated to 0 and 10, respectively. The ANCOVA analysis described in Section 8.1.1 (with covariates are in Section 6.4) will be used for each imputation dataset, and the overall results will be calculated to take account of the variability both within and between imputation datasets.

This analysis will be used for the primary efficacy endpoint at Week 16, plus secondary analyses at other time points up to Week 56, and also for a range of secondary efficacy endpoints at all time points up to Week 56.

Three additional methods will explore the sensitivity of the effect of missing data. The first method of Baseline Observation Carried Forward (BOCF) for missing data at the primary time point of Week 16 will impute the subject's Baseline value for the Week 16 time point, and therefore a zero change from baseline. If a subject's baseline data is also missing then that subject's data remain missing for the post-baseline time point. The second method of Last Observation Carried Forward (LOCF) for missing data at the primary time point of Week 16 will impute the subject's last observed data value for the efficacy endpoint. With LOCF, if a subject is missing all post-baseline efficacy data for a given efficacy endpoint, then baseline will be carried forward (if baseline is missing then the subject would have no contributing data to be included in the analysis). In both the BOCF and LOCF imputation analyses, the same main effects ANCOVA model as described below will be used. The third

method will use Mixed Model for Repeated Measurements (MMRM) utilizing all observed data up to and including Week 16 (see Appendix 1.1 for details on windows).

Analyses of the primary endpoint aLBPI at secondary time points up to Week 56, of the secondary endpoint RMDQ, and of PGA-LBP at all time points up to Week 56, will use the BOCF and LOCF imputation methods for missing data, and use the same (main effects) ANCOVA model as described for the primary analyses.

The responder endpoints will be analyzed using logistic regression for binary data, using both BOCF and LOCF separately for missing data of the response endpoint at a particular time point. Imputation using BOCF will lead to the subject being assessed as a non-responder. In addition, in order to closely match the primary imputation analysis, a mixed BOCF/LOCF imputation for response endpoints will be used. In this analysis BOCF imputation (ie, a subject would be a non-responder) would be used for missing data due to discontinuation for reasons of lack of efficacy, adverse event or death up to the time point of interest, and LOCF imputation would be used for missing data for any other reason.

Note, if Baseline is missing then the subject data for the change from Baseline will be set to missing for all efficacy analyses for that parameter. A subject who has a missing Baseline score will be missing for the response criteria for endpoints where the response is based on one parameter. The CLBP-RI is based on 3 parameters (aLBPI, RMDQ and PGA-LBP). It is set to missing if any one out of these three parameters are missing at baseline or for the relevant post-Baseline timepoint (per its definition, a response can still be achieved if all three component parameters meet certain criteria).

For the RMDQ, any missing items are treated as the patient not having that symptom, and so not included in the RMDQ total score.

The BPI-sf Pain interference index score is calculated from the seven BPI-sf pain interference items. The composite index score is calculated as the mean of the non-missing items as long as ≥4 of the 7 items are non-missing, otherwise (≤3 items non-missing) the index score is missing.

For the analysis of the rescue medication endpoints while subjects are still in the study any missing data will be imputed by carrying forward the last recorded daily data up to Week 16 (LOCF daily data). Imputation using the daily data will occur up to the end of the last week when the subject is in the study (see Appendix 1.1 for definitions of the last study day in each week). For example if a subject discontinues on study day 10, then data up to the end of Week 2 will be imputed in this way. The weekly scores for the rescue medication endpoints can then be calculated for each week the subject is in the study. Rescue medication endpoints are summarized and analyzed using LOCF, and so the last weekly score for the rescue medication will be used for LOCF after the subject has discontinued from the study (note, imputation is taken from the last week with non-missing data and not necessarily from the last available study week, eg, if Week 8 is missing then Week 7 data can be used). The baseline observation will not be carried forward in the case where a post-baseline observation is not available for the LOCF imputation. In the example above, the subject who discontinued in Week 2 (Study Day 10) will have their Week 2 value used as the LOCF

value for all Weeks 3-16. The BOCF imputation rule will not be used for the subject because rescue medication is collected during the Initial Pain Assessment Period only (days -5 to -1) and subjects should not be taking rescue medication within 24 hours of the Baseline visit (so part of day -1), therefore Baseline rescue medication use is not an accurate reflection of subjects true Baseline use of rescue medication. Imputation of weekly diary data after Week 16 will use LOCF based on the last available weekly diary data score available.

The electronic diary data is a mix of daily and weekly average pain assessments, although the recall assessment period is the past 24 hours for both daily and weekly assessments. A weekly mean score will be calculated from the available daily pain scores where that is available. Any missing daily pain scores will be left as missing in the weekly pain score calculated. If there are no non-missing observations then the weekly score will be missing. The Baseline mean will be calculated using equivalent rules from potential five values of the Initial Pain Assessment Period (IPAP). The weekly pain scores (either calculated from the daily scores when available or directly from the weekly pain assessments) will then be utilized for the multiple imputation, and the LOCF and BOCF imputations in the standard way. Note, for the weekly pain score, a pain score being carried forward with LOCF might not be a visit week assessment (eg, carry forward Week 3 for missing Week 4 data). For the purposes of the imputation analyses, where there is no post-baseline observation available to carry forward, then the baseline score carried forward will be the baseline average pain score, being the mean of the expected five pain scores in the baseline assessment period. If any of the baseline average pain scores are missing (or there are less than 5 pain scores) then the baseline is calculated over the remaining non-missing values.

Missing values in standard summaries of AEs, lab values, vital signs and ECGs will be handled per Pfizer standard algorithms. For the analysis of NIS the Baseline observation will not be carried forward in the case where a post-baseline observation is not available for the LOCF imputation.

8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

8.1. Statistical Methods

8.1.1. Analyses for Continuous Data

The mixed model ANCOVA will be used with continuous change from Baseline endpoints for landmark (single time point) analyses. The model will include the covariates described in Section 6.4, including study site as a random effect. Estimates of treatment effects and pair wise treatment comparisons will be done using least squares means (LS means) and 95% confidence intervals (CI).

Under the primary analysis using multiple imputation defined in Section 7, the multiple ANCOVA results will be combined using standard methods (Little & Rubin, 2002), which are described in Appendix 2.2.

8.1.2. Analyses for Categorical Data

Statistical approaches for analyses of categorical data are included Section 8.2.2 and Section 8.2.3.

8.1.3. Analyses for Binary Endpoints

Binary response parameters, and the incidence of rescue medication use and discontinuation from treatment due to lack of efficacy will be analyzed using logistic regression for binary data, with covariates described in Section 6.4. Output will show the number and percentage of subjects in each response category, and odds ratio's (with 95% CIs) for the treatment comparisons shown in Section 4.1.

8.2. Statistical Analyses

A modified treatment-policy estimands strategy is applied as the main strategy to assess effectiveness of tanezumab. Data collected will be included for efficacy assessment regardless of rescue medication being used or not.

The general study design for efficacy includes a planned treatment period through the Week 56 visit, and a planned 24-week post-treatment safety follow-up period. For patients who discontinue early from treatment, they are followed in a 24-week early termination safety follow-up. Efficacy data planned to be collected during the post-treatment safety follow-up period or during the early termination safety follow-up period are intended to measure response a significant time after treatment has been discontinued. They are not intended to assess treatment effects or compare treatment groups. All efficacy data collected up to Week 80 will be summarized (where available), and efficacy data up to Week 56 will be analyzed.

All efficacy assessments during the treatment period are made on the analysis windows defined in Appendix 1.1. Using these windows we find the analysis window for a patient's last subcutaneous (SC) dose. Any data included in a window that is up to 8 weeks from this last SC dose window is 'on-treatment', and any data included in a window that is more than 8 weeks after the last SC dose window is off treatment. Data in on-treatment analysis windows will be used in summaries and analyses, while data in off-treatment analysis windows will be excluded from all summaries and analyses of treatment period efficacy data, ie, up to Week 56.

Efficacy data collected via subject diary (aLBPI and rescue medication use) are collected daily or weekly, not at study visits. Diary efficacy data will be considered on-treatment if it is collected up to 84 calendar days after the last SC dose. Diary efficacy data collected more than 84 calendar days after the last SC dose will be considered off-treatment and excluded from summaries and analyses of treatment period efficacy data.

Subjects are randomized at Baseline to one of five treatment groups: placebo→tanezumab 5mg (at Week 16), placebo→tanezumab 10mg (at Week 16), tanezumab 5mg, tanezumab 10mg, and tramadol. Unlesss otherwise noted, efficacy data up to Week 16 will summarized and analyzed by 4 groups: placebo, tanezumab 5mg, tanezumab 10mg, and tramadol, where placebo includes placebo→tanezumab 5mg (at Week 16) and placebo→tanezumab 10mg (at Week 16). Efficacy data up to Week 80 will be summarized by 5 groups. For analyses after Week 16 where multiple imputation is used, only 3 groups (tanezumab 5mg, tanezumab 10mg, and tramadol) will be included. This is because patients who are randomized to placebo→tanezumab 5mg (at Week 16) or placebo→tanezumab 10mg (at Week 16) will

receive placebo for the first 16 weeks, and their data before week 16 will not be imputed into analyses after week 16. However, a supplementary analysis will be performed for aLBPI, RMDQ, and PGA-LBP with patients who are on placebo for the first 16 weeks and dosed at Week 16 being combined with either tanezumab 5mg and tanezumab 10mg per their dose assignment at Week 16.

8.2.1. Primary Analysis

The primary analysis for the primary endpoint will use ANCOVA with covariates of Baseline aLBPI score, Treatment Group, and study site as a random effect. The primary analysis set is the ITT analysis set. The primary analysis will use multiple imputation as described in Section 7, and analysis using the ANCOVA model, with combination of results from imputation analyses using standard methods as described in Section 8.1.1.

Primary Endpoint Sensitivity Analyses

A number of sensitivity analyses will be performed on the primary efficacy endpoint in order to assess the robustness of the conclusions for the primary objective. These relate to the analyses for missing data and the analysis population, the homogeneity of the results across factors that may influence efficacy, and influence of covariates. The analyses described below will not be subject to the testing strategy described for multiple comparisons of the primary analysis (given in Section 4.2). As such, assessment of all treatment comparisons will be made independent of results over the primary endpoint or the specific treatment comparisons for each analysis.

The ITT analysis set is used in the analyses numbered 2 to 4 below, and Per-Protocol analysis set used in analysis number 1 below.

1. Per-Protocol Analysis Set.

The primary analysis described above will be repeated, but using the Per-Protocol analysis set (as described in Section 5.2 and 5.6) in place of the Full analysis set. This analysis will assess the robustness of the efficacy conclusions to subjects who have more strictly adhered to protocol inclusion and exclusion criteria, and to protocol defined study procedures.

2. Alternative Missing Data Analyses.

There are three additional analyses that will assess the robustness of the efficacy conclusions to the choice of multiple imputation as the primary method for accounting for missing data. These analyses are described in detail in Section 7.

In the first and second analyses, the primary ANCOVA analysis model described in Section 8.1.1 will be repeated, but using BOCF and LOCF respectively for missing data (note these are single imputation analyses).

In the third analysis, the mixed model Repeated Measures (MMRM) analysis will be performed, with covariate terms for Time (study week, treated as a categorical variable), Treatment Group and Time-by-Treatment interaction, as well as the covariates described in Section 6.4. The unstructured covariance will be used in the modeling of the within-subject errors in the analysis. Even though this is a sensitivity analysis for the primary endpoint, estimates for the time points of Weeks 2, 4, 8 and 12, in addition to Week 16 will be shown from this analysis. This analysis will use the observed data up to Week 16.

A summary of the missing data pattern will be shown for the aLBPI for Baseline and Weeks 2, 4, 8, 12 and 16. This summary will show the incidence of subjects with each pattern of observed and missing data over these visits. This summary will be shown overall, and split by treatment group.

3. Sensitivity Analysis Excluding Subjects from Sites with Potential GCP Compliance Issues and Excluding Subjects Who May Have Been Enrolled Twice.

During the conduct of the study, it was identified that there may be GCP compliance issues at Sites PPD and PPD. Also one subject may have been enrolled twice at two different sites in the study (Patient ID =PPD and Patient ID=PPD). A sensitivity analysis will be performed excluding the patients from both sites and the patient who was potentially enrolled twice, using the same approach as the primary analysis for the primary endpoint.

4. Interaction Analyses.

There will be two analyses to explore the interaction of treatment with Study Site or Country. Estimates for Study sites and Countries with 40 or more subjects will be shown. Estimates will be shown within each level of each factor, ie:

- Study Site: Individual sites with ≥40 subjects in the ITT set.
- Country: Individual countries with ≥40 subjects in the ITT set.

Interaction analyses for the primary endpoint is performed to explore the effect of Study site and Country. These analyses will fit the covariate terms described in Section 6.4 (except in the Country interaction analysis, Study site will be replaced by Country as a fixed term), in addition to the interaction term of treatment group by factor. The Treatment by Study site interaction term will be fitted as a random effect. The Treatment by Country interaction term will be fitted as a fixed effect as there are a limited number of counties in the study.

The study sites to be examined in this way will be any site with an average of ten or more subjects per treatment group within the site (approximately 9 for placebo and tanezumab treatment groups and 13 for tramadol), which for this study relates to any site with 40 or more subjects in total. This assessment will be made prior to unblinding, therefore a study site in this group may still have fewer than approximately ten subjects in one or more of the treatment groups, however that site will still be included in this summary of efficacy of the larger study sites. This rule will also be applied to the Treatment by Country interaction analysis.

To aid the interpretation of the treatment-site and treatment-country interactions, a summary of the efficacy data for the primary endpoint by treatment group will be shown for the sites with ≥40 subjects and for the countries with ≥40 subjects over all treatment groups.

8.2.2. Secondary Analyses

The following secondary endpoint analyses support the primary endpoint in the assessment of efficacy. All analyses in this section use the ITT analysis set only. Unless otherwise stated, efficacy data will be summarized up to Week 80, and analyzed up to Week 56. The assessments for efficacy at and after Week 64 are at least 16 weeks after the last SC dose, and so are 'off treatment' for tanezumab/placebo SC treatment, and as such there is no hypothesis testing for the Week 64 or Week 80 data. Any efficacy data collected at the Week 64 or Week 80 Safety follow-up visit will be excluded from analyses of efficacy.

1. Other time points for the primary efficacy measure.

The ANCOVA model described above for the primary endpoint, using covariates of Baseline aLBPI score, Treatment, with Study Site as a random effect, will be used in the analysis of aLBPI for the change from Baseline to Days 1, 2, 3, 4, 5, 6, and 7, and to Weeks 1, 2, 4, 8, 12, 24, 32, 40, 48 and 56. This analysis will be produced using multiple imputation (as described in Sections 7 and 8.1.1), and BOCF and LOCF for missing data.

5. Secondary endpoints analyzed using ANCOVA.

The ANCOVA model described above for the primary endpoint using covariates of Baseline score, Baseline aLBPI and Treatment, with Study Site as a random variable, will be used in the analysis of RMDQ, PGA-LBP, and the BPI-sf questions (Worst Pain; Average Pain; Pain interference index; Pain interference with general activity, with walking ability, with sleep and with normal work) for the change from Baseline to Weeks 2, 4, 8, 16, 24, 32, 40, 48, and 56. This analysis will be produced using multiple imputation, and BOCF and LOCF for missing data for RMDQ and PGA-LBP. For others, only multiple imputation will be used.

Supplementary analyses with multiple imputation for missing data will be performed for aLBPI, RMDQ, and PGA-LBP with patients who are on placebo for the first 16 weeks and dosed at Week 16 being combined with either tanezumab 5mg and tanezumab 10mg per their dose assignment at Week 16.

6. Response and Incidence endpoints.

The response endpoints of CLBP-RI, improvement in PGA-LBP ≥2 and aLBPI ≥30, 50, 70 and 90% improvements are analyzed using logistic regression with covariates as defined in Section 6.4, for response at Weeks 2, 4, 8, 12 (aLBPI response only), 16, 24, 32, 40, 48, and 56. These analyses use mixed BOCF/LOCF. For CLBP-RI and aLBPI ≥30, 50, 70 and 90% improvements, the analyses will also be performed using BOCF and LOCF for missing data. The use of BOCF for missing data implies subjects with missing data are included in the analysis as non-responders. Similarly the use of LOCF in the case where subjects have no post-Baseline data (and Baseline would be carried forward) again implies those subjects are included in the analysis as non-responders.

Incidence of rescue medication use will be analyzed using the logistic regression as described above up to Week 56, but only using LOCF imputation for missing data. That is, the incidence of rescue medication use will be analyzed using logistic regression for binary data, with covariates described in Section 6.4.

The incidence of discontinuation from treatment due to 'INSUFFICIENT CLINICAL RESPONSE' and 'PATIENT MEETS PROTOCOL-SPECIFIED PAIN CRITERIA FOR DISCONTINUATION' on the End of Treatment Subject Summary Case Report Form will be analyzed using logistic regression with model terms of Baseline aLBPI and treatment group, and using incidence up to the end of treatment period (the Week 56 visit or early termination). Discontinuation in the post-treatment safety follow-up period will not be included in this endpoint for analysis, but will be summarized as part of the safety tables. A supplementary analysis for the incidence of discontinuation from treatment due to 'INSUFFICIENT CLINICAL RESPONSE' will also be performed.

The cumulative aLBPI response at Weeks 16, 24 and 56 using response definitions from a reduction of >0% to =100% (in steps of 10%) will be summarized, using mixed BOCF/LOCF, and also LOCF and BOCF imputation for aLBPI. Imputation with BOCF for subjects with missing data at that timepoint will lead to the subjects being assessed as non-responders for the response endpoint.

Similar response analyses will be provided for RMDQ scores at Weeks 2, 4, 8, 16, 24, 32, 40, 48, 56, 64 and 80. The cumulative RMDQ response at Weeks 16, 24 and 56 using response definitions from a reduction of >0% to =100% (in steps of 10%) will be summarized, using mixed BOCF/LOCF, LOCF and BOCF imputation for RMDQ.

7. Time to Event.

The time to discontinuation from treatment due to'INSUFFICIENT CLINICAL RESPONSE' and 'PATIENT MEETS PROTOCOL-SPECIFIED PAIN CRITERIA FOR DISCONTINUATION' on the End of Treatment Subject Summary Case Report Form will use the log-rank test. Survival curve estimates (time to 1st, 2nd, 5th, 10th and 25th percentiles, and minimum and maximum values) and a plot of the time to discontinuation (failure) will be shown using the Kaplan-Meier estimates. Only discontinuation up to the end of treatment period (Week 56 visit or early discontinuation) will be used in this analysis. Discontinuation due to lack of efficacy after the end of treatment visit will be included in the standard safety tables. Imputation of time to event for discontinued subjects (discontinuing for reasons other than lack of efficacy) prior to the Week 56 visit uses censoring at the time of discontinuation. Imputation of time to event for completed subjects or discontinued subjects (for any reason) after the Week 56 visit uses censoring at the Week 56 visit time point. A supplementary analysis for the time to discontinuation from treatment due to 'INSUFFICIENT CLINICAL RESPONSE' will also be performed.

8. Number of Days and Amount of Rescue Medication Use.

The rescue medication data will be converted to Weekly scores for the week prior to the timepoint of interest. Calculation of the endpoints for both the IPAP and the concomitant medication log data collection is described in Appendix 1.4.

The number of days and amount of rescue medication endpoints will be analyzed using the Negative Binomial model, with model terms for Baseline aLBPI score and Treatment Group. In this model the error term is defined with a negative binomial distribution, and 'log' is used as the link function. Output from this analysis will be the estimated number/amount of rescue medication use per week in each treatment group, and (following the exponential back transformation) the ratio of rescue medication use for the treatment comparisons shown in Section 4.1. The 95% CIs will be given for the estimates of both the individual treatment groups and the treatment group ratio's.

The number of days of rescue medication use per week endpoint will be analyzed for the Weeks 2, 4, 8, 12, 16, 24, 32, 40, 48, and 56. The amount of rescue medication use per week will be analyzed for the Weeks 2, 4, 8, 12 and 16. Missing data will be imputed using LOCF only. For this analysis, Baseline data will not be carried forward in the case of a post-Baseline observation not being available for use in LOCF. In addition, imputation using BOCF will not be performed.

9. CMH test for PGA.

The change from Baseline in the PGA-LBP to Weeks 2, 4, 8, 16, 24, 32, 40, 48 and 56 will also be analyzed using the CMH test. Changes by each level of improvement will be summarized, as well as any improvement (change<0), no change (change=0), any worsening (change>0). This analysis will provide a sensitivity analysis for the ANCOVA analysis of the PGA. The missing data imputation used for this analysis will be mixed BOCF/LOCF.

10. EuroQol 5 Dimensions (EQ-5D).

The Baseline and Weeks 8, 16, 24, 40, 56 and 64 responses in the five dimensions (mobility; self-care; usual activities; pain/discomfort; anxiety/depression) and overall health utility score from the EuroQol 5 Dimensions (EQ-5D) will be summarized by treatment group. This summary will use observed data only (no imputation for missing data). The calculation of the overall health utility score is described in Appendix 1.4.

An additional question, called the EQ-VAS asks the patient to rate their health today using a VAS scale from 0 (the worst health you can imagine) to 100 (the best health you can imagine). This will be summarized along with the health utility score.

A table showing number and percentage of subjects will summarize the response for each dimension (item) of the EQ-5D at each timepoint. These summary tables will be shown by treatment group. In addition, for each treatment and each time point assessed, descriptive statistics (mean, standard deviation, median, number of subjects) will be shown for the health utility score, and the EQ-VAS measure of health today.

11. Work Productivity and Activity Impairment Questionnaire: Low Back Pain (WPAI:LBP).

The change from Baseline to Weeks 16, and 56 in the impairment scores of the Work Productivity and Activity Impairment Questionnaire: Low Back Pain (WPAI:LBP) will be summarized by treatment group.

This summary will use observed data only (no imputation for missing data). The calculations of these endpoints are described in Appendix 1.4.

The summary will show number and percentage of patients with a decrease, no change, and an increase in score for the change from Baseline to each timepoint, as well as descriptive statistics (mean, standard deviation, median, number of subjects) of the change in addition to the individual timepoints of Baseline and Weeks 16 and 56. The 4 WPAI parameters will be analyzed using the ANCOVA model described above for the primary endpoint using covariates of the corresponding Baseline score, Baseline diary average pain, and Treatment, with Study Site as a random effect.

12. Health Care Resource Utilization at Baseline, Weeks 64 and 80 will be summarized.

13. Treatment Satisfaction Questionnaire Medicine v.II (TSQM v.II).

The Weeks 16 and 56 responses in the 4 TSQM parameters of satisfaction with effectiveness, side effects and convenience, and overall satisfaction will be summarized. The 11 questions of the TSQM are used to calculate the 4 endpoints of Effectiveness, Side Effects, Convenience and Global Satisfaction, each scored on a 0-100 scale with 100 being the best level of satisfaction.

The four parameters of the TSQM will be analyzed using the Cochran-Mantel-Haenszel test at both Weeks 16 and 56. Summary tables showing number and percentage of patients by value and treatment group will be shown for TSQM items 1-11, and the four satisfaction parameters.

This summary and analysis will use observed data only (no imputation for missing data). The calculations of these endpoints are described in Appendix 1.4.

14. Patient Reported Treatment Impact Assessment-Modified (mPRTI) at Weeks 16 and 56.

The mPRTI is collected at Weeks 16 and 56. The two endpoints derived from this questionnaire are described below:

- Patient willingness to use drug again. This comes from the question “In the future, would you be willing to use the same drug that you have received in this study for your low back pain?” This is rated on a 5 point Likert scale from 1 (‘Yes, I would definitely want to use the same drug again’) to 5 (‘No, I definitely would not want to use the same drug again’).

- Patient preference of drug versus prior treatment. This comes from the question "Overall, do you prefer the drug that you received in this study to the treatment you received before this clinical trial?" This is rated on a 5 point Likert scale from 1 ('Yes, I definitely prefer the drug I am receiving now') to 5 ('No, I definitely prefer my previous treatment').

The two parameters of the mPRTI will be analyzed using the CMH test at both Weeks 16 and 56. Summary tables showing number and percentage of patients by value and treatment group will be shown for all mPRTI questions.

This summary and analysis will use observed data only (no imputation for missing data).

15. NIH Pain Consortium Chronic Lower Back Pain Minimum Dataset.

Responses by question and domain will be reported by treatment group for Baseline, Week 16 and 56. Categorical summaries of severity in the pain impact domain and other domains will be reported by treatment for Baseline, Weeks 16 and 56.

8.2.3. Safety Analyses

Unless specified otherwise, all summaries will be provided for data collected up to Week 16 and for the entire treatment period or for the entire study. For the summaries up to Week 16, data will be summarized by 4 treatment groups: placebo, tanezumab 5mg, tanezumab 10mg, and tramadol. For summaries for the entire treatment period, data will be summarized by 3 treatment groups: tanezumab 5mg, tanezumab 10mg, and tramadol, where tanezumab 5 mg (or 10 mg) will include patients who were randomized to placebo and received tanezumab 5 mg (or 10 mg) at Week 16 and patients who were randomized to tanezumab 5 mg (or 10 mg) at Baseline. For summaries for the entire study, data will be summarized by 4 treatment groups: placebo, tanezumab 5mg, tanezumab 10mg, and tramadol, where placebo will include patients who received only placebo in the treatment period, tanezumab 5 mg (or 10 mg) will include patients who were randomized to placebo and received tanezumab 5 mg (or 10 mg) at Week 16 and patients who were randomized to tanezumab 5 mg (or 10 mg) at Baseline.

For overall treatment-emergent adverse events, incidence of treatment-emergent adverse events, incidence of treatment-emergent adverse events leading to discontinuation from treatment, and incidence of treatment-emergent serious adverse events, summaries for the entire treatment period and for the entire study will be provided, in addition to the following:

a. Summaries up to the end of treatment by 4 treatment groups: placebo -> tanezumab 5 mg, placebo -> tanezumab 10 mg, tanezumab 5 mg, and tanezumab 10 mg, where placebo ->tanezumab 5 mg (placebo -> tanezumab 10 mg) includes only patients who were randomized to placebo -> tanezumab 5 mg (placebo -> tanezumab 10 mg) at Baseline and received tanezumab 5 mg (tanezumab 10 mg) at Week 16,

b. Summaries for the follow-up period by 4 treatment groups for: placebo, tanezumab 5mg, tanezumab 10mg, and tramadol, where placebo will include patients who received only placebo in the treatment period, tanezumab 5 mg (or 10 mg) will include patients who were randomized to placebo and received tanezumab 5 mg (or 10 mg) at Week 16 and patients who were randomized to tanezumab 5 mg (or 10 mg) at Baseline.

8.2.3.1. Safety Endpoints

1. Neuropathy Impairment Score.

The change from Baseline in the NIS for Weeks 8, 16, 24, 32, 40, 48, 56, 64 and 80 will be analyzed using a CMH test for 'row mean scores differ' with change score categories. Output will show number and percentage of subjects whose NIS score worsened (change>0), improved (change<0) or had no change, in addition to the mean (with standard deviation) and median change, and minimum and maximum change. Missing data will be imputed using LOCF only. For this analysis, Baseline data will not be carried forward in the case of a post-Baseline observation not being available for use in LOCF. An additional analysis will use the change from Baseline to the largest (worst) post-Baseline value.

2. Joint Safety and Surgical Outcome and Recovery Endpoints.

The incidence of subjects with any of the joint safety adjudication outcomes, including outcomes of rapidly progressive osteoarthritis (type-1 only), rapidly progressive osteoarthritis (type-2 only), rapidly progressive osteoarthritis (type-1 or type-2 combined), subchondral insufficiency fracture, primary osteonecrosis, and pathological fracture, and for occurrence of total joint replacement will be shown by number of subjects treated and subject years of observation.

For the joint safety event analyses, the observation period is defined as the time from first SC dose to study completion or discontinuation for subjects who did not have an event, or time from first SC dose to the earliest event for subjects who did have at least one event.

Reporting of total joint replacement events including surgery will be described in a separate Statistical Analysis Plan that will cover patients in this subset from Studies 1059, 1061 and 1063. Corresponding data from Studies 1056, 1057 and 1058 will be reported under study 1064. Summary listings for the data from the 1059 substudy will be part of the 1059 study CSR as there are <10 total joint replacement events expected in the substudy.

3. Other Safety Endpoints.

Pfizer standard safety data presentations will be made for demography data, discontinuation data, adverse event data, laboratory test data, vital signs data and ECG data.

For the 3-tier adverse event reporting, tier 1 adverse events are defined in the tanezumab Safety Review Plan, and this definition of tier-1 adverse events for the report of study 1059 tables will be finalized prior to the unblinding of this study.

Tier 2 AEs are those with a frequency of ≥3% in any treatment group that are not in tier 1.

Tier 3 AEs are those not in Tier 1 or Tier 2, and will be summarized using standard Pfizer data standards tables, where all Adverse Events will be included (ie, Tier 3 AEs will not be shown separately).

Treatment-emergent adverse events within tier 1 and 2 will be summarized using Risk Differences between each tanezumab group and tramadol, together with 95% confidence interval, using exact methods. Significance tests will be performed for the tier 1 adverse events. There will be no multiplicity adjustment for these significance tests. These tables will be produced for the comparisons of tanezumab 5mg versus tramadol and tanezumab 10mg versus tramadol. In addition, a summary of the adverse events which have an onset in the first 16 weeks of the study will be shown, for corresponding comparisons of tier 1 and 2 adverse events for the comparison of tanezumab 5mg versus placebo and tanezumab 10mg versus placebo.

The following footnote will be used in the Tier 1 AE tables: "P-values and confidence intervals are not adjusted for multiplicity and should be used for screening purpose only. 95% Confidence intervals are provided to help gauge the precision of the estimates for Risk Difference. Similarly the following footnote will be used in the Tier 2 AE tables: "Confidence intervals are not adjusted for multiplicity and should be used for screening purpose only. 95% Confidence intervals are provided to help gauge the precision of the estimates for Risk Difference."

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. As such, safety analysis is generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation. The 3-tier approach facilitates this exploratory analysis.

Pfizer standard safety data presentations will be made for demography data, concomitant medications, discontinuation data, adverse event data, laboratory test data, vital signs data and ECG data.

The following standard/non-standard safety tables will also be included

- A summary of baseline characteristics. This summary includes BPI-sf average pain and worst pain, RMDQ total score, PGA-LBP, Low back pain duration and etiology, Quebec Task Force (QTF) category, diabetes status (from medical history and/or pre-treatment HbA1c≥6.5%), and pain Detect score. This summary will also include a summary of the number of subjects who are ≥75 years old.

- Summaries of Baseline painDetect score in continuous scale (Mean/SD/Min/Max) and by painDetect category (painDetect score <=12, 13 to 18, >=19).

- The cumulative distribution function of painDetect score, overall and by region (EU vs Non-EU).

- Summary of number of patients treated by country and treatment group.

- Incidence and severity of adverse events leading to discontinuation.

- Summary of AEs, Incidence of AEs, Incidence of AEs leading to discontinuation and summary of Serious AEs. .

- Summary of evidence of neurological examination abnormalities by visit and final assessment, and incidence of neurological findings over consecutive visits. Further details of this summary are given below.

- Summary of final outcome of neurological consultation. Further details of this summary are given below.

- Summary of incidence of sympathetic neuropathy based on investigator assessment and, if performed, expert consultant assessment.

- 'Incidence and severity' tables of treatment-emergent adverse events of Abnormal Peripheral Sensation (APS), and Sympathetic Nervous Function, as defined above. Other adverse events may be added to these groupings if they are observed in this study or other studies in the tanezumab program.

- Summary table and listing of inclusion and exclusion criteria that are not met by subjects who were screened (but not treated); listing of subjects randomized but not treated.

- Summary of discontinuation by treatment group and reason, and study week of discontinuation for the treatment period (Weeks 1-2, 3-4, 5-8, 9-12, 13-16, 17-20, 21-24, 25-32, 33-40, 41-48, 49-56) and for the safety follow-up period (Weeks 1-8, 9-16, 17-24, >24).

- A summary of the maximum increase from baseline in the sitting systolic and diastolic blood pressure. The categories used are: (systolic BP) only decreases or no change, >0 to 10, >10-20, >20-30, >30-40, and >40, and (diastolic BP) only decreases or no change, >0 to 10, >10-20, >20-30, and >30.

- A summary of the maximum decrease from baseline in the sitting systolic and diastolic blood pressure. The categories used are: (systolic BP) <-40, -40 to <-30, -30 to <-20, -20 to <-10, -10 to <0, only increases or no change, and (diastolic BP) <-30, -30 to <-20, -20 to <-10, -10 to <0, only increases or no change.

- A summary of the change from baseline to last observation in the sitting systolic and diastolic blood pressure. The categories used for these summaries are: (systolic BP) ≤-40, >-40 to -30, >-30 to -20, >-20 to -10, >-10 to 0, >0 to <10, 10-<20, 20-<30, 30-<40, ≥40, and (diastolic BP) ≤-30, >-30 to -20, >-20 to -10, >-10 to 0, >0 to <10, 10-<20, 20-<30, ≥30.

- A summary of incidence of subjects with orthostatic hypotension (defined in Section 6.2 above), for each visit and any post-baseline incidence of orthostatic hypotension. An additional summary will be provided of outcomes of assessments resulting from an incident of orthostatic hypotension or other events of interest, using data from both the CRF database and the consultation database, as appropriate.

- A summary of discontinuation up to End of Treatment period, and up to End of Study period.

- Incidence of musculoskeletal physical examination by visit.

- Summary of the Survey of Autonomic Symptoms (SAS) number of symptoms reported and total symptom impact score, at each visit, and for the change from Baseline score (scores defined in Section 6.2 above).

- Summary of concomitant medications for Chronic Low Back Pain for non-NSAID and NSAID medications (shown separately).

- Summary of number of days of non-study NSAID use and mg dose amount of non-study NSAID use per dosing interval (eg, Baseline to Week 8 and Week 8 to Week 16, up to Week 56) and for the first 8-week interval in the safety follow-up period. This will show the number and percentage of subjects in a dosing interval who exceeded the limit of 10 days of NSAID use. If a dosing interval exists, the visits will be used to define the interval, otherwise calendar time (0-8 weeks, 9-16 weeks post last dose, etc.) will be used. A summary of average number of days of NSAID use and the mg dose amount of NSAID use will be displayed by dosing interval through the first 8 week of the follow-up period. Also, a summary of the overall total amount up to 16 weeks after the last SC dose will be shown, as well as the number and percentage of subjects who exceeded the limit of 80 days of NSAID use.

- Summary of failed drug treatments for protocol qualification, with reasons for discontinuation.

- Summary of oral study medication compliance. This is calculated for each interval Baseline to Week 2, Visit Weeks 2 to 4, 4 to 8, 8 to 16, 16 to 24, 24 to 32, 32 to 40, 40 to 48 and 48 to 56, as well as the entire post-Baseline period up to Week 56 (or end of treatment visit). Compliance is calculated as number of tablets dispensed minus the number returned divided by the number of days in the interval multiplied by 100, to get a percentage compliance for each patient for each time period.

Neurological-related safety data

The "conclusion from the neurological examination" data will be summarized for each timepoint and the final assessment over all neurological examinations for each subject. In addition the persistence of any neurological examination finding will be summarized, showing the incidence of subjects with new or worsened neurological examination abnormality (both clinically significant only and also for any finding) for 2, 3, 4 and ≥5 consecutive visits.

The number of subjects with at least 1 adverse event requiring neurologist consult, and whether or not a neurologic consult performed, and the expert primary diagnosis will be summarized.

Immunogenicity

The following assessments of ADA data will be made:

- A listing of individual serum ADA results sorted by treatment group, subject ID and planned visit. The listing will also include the actual test date/times.
- The proportion of subjects who test positive (ie, develop anti-tanezumab antibodies) and negative will be summarized by treatment group and planned visit. The summary will also include the proportion of subjects who test positive and negative overall in the study.
- Subjects who develop anti-tanezumab antibodies after treatment will be evaluated for the presence of anti-tanezumab neutralizing antibodies, and individual results will be listed.
- Individual subjects with positive ADA results will be evaluated for potential ADA impact on the individual's PK, efficacy and safety profile.

8.2.3.2. Pharmacokinetics

The following reporting of PK data will be done using all available data:

- A listing of all plasma tanezumab concentrations sorted by subject, active treatment group and nominal time post dose. The listing of concentrations will also include the actual times post dose.
- A descriptive summary of the plasma tanezumab concentrations based on nominal time post dose for each treatment group.
- Boxplots of tanezumab plasma trough concentrations at the nominal times for the tanezumab treatment groups.

8.2.3.3. Pharmacodynamics (NGF)

Serum samples from a subset of patients will be run in the bioanalytical assays for assessment of NGF and the measurements will be summarized in the following tables and figures.

- A listing of individual NGF concentrations sorted by subject, active treatment group and time post dose.
- Descriptive statistics of NGF concentrations based on time post dose for each treatment group.
- Boxplots of NGF over time post dose for each treatment group.

9. REFERENCES

1. Alosh M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Statist. Med.* (2014):33:693–713.

2. Bretz F, Posch M, Glimm E, Klingmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures usingweighted Bonferroni, Simes, or parametric tests *Biometrical Journal*. 53(2011) 6, 894–913.

3. Atkinson, MJ et al (2005). Hierarchical construct validity of the treatment satisfaction questionnaire for medication (TSQM Version II) among outpatient pharmacy consumers. *Value in Health*. **8(Supp 1)**, S9-24.

4. Atkinson, MJ et al (2005). Hierarchical construct validity of the treatment satisfaction questionnaire for medication (TSQM Version II) among outpatient pharmacy consumers. *Value in Health*. **8(Supp 1)**, S9-24.

5. EuroQol Group. EuroQol: a new facility for the measurement of health related quality of life. *Health Policy* 1990; 16:199-208.

6. Little RJ & Rubin DB (2002). *Statistical Analysis with Missing Data*. New Jersey: Wiley.

7. Tudor-Locke C, et al. A Catalog of Rules, Variables, Definitions Applied to Accelerometer Data in National Health and Nutrition Examination Survey, 2003-2006. *Prev Chronic Dis* 2012,9:110332. DOI: http://dx.doi.org/10.5888/pcd9.110332).

10. APPENDICES

Appendix 1. DATA DERIVATION DETAILS

Appendix 1.1. Definition and Use of Visit Windows in Reporting

Study visits are planned at Screening, Baseline and then at post-baseline Weeks 2, 4, 8, 16, 24, 32, 40, 48, 56, 64 and 80. If a subject discontinues from the trial then there will be an Early Termination Follow-Up period and for those who refuse, a complete Early termination visit. To account for this visit and any early or late scheduled visits (compared to the target study days) we define 'windows' to be able to allocate each efficacy observation to a single specific study visit. For the assessments made at study visits (eg, BPI-sf, PGA-LBP etc.) these visit windows are shown below. When multiple observations occur in a visit window, the observation closest to the protocol specified target day will be used, noting that the latter will be used in the case of a tie.

Visit	Target Study Day	Window
Screening [1]	Variable (up to 30 days prior to baseline visit)	[No lower limit, Day -6]
Baseline	1 (defined as initial day of study drug administration)	[-5,1]
Week 2	15	[2,22]
Week 4	29	[23,43]
Week 8	57	[44,85]
Week 16	113	[86,141]
Week 24	169	[142,197]
Week 32	225	[198, 253]
Week 40	281	[254, 309]
Week 48	337	[310, 365]
Week 56	393	[366, 421]
Week 64	449	[422, 477]
Week 80	561	[532, 589]

Besides the windows specified above, there is one additional window as defined as follows:

1. "16 Weeks after Last Dose". This window will include data from 16 ± 4 weeks past the date of the last SC dose. The target day is 113 days after the last SC dose, with a window of [86, 141] days after the last SC dose. If multiple observations occur in this visit window, the observation closest to the specified target day will be used, noting that the latter will be used in the case of a tie.

EQ-5D-5L is collected at Baseline, Weeks 8, 16, 24, 40, 56, and 64, and ET1 and ET2. WPAI:LBP is collected at Baseline, Week 16, 56, and 64, and ET1 and ET2. HCRU is collected at Week 64 and Week 80, and ET2 and ET3. TSQM and mPRTI are collected at Weeks 16 and 56, and ET1. The visit window for these visits will be defined below.

EQ-5D-5L

Visit	Target Study Day	Window
Baseline	1 (defined as initial day of study drug administration)	[-7, 1]
Week 8	57	[2, 85]
Week 16	113	[86, 141]
Week 24	169	[142, 225]
Week 40	281	[226, 337]
Week 56	393	[338, 421]
Week 64	449	[422, no upper limit]

WPAI: OA

Visit	Target Study Day	Window
Baseline	1 (defined as initial day of study drug administration)	[-7,1]
Week 16	113	[2,141]
Week 24	169	[142,281]
Week 56	393	[282, 421]
Week 64	449	[422, no upper limit]

Actigraphy, TSQM, mPRTI

Visit	Target Study Day	Window
Baseline	1 (defined as initial day of study drug administration)	[no lower limit, 1]
Week 16	113	[2,253]
Week 56	393	[254, no upper limit]

Data in on-treatment analysis windows will be used in summaries and analyses of treatment period efficacy data, ie, up to Week 56, while data in off-treatment analysis windows will be only included in summaries (at Week 64, Week 80, 16 Weeks after Last Dose, and End of Study). Data in off-treatment analysis windows will be imputed, if needed, in analyses.

Any data collected up to 8 weeks (per the windowing rule above) from the last SC dose window is 'on-treatment', and any data collected more than 8 weeks (per the windowing rule above) after the last SC dose window is off treatment. For example, for patients whose last SC dose is at Week 8 (per the windowing rule above), any data collected after Week 16 (per the windowing rule above) will be off-treatment. For the aLBPI, the data is collected daily up to the end of Week 16 via electronic diary and thereafter weekly up to Week 64. Data up to Week 64 will be reported as part of the efficacy assessment (summary up to Week 64;

analysis up to Week 56). The Baseline score is the mean of the non-missing pain scores over study days -5 to -1.

The table below describes the visit days for each week (weeks 1-16). All available on-treatment diary data in each of the weekly intervals will be used to calculate the mean daily pain score for that study week.

Study Week	Days	Study Week	Days
1	1-7	9	57-63
2	8-14	10	64-70
3	15-21	11	71-77
4	22-28	12	78-84
5	29-35	13	85-91
6	36-42	14	92-98
7	43-49	15	99-105
8	50-56	16	106-112

However, if a subject receives the Week 16 injection dose prior to Day 113, the Week 16 score will be calculated using the mean of the available scores from the 7 calendar days immediately prior to the Week 16 injection dose. Any scores used in this calculation of Week 16 will not also be used in an earlier week calculation, eg, if the Week 16 dose occurs on Day 109, the available scores from Days 102-108 will be used to calculate the average score for Week 16, and the available scores from Days 99-101 will be used to calculate the average score for Week 15.

After the Week 16 visit, weekly pain scores are captured weekly in the diary. These are grouped in 4-week intervals using visit windows as shown below. If a subject comes in late for a Week 16 visit (or weekly diary is not activated at the visit), and so has daily diary data collected past Day 112, these data will be averaged with any data obtained weekly for any given interval.

All available on- or off-treatment data will be used for these windows after the planned treatment period.

Summary Week	Includes Weeks	Days
20	17 - 20	113-140
24	21 - 24	141-168
28	25 - 28	169-196
32	29 - 32	197-224
36	33 - 36	225-252
40	37-40	253-280
44	41-44	281-308
48	45-48	309-336
52	49-52	337-364
56	53-56	365-392
60	57-60	393-420
64	61-64	421-448

When multiple observations occur in a day for pain score collected daily or in a week for pain score collected weekly, the average of scores will be used.

For rescue medication, data are collected until the end of study (Week 80 or final ET). Additional windows are defined below

Summary Week	Includes Weeks	Days
68	65 - 68	449-476
72	69 – 72	477-504
76	73 – 76	505-532
80	77 – 80	533-560

Besides the windows specified above, there is one additional window as defined as follows:

> "16 Weeks after Last Dose". This window will include the average of all data collected from 13 to 16 calendar weeks (85 to 112 calendar days) past the date of the last SC dose. All available on- or off-treatment data will be used for this window after the planned treatment period.

Note that, similar to data collected at clinic visit, any data collected via diary up to 84 calendar day (12 weeks, 8 weeks + a 4 week window) from the date of the last SC dose are 'on-treatment', and any data collected more than 84 calendar day (12 weeks, 8 weeks + a 4 week window) from the date of the last SC dose are off treatment.

Appendix 1.2. Definition of Protocol Deviations that Relate to Statistical Analyses/Populations

Not applicable.

Appendix 1.3. Definition of Analysis Populations/Sets

Not applicable.

Appendix 1.4. Further Definition of Endpoints

Health State Utility of the EQ-5D-5L

The EQ-5D-5L contains five questions that measure the following dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each of the five dimensions has five levels: (1) no problems; (2), slight problems; (3) moderate problems; (4) severe problems; and (5) extreme problems.

The health utility scores are defined for every possible set of outcome combinations of the five dimensions for the following countries:

- Denmark, France, Germany, Japan, the Netherlands, Spain, Thailand, UK, US and Zimbabwe

It is intended that this study will recruit patients from the following 8 countries.

- Canada, Denmark, France, Hungary, Japan, South Korea, Spain, Sweden, US.

Some of these may not actually recruit or treat patients, and other countries may be added. As there is a mismatch between countries where patients are being recruited and the currently available EQ-5D-5L health utility scoring, we will assign patients to the following scoring countries based on the following assignments.

EQ-5D-5L Scoring Country	Study Recruitment Country
Denmark	Denmark, Sweden
France	France
Germany	Hungary,
Japan	Japan, South Korea
The Netherlands	-
Spain	Spain
Thailand	-
UK	-
US	Canada, US
Zimbabwe	-

If more EQ-5D-5L utility scores become available or other countries are added, then this assignment may be modified.

The health utility for a patient with no problems in all 5 items is 1 for all countries (except for Zimbabwe where it is 0.9), and is reduced where a patient reports greater levels of problems across the five dimensions. The minimum score across the countries is -0.654.

WPAI:LBP Endpoints

The tables below summarizes the 6 questions of the WPAI:LBP questionnaire, and the four endpoints of the effect of impairment on activity and impairment.

Question	Question Wording	Scoring
1	Are you currently employed? [if No skip to question 6]	Yes, No
2	During the past seven days, how many hours did you miss from work due to problems associated with your Low Back Pain	number of hours (free text)
3	During the past seven days, how many hours did you miss from work because of any other reason, such as vacation, holidays, time off to participate in this study?	number of hours (free text)
4	During the past seven days, how many hours did you actually work (if '0' skip to Question 6)	number of hours (free text)
5	During the past seven days, how much did your Low Back Pain affect productivity while you were working?	0 to 10 scale with 0 being 'Low Back Pain had no effect on my work' and 10 being 'Low Back Pain completely prevented me from working'
6	During the past seven days, how much did your Low Back Pain affect your ability to do your regular daily activities, other than work at a job?	0 to 10 scale with 0 being 'Low Back Pain had no effect on my daily activities' and 10 being 'Low Back Pain completely prevented me from doing my daily activities'

WPAI endpoint	Calculation
Percent activity impairment due to Low Back Pain	Q6*10
Percent overall work impairment due to Low Back Pain	Q5*10
Percent impairment while working due to Low Back Pain	$\left\{\frac{Q2}{Q2+Q4}+\left[1-\left(\frac{Q2}{Q2+Q4}\right)\right]\left(\frac{Q5}{10}\right)\right\}\times 100$
Percent work time missed due to Low Back Pain	$\frac{Q2}{Q2+Q4} * 100$

TSQM vII

The 11 questions of the TSQM and the scoring are shown below:

Item	Question wording	Likert Scoring
1	How satisfied or dissatisfied are you with the ability of the medication to prevent or treat the condition?	1 (Extremely dissatisfied) to 7 (Extremely satisfied)
2	How satisfied or dissatisfied are you with the way the medication relieves symptoms?	1 (Extremely dissatisfied) to 7 (Extremely satisfied)
3	As a result of taking this medication, do you experience any side effects at all?	0 (No), 1 (Yes)
4	How dissatisfied are you by side effects that interfere with your physical health and ability to function (eg, strength, energy levels)?	1 (Extremely dissatisfied) to 5 (Not at all dissatisfied)
5	How dissatisfied are you by side effects that interfere with your mental function (eg, ability to think clearly, stay awake)?	1 (Extremely dissatisfied) to 5 (Not at all dissatisfied)
6	How dissatisfied are you by side effects that interfere with your mood or emotions and ability to function (eg, anxiety/fear, sadness, irritation/anger)?	1 (Extremely dissatisfied) to 5 (Not at all dissatisfied)
7	How satisfied or dissatisfied are you with how easy the medication is to use?	1 (Extremely dissatisfied) to 7 (Extremely satisfied)
8	How satisfied or dissatisfied are you with how easy it is to plan when you will use the medication each time?	1 (Extremely dissatisfied) to 7 (Extremely satisfied)
9	How satisfied or dissatisfied are you by how often you are expected to use/take the medication?	1 (Extremely dissatisfied) to 7 (Extremely satisfied)
10	How satisfied are you that the good things about this medication outweigh the bad things?	1 (Extremely dissatisfied) to 7 (Extremely satisfied)
11	Taking all things into account, how satisfied or dissatisfied are you with this medication?	1 (Extremely dissatisfied) to 7 (Extremely satisfied)

The scoring of the 4 satisfaction parameters are shown in the table below.

TSQM Parameter	Scoring
Effectiveness	[(Item 1+Item 2) – 2]/12 *100
Side Effects	[(Item 4 + Item 5 + Item 6) – 3]/12 *100 If one item is missing then: [(Sum of two completed items)-2]/8 *100
Convenience	[(Item 7 + Item 8 + Item 9) – 3]/18 *100 If one item is missing then: [(Sum of two completed items)-2]/12 *100
Global Satisfaction	[(Item 10+Item 11) – 2]/12 *100

The four parameters have a scale of 0-100, with 100 being the best (most satisfied) score.

Healthcare Resource Utilization (example using 3 month recall – 8 week recall is also used in the study)

Question	Response	Scoring
During the last 3 months, what services did you receive directly related to your osteoarthritis? • Primary Care Physician. • Neurologist. • Rhematologist. • Physician Assistant or Nurse Practioner. • Pain Specialist. • Orthopedist. • Physical Therapist. • Chiropractor. • Alternative Medicine or Therapy. • Podiatrist. • Nutritionist/Dietician. • Radiologist. • Home healthcare services. • Other.	Number of Visits.	Response not selected = 0. Number of visits = 1-999.
During the past 3 months, have you visited the emergency room due to your osteoarthritis?	Yes, No	No = 0 Yes = 1
How many times?	Number of visits	0-999
During the past 3 months, have you been hospitalized due to your osteoarthritis?	Yes, No	No = 0 Yes = 1
How many nights in total did you stay in hospital due to your	Number of Nights	0-999 (max should be 92)

osteoarthritis in the last 3 months?		
Did you use these aids or devices to help you in doing things because of your osteoarthritis in the last 3 months? • Walking Aid. • Wheelchair. • Devices or utensils to help you dress, eat or bathe. • Other.	Did not use any aids or devices Never, rarely, sometimes, often, always	Did not use any aids or devices = 0 Device not selected = 0 Never = 1 Rarely = 2 Sometimes = 3 Often = 4 Always = 5
Did you quit your job because of your osteoarthritis?	Yes, No	No = 0 Yes = 1 Not applicable = 2
How long ago did you quit your job because of your osteoarthritis?	Years and Months	0-99 Years and 0-99 Months (should be max of 11 months)

Rescue Medication Endpoints

Rescue medication data is collected daily using an electronic system up to Week 16, and weekly after Week 16 and up to Week 64. Daily and Weekly collected data will be assigned to a specific study week for summary and reporting. The assignment of daily and weekly data to weeks will use the same principle as described above in Appendix 1.1 for the daily and weekly aLBPI data.

The incidence of rescue medication use will look for any incidence in the week of interest (collected through daily or weekly diary data). The number of days of RM use (using daily and weekly data) and the total amount taken (using daily data up to Week 16 only) over the week will be calculated for the assigned week algorithm described above.

Imputation is described in Section 7 above. Imputation occurs for daily data up to Week 16 where the patient is in the trial and up to the end of that particular week. An example of imputation and calculating the three endpoints using the daily diary data is shown below.

Example of calculating rescue medication data from Daily Diary Data (Patient does not discontinue)

In this example, a patient has a Week 2 visit on study day 14 (slightly earlier than the nominal day 15). Study days 8-14 would represent Week 2 data.

Using the Week 2 interval described above for a subject, ie, study days [8-14], we have the following rescue medication example data. The amount taken and number of days of rescue medication use is adjusted for the duration of the Weekly interval.

Study Day (Week)	Number of Doses of RM taken [1]	Number of Doses of RM taken [1] with LOCF imputation
8 (Week 2)	2	2
9 (Week 2)	Missing	2 [2]
10 (Week 2)	0	0
11 (Week 2)	1	1
12 (Week 2)	Missing	1 [2]
13(Week 2)	2	2
14 (Week 2)	0	0

1. 500mg tablets of acetaminophen; [2] Using LOCF imputation for missing data.

For this subject the following data will be calculated for Week 2:

- Incidence of rescue medication taken in Week 2: Yes. Rescue medication taken on days 8, 9 (imputed), 11,12 (imputed), 13.
- Number of days of rescue medication use in Week 2: 5. For days 8-14 we have rescue medication taken on days 8,9 (imputed), 11, 12 (imputed), and 13. The number of days taken for the 7 day period is 5/7*7=5.
- Amount (mg) of rescue medication use in Week 2: For days 8-14 we have the number of doses taken of 2, 2 (imputed), 0, 1, 1 (imputed), 2, and 0. The number of doses taken for the 7 day period is 8/7*7=8, making the amount of acetaminophen dosage of 4000 mg.

Example of calculating rescue medication data from Daily Diary Data (Patient discontinues)

In this example, a patient discontinues on study day 62, a few days after a Week 8 visit (which was on study day 60). The Week 5-8 data is calculated as described above (eg, Week 8 using days [50, 56]). The patient has rescue medication data as shown below.

Study Day (Week)	Number of Doses of RM taken [1]	Number of Doses of RM taken [1] with LOCF imputation
57 (Week 9)	1	1
58 (Week 9)	1	1
59 (Week 9)	Missing	1 [2]
60 (Week 9)	Missing	1 [2]
61 (Week 9)	Missing	1 [2]
62 (Week 9)	Missing	1 [2]
63 (Week 9)	Missing	1 [2]

[1] 500mg tablets of acetaminophen; [2] Using LOCF imputation for missing data.

Week 9 is calculated as days 57 to 63. The data up to the end of the last week the patient was in the trial is imputed using LOCF as shown above. Therefore the Week 9 scores are then used to impute the Weekly data for summary and analysis for Weeks 10 to 56.

As above the incidence of rescue medication for Week 9 would be 'Yes'. The number of days of rescue medication use would be 7, and the average dose would be 7/7*7*500=3500 mg for this week.

Appendix 2. STATISTICAL METHODOLOGY DETAILS

Appendix 2.1. Further Details of Interim Analyses

Details of the ongoing review of safety data (including joint safety events) are given in a separate statistical analysis plan for the Data Monitoring Committee.

Appendix 2.2. Further Details of the Statistical Methods

A description of the combination of the ANCOVA results from each of the multiple imputed datasets is given below, and taken from Little & Rubin (2002), page 86-7.

In this analysis we have defined the number of imputations (D) to be 100.

The treatment estimates for individual treatment groups and treatment contrasts are defined as θ_i for $i = 1 \ldots D$. . The combined estimate is $\bar{\theta}_D = \frac{1}{D}\sum_{i=1}^{D}\theta_i$. The variability of the combined estimate contains components of both Within- (W) and Between- (B) imputation dataset variability. These are shown below:

$$\bar{W}_D = \frac{1}{D}\sum_{i=1}^{D} W_i \text{ and } B_D = \frac{1}{D-1}\sum_{i=1}^{D}\left(\hat{\theta}_i - \bar{\theta}_D\right)^2$$

where W_i is the variance for the parameter θ_i.

The total variance for $\bar{\theta}_D$ is shown below:

$$T_D = \bar{W}_D + \frac{D+1}{D} B_D .$$

The test statistic $\frac{\left(\theta - \bar{\theta}_D\right)}{\sqrt{T_D}}$ has a t-distribution with v^* degrees of freedom, which is defined below:

$$v^* = \left(\frac{1}{v} + \frac{1}{\hat{v}_{obs}}\right)^{-1}$$

,

using

$$v = (D-1)\left(1 + \frac{1}{D+1}\frac{\bar{W}_D}{B_D}\right)^2$$

$$\hat{v}_{obs} = \left(1 - \hat{\gamma}_D\right)\left(\frac{v_{com}+1}{v_{com+3}}\right)v_{com}$$

$$\hat{\gamma}_D = \left(1 + \frac{1}{D}\right)\frac{B_D}{T_D}.$$

$$v = (D-1)\left(1+\frac{1}{1+D^{-1}}\frac{\bar{W}_D}{B_D}\right)^2$$

$$\hat{v}_{obs} = (1-\hat{\gamma}_D)\left(\frac{v_{com}+1}{v_{com}+3}\right)v_{com}\hat{\gamma}_D = \left(1+\frac{1}{D}\right)\frac{B_D}{T_D}$$

This distribution can be used to construct the test statistics and 95% confidence intervals for θ.

Appendix 2.3. Schedule of Activities: Baseline through Week 56

	Screen		Treatment Phase										
Visit Indentifiers		**IPAP**	**Baseline**[b]	**Weeks 1, 3**	**Week 2**	**Week 4**	**Week 8**	**Week 12**[d]	**Week 16 Primary Efficacy timepoint**	**Week 24**	**Weeks 32, 40, 48**	**Weeks 20, 28, 36, 44, 52**[d]	**Week 56/End of Treatment**
Study Activities[a]	**Day -37 to -6**	**Day-5 to -1**	**Day 1 Dosing Visit**	**(±2 days) Telephone Contact**[c]	**Day 15 (±2 days)**	**Day 29 (±3 days)**	**Day 57 (±7 days) Dosing Visit**	**(±7 days) Telephone Contact**	**Day 113 (±7 days) Dosing Visit**	**(±7 days) Dosing Visit**	**(±7 days) Dosing Visits**	**(±7 days) Telephone Contact**	**Day 393 (±7 days)**
Informed Consent	X												
Inclusion/Exclusion Criteria/subject eligibility	X		X										
Demographics, General and Musculoskeletal Specific Medical History and Prior/Current Medication Use	X												
Primary Diagnosis/	X												
Quebec Task Force Category	X												
Height/BMI/ Smoking Status/Female Hormonal Status/Alcohol Use	X												
Body Weight	X												X
Health Care Resource Utilization (HCRU)			X										
Vital Signs (sitting BP, HR)	X		X		X	X	X		X	X	X		X
Orthostatic Blood Pressure (supine/standing)	X		X		X	X	X		X	X	X		X
Electrocardiogram (12-lead)	X								X				X
General Physical Examination	X												X
Musculoskeletal Physical Examination	X		X		X	X	X		X	X	X		X
Neurologic Exam/Neuropathy Impairment Score (NIS)[e]	X		X		X	X	X		X	X	X		X
Adverse event assessment			X	X	X	X	X	X	X	X	X	X	X
Review weekly joint pain scores			X	X	X	X	X	X	X	X	X	X	X
Concomitant medication review			X	X	X	X	X	X	X	X	X	X	X
Patient Reported Assessments Completed at Study Visits (collected via tablet device at site)													
BPI-sf			X		X	X	X		X	X	X		X
RMDQ			X		X	X	X		X	X	X		X

	Screen		Treatment Phase										
Visit Indentifiers		**IPAP**	**Baseline[b]**	**Weeks 1, 3**	**Week 2**	**Week 4**	**Week 8**	**Week 12[d]**	**Week 16 Primary Efficacy timepoint**	**Week 24**	**Weeks 32, 40, 48**	**Weeks 20, 28, 36, 44, 52[d]**	**Week 56/End of Treatment**
Study Activities[a]	**Day -37 to -6**	**Day-5 to -1**	**Day 1 Dosing Visit**	**(±2 days) Telephone Contact[c]**	**Day 15 (±2 days)**	**Day 29 (±3 days)**	**Day 57 (±7 days) Dosing Visit**	**(±7 days) Telephone Contact**	**Day 113 (±7 days) Dosing Visit**	**(±7 days) Dosing Visit**	**(±7 days) Dosing Visits**	**(±7 days) Telephone Contact**	**Day 393 (±7 days)**
Patient Global Assessment of Low Back Pain			X		X	X	X		X	X	X		X
Pain DETECT			X										
WPAI:LBP			X						X				X
EQ-5D-5L			X				X		X	X	X[f]		X
NIH CLBP Min Dataset			X						X				X
TSQM									X				X
mPRTI									X				X
Survey of Autonomic Symptoms (SAS)	X									X			X
Subject Daily and Weekly Assessments (via IRT)													
LBPI score and rescue medication usage[g]	X[h]	---------- (Daily via IRT [handheld device])----------								---------- (Weekly via IRT [handheld device])------			
Record joint pain in major joints, if applicable[i]	X	---------- (Weekly via IRT [handheld device])----------											
Concomitant NSAID usage			---------- (Weekly via IRT [handheld device])----------										
Radiographic Assessments													
X-rays of the hips, knees and shoulders	X									X[j]			X
Central reader to confirm radiologic eligibility	X									X[j]			
Compliance assessments													
Assess compliance with oral study medication				X	X	X	X	X	X	X	X	X	X
Compliance with daily and weekly diary entries via IRT			X	X	X	X	X	X	X	X	X	X	X
Rescue medication compliance			X	X	X	X	X	X	X	X	X	X	X
NSAID limit compliance				X	X	X	X	X	X	X	X	X	X
Remind subject of contraceptive requirements	X		X	X	X	X	X	X	X	X	X	X	X
Laboratory													
Hepatitis Screen (Hepatitis B & C); HIV, Urine Toxicology screen	X												
Hemoglobin A1c	X												

	Screen		Treatment Phase										
Visit Indentifiers		**IPAP**	**Baseline[b]**	**Weeks 1, 3**	**Week 2**	**Week 4**	**Week 8**	**Week 12[d]**	**Week 16 Primary Efficacy timepoint**	**Week 24**	**Weeks 32, 40, 48**	**Weeks 20, 28, 36, 44, 52[d]**	**Week 56/End of Treatment**
Study Activities[a]	**Day -37 to -6**	**Day-5 to -1**	**Day 1 Dosing Visit**	**(±2 days) Telephone Contact[c]**	**Day 15 (±2 days)**	**Day 29 (±3 days)**	**Day 57 (±7 days) Dosing Visit**	**(±7 days) Telephone Contact**	**Day 113 (±7 days) Dosing Visit**	**(±7 days) Dosing Visit**	**(±7 days) Dosing Visits**	**(±7 days) Telephone Contact**	**Day 393 (±7 days)**
Serum FSH testing[k]	X												
Serum/Urine Pregnancy Test[l]	X		X				X		X	X	X		X
Hematology	X		X						X				
Blood Chemistry	X		X						X				
Urinalysis	X												
Serum/Plasma Retention Sample			X						X				X
Plasma Pharmacokinetic sample[m]			X		X[m]	X[m]	X		X		X[n]		X
Serum Pharmacodynamic sample (NGF)[m]			X		X	X	X				X[n]		X
Serum Anti-Drug Antibody[m]			X				X		X		X[n]		X
Serum and urine biomarkers[o]			X										
Banked biospecimen (whole blood)			X										
Trial Treatment													
Assess treatment response and eligibility to continue in the trial[p]									X[p]		X[p]		
Randomization			X										
SC Study medication			X				X		X	X	X		
Blinded Oral Study Medication			----------X----------										
Adjust oral study medication based on pain relief and tolerability[q]				X	X	X			(X[r])	(X[r]))	(X[r]))		
Dispense Blinded Oral Study Medication			X		X	X	X		X	X	X		
Dispense rescue medication	X		X		X	X	X		X	X	X		X

Schedule of Activities: Follow-Up Period

Study Activities	Follow-Up Period				Early Termination (ET) Procedures				
	Week 60	Week 64	Week 68, 72, 76	Week 80 End of Study	ET Visit 1	ET Telephone Contact	ET Visit 2	ET Telephone Contact	Final ET Visit
	(±7 days) Tele phone Contact	Day 449 (±7 days)	(±7 days) Tele phone Contact	Day 560 (±7 days)	8 weeks after last dose of SC Study Med (±7 days)	12 weeks after last dose of SC Study Med (±7 days)	16 Weeks after last dose of SC Study Med (±7 days)	20 weeks after last dose of SC Study Med (±7 days)	24 Weeks after last dose of SC Study Med (±7 days)
Vital Signs (sitting BP, HR)		X		X	X		X		X
Orthostatic Blood Pressure (supine/standing)		X		X	X		X		X
12-lead ECG				X	X				X
Body weight					X				
General Physical Examination					X				
Musculoskeletal Physical Examination		X		X	X		X		X
Neurologic Exam/Neuropathy Impairment Score (NIS)[e]		X		X	X		X		X
Adverse event assessment	X	X	X	X	X	X	X	X	X
Review weekly joint pain scores	X	X	X	X	X	X	X	X	X
Concomitant medication review	X	X	X	X	X	X	X	X	X
Patient Reported Assessments Completed at Study Visits (collected via tablet device at site)									
BPI-sf		X			X		X		
RMDQ		X		X	X		X		X
Patient Global Assessment of Low Back Pain		X			X		X		
WPAI:LBP		X			X		X		
EQ-5D-5L		X			X		X		
NIH CLBP Min Dataset					X				
HCRU		X		X			X		X
TSQM					X				
mPRTI					X				
Survey of Autonomic Symptoms questionnaire				X	X				X
Subject Weekly Assessments (IRT)									
LBPI score[g]	----------X------				----------X----------				
Rescue medication use[g]	----------X----------				----------X----------				
Record joint pain, if applicable[i]	----------X----------				----------X----------				
Concomitant NSAID usage	----------X----------				----------X----------				
Radiographic Assessments									
X rays of the hips, knees and shoulders				X	X				X
Compliance assessments									
Compliance with daily and weekly diary entries via IRT	X	X	X		X	X	X	X	
Rescue medication compliance	X	X			X	X	X		
NSAID limit compliance	X	X			X	X	X		
Remind subject of contraceptive requirement	X	X			X	X	X		
Laboratory									
Serum/Urine Pregnancy Test[l]		X			X		X		
Hematology		X					X		

Study Activities	Follow-Up Period				Early Termination (ET) Procedures				
	Week 60	Week 64	Week 68, 72, 76	Week 80 End of Study	ET Visit 1	ET Telephone Contact	ET Visit 2	ET Telephone Contact	Final ET Visit
	(±7 days) Tele phone Contact	Day 449 (±7 days)	(±7 days) Tele phone Contact	Day 560 (±7 days)	8 weeks after last dose of SC Study Med (±7 days)	12 weeks after last dose of SC Study Med (±7 days)	16 Weeks after last dose of SC Study Med (±7 days)	20 weeks after last dose of SC Study Med (±7 days)	24 Weeks after last dose of SC Study Med (±7 days)
Blood Chemistry		X					X		
Serum/Plasma Retention Sample		X			X		X		
Plasma Pharmacokinetic sample		X			X		X		
Serum Pharmacodynamic sample (NGF)		X			X		X		
Serum Anti-Drug Antibody		X		X	X		X		X
Trial Treatment									
Dispense rescue medication		X			X		X		
Assign standard of care treatment as needed		X					X		

a. Refer to Protocol Section 6 for the order in which study procedures are to be conducted.
b. All study activities at Baseline (Day 1) should be performed prior to dosing with study medication, unless otherwise noted.
c. Telephone contact with the subject should be made at Weeks 1 and 3 to counsel the subject regarding adjustment of the oral study medication.
d. Telephone Contact Visit: If adverse events dictate that the subject should be seen, an unscheduled visit may be conducted and pertinent exams conducted (eg, physical exam, neurological exam, ECG, clinical laboratory testing) depending on the nature of the event and the Investigator's clinical judgment.
e. A neurological examination (NIS) will be performed by the Investigator (or designated physician) and assessed for clinically significant changes from Baseline.
f. EQ-5D-5L to be conducted at Week 40 only.
g. LBPI scores and rescue medication use are collected daily using IRT up to Week 16. After Week 16 LBPI scores and rescue medication will be collected weekly via IRT. LBPI scores are collected until Week 64, rescue medication usage is collected to Week 80.
h. At the Screening visit only, in order to determine eligibility, the LBPI score will be collected via the IRT.
i. Collected at Screening and then weekly thereafter as described Protocol Section 7.3.2.
j. At Week 24, sites must receive confirmation of continued radiologic eligibility from the Central Reader prior to administering the Week 24 SC study medication. Refer to Protocol Section 6.1.1.1.
k. FSH testing in female subjects as described in Protocol Section 7.3.4.4.
l. Serum pregnancy tests are obtained at Screening, Weeks 56 and 64 or Early Termination Visits 1 and 2 for subjects who discontinue (Refer to Protocol 6.20). A urine pregnancy test will be obtained at Baseline prior to initial dosing, and pre-dose at Weeks 8, 16, 24, 32, 40, and 48.
m. On dosing visits, samples for ADA, PK, and NGF should be obtained pre-dose. At Weeks 2 and 4 PK and NGF will be at measured at selected sites only.
n. Of the 3 dosing visits listed, obtain a sample for ADA, PK, at Weeks 32 and 48 and for PD (NGF) at Week 48 only.
o. Biomarker samples should be collected prior to dosing and, if possible, following a fasting period of at least 8 hours at approximately the same time of day at all scheduled timepoints. Urine collected for biomarkers should be the second or later void of the day.
p. At the Week 16 visit, all subjects must have at least a 30% reduction in average LBPI score relative to Baseline and at least a 15% reduction in mean weekly LBPI score from Baseline at any week from Week 1 to Week 15 in order to continue study treatment. In addition, at the Week 32 visit, all subjects must have at least a 30% reduction in average LBPI score relative to Baseline in order to continue study treatment. Subjects who do not meet these response criteria will be discontinued from the Treatment Period and enter Early Termination Follow-up Period (See Protocol Section 6.20.1).
q. Refer to Protocol Section 6.5.
r. For subjects participating in Europe, following the completion of the Week 16 visit through the Week 56 visit, the dose of tramadol PR or oral placebo may be decreased to a minimum of 100 mg per day, if clinically indicated. If the dose of tramadol PR or oral placebo is reduced, it may later be re-escalated to a maximum of the previous individually titrated dose (See Protocol Section 5.5.2).