Protocol A4091063

A PHASE 3 RANDOMIZED, DOUBLE BLIND, ACTIVE CONTROLLED, MULTICENTER STUDY OF THE LONG TERM SAFETY AND EFFICACY OF SUBCUTANEOUS ADMINISTRATION OF TANEZUMAB IN JAPANESE ADULT SUBJECTS WITH CHRONIC LOW BACK PAIN

Statistical Analysis Plan
(SAP)

Version: 3

Date: 4-Jul-2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
Original 25 Dec 2015	Original 28 Aug 2015	N/A	N/A
Version 2 14 Jun 2019	Amendment 2 14 Sep 2017	The changes reflect protocol amendment, updates from blinded data review and program decisions for alignment of analysis. Additionally, clarifications, removal of redundant/duplicated text have been implemented.	• Various formatting and editorial clarifications were made (throughout). • Section 2.2: Updated the study design according to the change made in Protocol amendment 2. • Section 2.2: Clarified that most safety results will be presented for the treatment period. • Section 3.1 and 6.1: Specified the outcome of consultation as safety endpoints. • Section 3.2: Added change from baseline to Week 1 in low back pain intensity (LBPI). Added cumulative distribution of percent change from Baseline in RMDQ score to Weeks 16 and 56. Added usage of rescue medication (incidence, number of days of usage and amount taken) during Week 1. • Section 3.4: Modified description. Moved the description on how to summary baseline variables to Section 6.5.1. • Section 3.5.1: Clarified the Tier-1 events. Section 3.5.1: Changed the Tier-2 AE cutoff from 1% to 3%. Updated definitions for Abnormal Peripheral Sensation and Sympathetic Nervous System. A smaller set of the AEs of Sympathetic Nervous System (AEs of Decreased Sympathetic Function) is deleted. • Section 4: Clafied the definitions of analysis sets. • Section 5.2.1: Removed the calculation of treatment difference. Changed the calculation of 95% CI for response rate. • Section 5.2.2: Added mixed model ANCOVA for aLBPI and RMDQ. Removed the calculation of treatment differences, except for aLBPI and RMDQ ANCOVA analyses. • Section 5.2.3: Moved the description on NIS and PGA-LBP to section 6 with some modification. Removed the summary of WPAI parameters using categories.

			• Section 5.3: Added the analysis with multiple imputation, BOCF and LOCF for the summary for LBPI, RMDQ, PGA and BPI-sf. • Section 6.1.1: P-value and 95% CI for risk difference will be shown for pooled Tier 1 adverse events and will not be shown for each adverse event within Tier 1 event. Updated the description of footnote for 3-tier AE tables. Removed the following analyses: Tier 1 and Tier 2 AE graphs, the comparison between the number of subjects with AEs of APS against the number of subjects with the AE of Oedema peripheral, the plots of the start and stop study day of specified individual AES of APS, the summary of start day and AE duration for five AEs of APS (Paraesthesia, Hyperaesthesia, Hypoaesthesia, Allodynia and Dysaesthesia), AEs of DSF and Oedema peripheral, summary of AEs and serious AEs by 1000 patient years of exposure, 'incidence and severity' table for decreased sympathetic function (DSF), and listing of AEs for subjects who are not randomized and those randomized and not treated. Added the summary for adverse events associated with laboratory abnormalities, vital signs, and ECG measurement. • Section 6.1.2: Changed the category of blood pressure. • Section 6.1.3: Removed the summary of mean change in postural blood pressure. Clarified the detail of analysis. • Section 6.1.6: Updated the analysis plan for joint safety adjudication outcomes and total joint replacement. • Section 6.1.7: Separate section (6.1.8) was made for outcome of consultation and moved the description on outcome of consultation to Section 6.1.8. • Section 6.1.9: Specified that ADA assessments will be made using all available data. • Section 6.1.10: Updated the analysis plan for physical examination. • Section 6.2: Clarified the definition of 'on-treatment' and 'off-treatment'.

			• Section 6.2.1: Added the ANCOVA analysis for change from baseline for aLBPI. Updated the description on handling of missing data based on the modification in section 5.3. Added plots of cumulative distribution of percent change from Baseline in aLBPI score. Removed the calculation of treatment differences in mean of change from baseline for aLBPI and in response rate in each aLBPI response category. Added the summary of change from baseline of aLBPI at Week 1. Removed the summary of proportion of subjects who meet a aLBPI response definition at Week 16. Added the description on the handling of aLBPI data obtaining after the early termination visit. • Section 6.2.2: Added the ANCOVA analysis for change from baseline for RMDQ. Added the description on handling of missing data based on the modification in section 5.3. Removed the calculation of treatment differences in mean of change from baseline for RMDQ. Added treatment response analysis for RMDQ. • Section 6.2.3: Updated the description on handling of missing data based on the modification in section 5.3. Removed the calculation of treatment differences in mean of change from baseline for PGA of LBP and in response rate in PGA of LBP response category. • Section 6.2.4: Updated the description on handling of missing data based on the modification in section 5.3. Removed the calculation of treatment differences in mean of change from baseline for BPI-sf. • Section 6.2.5: Removed the calculation of treatment differences in response rate in CLBP-RI. Added a plot of CLBP-RI analysis. • Section 6.2.7: Modified the description on analyses. • Section 6.2.8: Clarified the definition of event "discontinuation from treatment due to lack of efficacy". • Section 6.2.9: Added usage of rescue medication (incidence, number of days of usage and amount taken) during Week 1. Clafiried the summary of rescue medication.

			• Section 6.2.10: Clarified the analyses for health care resource utilization. • Section 6.2.12: Removed the summary splitted by response to question 1. • Section 6.3.1: Clarified PK data summaries. • Section 6.3.2: Clarified NGF data summaries. • Section 6.3.3: Removed as biomarker data are not generated. • Section 6.4: Removed Subset analysis using NGF subset analysis dataset and biomarker subset analysis dataset. • Section 6.5.1: Added baseline summaries for diabetes status and painDetect score. • Section 6.5.2: Removed cumulative summary of discontinuation by treatment group and reason, study week of discontinuation, and summary table and listing of demographic details for subjects who are not randomized and those randomized but not treated. • Section 6.5.3: Modified the summary of study treatment exposure. • Section 6.5.4: Added NSAID use summaries. • Section 8: Updated the references. • Appendix 1: Updated the summary of efficacy analyses. • Appendix 2.1: Clarified the windows for various data types. • Appendix 2.3: Some modifications and clarifications were made. • Appendix 4.2: Added the details of the statistical method. • Appendix 4.3: Added the detail of painDETECT questionnaire. • Delete Appendix 5 (immunogenicity data reporting detail).

Version 3 4 Jul 2019	N/A	The changes reflect updates from blinded data review. Additionally, clarifications, removal of redundant/duplicated text have been implemented.	• Section 5.2.2: Clafiried the explanation on the ANCOVA analyses. • Section 6.1.4: Modified the explanation. • Section 6.1.7: Added the analysis on Neuropathy impairment score (NIS). • Section 6.2.1: Added the treatment difference of proportion of responder in weekly average LBPI score. • Section 6.2.6: Modified the descriptive statistics shown in the summary table. • Section 6.5.2: Modified the study week categories in the summary of time to discontinuation. • Section 6.5.3: Added the description on the summary of number of SC doses taken and duration in phases. Updated the description on the summary of oral study medication compliance. • Section 6.5.4: Added the description on the summary for treatments prior to start of study treatment. Clarified the summary for concomitant medications and non-drug treatments. • Appendix 1: Modified the time point for efficacy analyses.

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study A4091063. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives

Primary Objective:

- *Evaluate the long-term safety of tanezumab 10 mg and 5 mg SC administered every 8 weeks (7 administrations).*

Secondary Objective:

- *Demonstrate the long-term analgesic efficacy of tanezumab 10 mg and 5 mg SC administered every 8 weeks (7 administrations).*

2.2. Study Design

This is a randomized, double-blind, active-controlled, multicenter, parallel-group Phase 3 study of the safety and efficacy of tanezumab when administered by SC injection for up to 56 weeks in subjects with chronic low back pain. Subjects will be randomized to 1 of 3 treatment groups in a 1:1:1 ratio. Treatment groups will include:

1. *Placebo SC matching tanezumab administered at an 8-week interval (total of 7 times) plus celecoxib 100 mg twice a day (BID) to be administered orally for 56 weeks.*

2. *Tanezumab 5 mg SC administered at an 8-week interval (total of 7 times) plus placebo matching celecoxib to be administered orally BID for 56 weeks.*

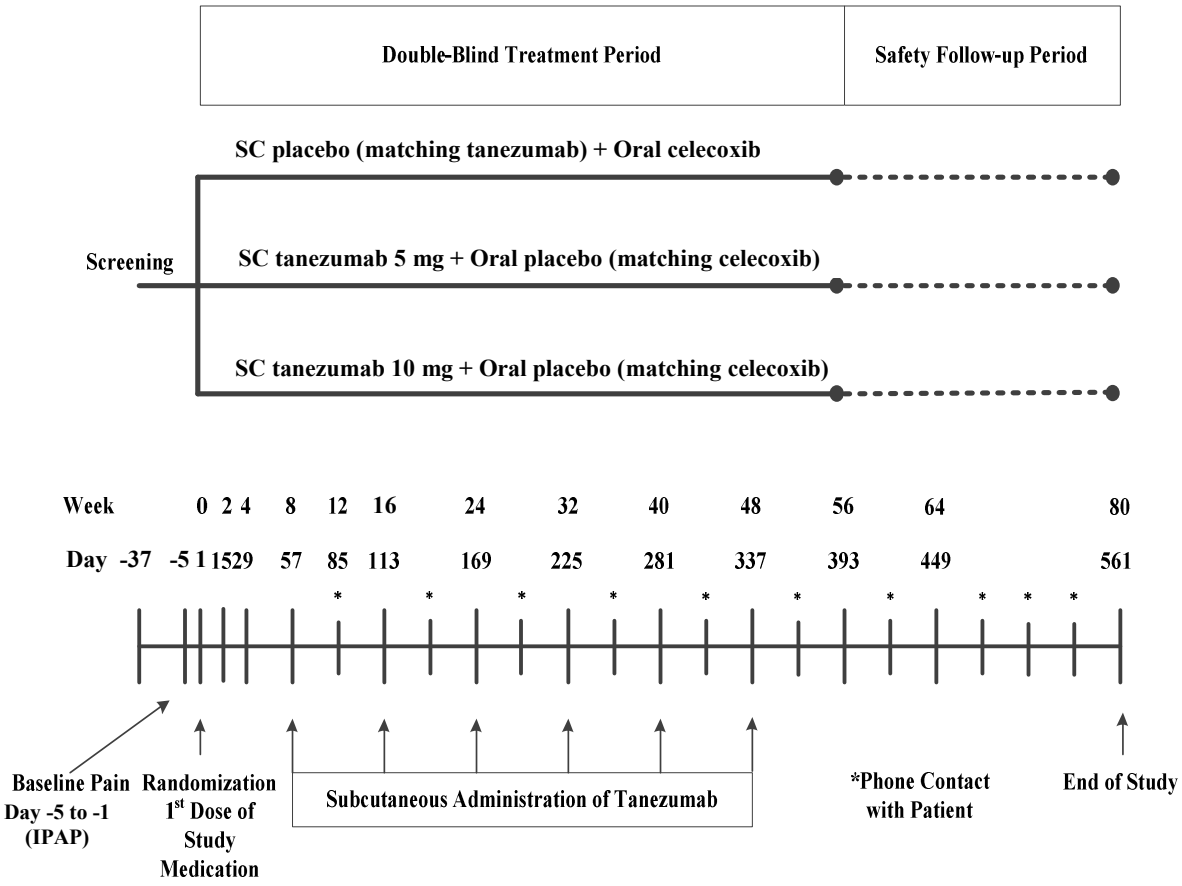
3. *Tanezumab 10 mg SC administered at an 8-week interval (total of 7 times) plus placebo matching celecoxib to be administered orally BID for 56 weeks.*

The study is designed with a total duration (post randomization) of up to 80 weeks and will consist of three periods: Screening (up to 37 days; includes a Washout Period and an Initial Pain Assessment Period), Double-blind Treatment (56 weeks) and Follow-up (24 weeks).

The total sample size will be approximately 200 subjects (170-220 subjects, approximately 66 subjects [56-73 subjects] per treatment group). However, it is acceptable to randomize more than 220 subjects from the safety perspective as described below because primary objective is to evaluate the long-term safety.

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.

Figure 1. Study Schematic



3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoints

Safety Measures

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

The details of the safety endpoints are described in Section 3.5.

3.2. Secondary Endpoints

Efficacy-Related Endpoints

- *Change from Baseline to Weeks* 1, *2, 4, 8, 12, 16, 24, 32, 40, 48, 56 and 64 in average Low Back Pain Intensity (LBPI) score as measured by an 11-point Numeric Rating Scale (NRS).*
- *Change from Baseline to Weeks 2, 4, 8, 16, 24, 32, 40, 48, 56 and 64 in the Roland-Morris Disability Questionnaire (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.*

- Cumulative distribution of percent change from Baseline in RMDQ score to Weeks 16 and 56.

- *Response as defined by a ≥30%, ≥50%, ≥70% and a ≥90% reduction from Baseline in weekly average LBPI score derived from the subject diary at Weeks 16, 24, 40 and 56.*

- *Change from Baseline to Weeks 2, 4, 8, 16, 24, 40, 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.*

- *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.*

- *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 16, 24, 40 and 56.*

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

  - A reduction of ≥30% in weekly 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.

- *Treatment Response: Improvement of ≥2 points in Patient's Global Assessment of Low Back Pain at Weeks 16, 24, 40 and 56.*

- *Euro Quality of Life Health State Profile (EQ-5D-5L™) dimensions and overall health utility score at Baseline, Weeks 16 and 56.*

- *Work Productivity and Activity Impairment Questionnaire: Low Back Pain (WPAI:LBP) change from Baseline to Week 16, 56 and 64, 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* 1, *2, 4, 8, 12, 16, 24, 32, 40, 48, 56 and 64.*

- *Usage of rescue medication (amount taken) during Weeks* 1, *2, 4, 8, 12 and 16.*

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

Treatment Satisfaction Measures

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

  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 4 parameters are described in Appendix 2.3.

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

3.3. Tertiary Endpoints

- *Plasma tanezumab concentrations.*

- *Serum NGF assessments.*

- *Serum and urine osteoarthritis biomarker concentrations*.

3.4. Baseline Variables

Baseline variables are demographics, general and musculoskeletal specific medical history, Musculoskeletal baseline status, prior/current medication use, primary diagnosis and Quebec Task Force (QTF) category.

The method and definition of reporting windows including Baseline for assigning efficacy data to particular time points is described in Appendix 2.1.

3.5. Safety Endpoints

3.5.1. Adverse Events

An adverse event is considered treatment emergent relative to a given treatment if:

- The event occurs for the first time during the effective duration of treatment and was not seen prior to the start of treatment, or

- The event was seen prior to the start of treatment but increased in severity during treatment.

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 (See Section 6.1.1).

Tier-1 events: These are pre-specified events of clinical importance. Tier-1 events include syncope, bradycardia, orthostatic hypotension, anhidrosis, hypohidrosis.

Tier-2 events: These are events that are not Tier-1 but are "common". A MedDRA PT is defined as a Tier-2 event with a frequency of ≥3% in any treatment group.

Tier-3 events: These are events that are neither Tier-1 nor Tier-2 events, 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).

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	

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.

3.5.2. Laboratory Tests

Larobatory tests will be summarized according to Pfizer standards.

3.5.3. Orthostatic (supine/standing) Blood Pressure Assessments

The incidence of orthostatic hypotension at each visit, at the last visit and for any post-baseline incidence will be summarized. 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.

3.5.4. Survey of Autonomic Symptoms (SAS) Scores

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 then 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.

3.5.5. Electrocardiogram (ECG, 12-lead) Data

Electrocardiogram data including PR, QRS, QT, QTcF, QTcB, RR intervals, and heart rate [HR] will be summarized according to Pfizer standards.

3.5.6. Joint Safety Adjudication Outcomes and Total Joint Replacements

The following adjudication outcomes and total joint replacement endpoints are defined.

- The composite endpoint consisting of adjudication outcomes of rapidly progressive osteoarthritis (type-1 or type-2), subchondral insufficiency fracture, primary osteonecrosis, or pathological fracture.

- The composite endpoint consisting of adjudication outcomes of rapidly progressive osteoarthritis (type-2 only), subchondral insufficiency fracture, primary osteonecrosis, or pathological fracture.

- Individual adjudication 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.

- Total joint replacements.

- Total joint replacements or any adjudication outcome (subjects who undergo total joint replacement plus subjects who have an adjudicated outcome of rapidly progressive osteoarthritis type-1 or type-2, subchondral insufficiency fracture, primary osteonecrosis, or pathological fracture whether they undergo total joint replacement or not).

3.5.7. Neurologic Examination (Neuropathy Impairment Score [NIS])

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.

3.5.8. Anti Drug Antibody Assessments (ADA)

The assessments of ADA data as described in Section 6.1.9 will be summarized.

3.5.9. Physical Examinations

The physical examination data as described in Section 6.1.10 will be summarized.

4. ANALYSIS SETS

Data for all subjects will be assessed to determine if subjects meet the criteria for inclusion in each analysis population prior to unblinding and releasing the database and classifications will be documented per standard operating procedures.

4.1. Intent To Treat Analysis Set

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

This analysis set is expected to be the same as the safety analysis set. If a subject is treated without having been randomized, then the analysis sets will be different, and safety analyses will use this different safety analysis set (see Section 4.2). If the sets are the same, safety analyses may still refer to it as the 'Safety Population'.

4.2. 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 population' in the corresponding safety data analyses, summaries, and listings. Subjects will be assigned per treatment they actually received.

4.3. Other Analysis Sets

TJR (Total Joint Replacement) 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.

4.4. 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 took 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 database release.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

No formal hypotheses are to be tested.

5.2. General Methods

5.2.1. Analyses for Binary Data

Output will show the number and percentage of subjects in each response category, and the response rate and its 95% confidence interval (CI) will be calculated. The confidence interval of the response rate will be calculated based on Agresti & Coull method.

5.2.2. Analyses for Continuous Data

The descriptive statistics n, mean, median, standard deviation, minimum, maximum and 95% CI will be used to summarize the endpoints.

The analysis for change from baseline of aLBPI (average LBPI) will use mixed model ANCOVA with a fixed effect of treatment group, Baseline aLBPI as a covariate, and study site as a random effect. The analysis for change from baseline of RMDQ will use mixed model ANCOVA with a fixed effect of treatment group, Baseline score and Baseline aLBPI as covariates, and study site as a random effect. Change from baseline at each time point will be estimated using least squares means (LS means) with standard error (SE) and 95% CI.

Estimates of treatment differences between each tanezumab treatment group versus celecoxib group will also be done using least squares means (LS means) with SE and 95% CI. P-value will not be shown. Under the analysis using multiple imputation defined in Section 5.3, the multiple ANCOVA results will be combined using standard methods (Little & Rubin, 2002),[4] which are described in Appendix 4.2.

5.2.3. Analyses for Categorical Data

Output will show number and percentage of subjects in each category.

5.2.4. Analyses for Time to Event Data

Survival curve estimates (time to 1st, 2nd, 5th, 10th and 25th percentiles, and minimum and maximum values) and a plot of the time to event 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.

Any subject who discontinues for any other reason 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.

5.3. Methods to Manage Missing Data

Missing values in standard summaries of AEs, lab values, vital signs and ECGs will be handled per Pfizer standard algorithms.

The summary of LBPI, RMDQ, PGA and BPI-sf will use multiple imputation for missing data (where the method for imputation will be dependent on the reason for missing data). As for PGA and BPI-sf, the imputation will be used only for Week 16 and Week 56 analyses. The imputation strategies are described in the following table. While the table describes the multiple imputation strategy specific for the Week 16 time point, it also applies to Week 56 time point, eg, 'Week 56' would be substituted for 'Week 16' in the table, to describe the Week 56 multiple imputation strategy.

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 visit reporting window*),	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 observed efficacy data at Week 16 over all ITT subjects. For example if last observation

• Subject discontinued for a different reason prior to or during the Week 16 visit reporting window*.	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 2.1 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 and 56 data for the PGA-LBP will be rounded to integer scores in the range 1 to 5. Imputed Week 16 and 56 data for the aLBPI that are <0 and >10 will be truncated to 0 and 10, respectively.

Two 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 Week 16 and 56 will impute the subject's Baseline value for the Week 16 and 56 time point, respectively, 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 Week 16 and 56 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).

As for LBPI and RMDQ, ANCOVA with covariates described in Section 5.2.2 will be conducted for all time points up to week 56. Study site will be fitted as a random effect in the ANCOVA model. This analysis will use multiple imputation for missing data as described above. While the table describes the multiple imputation strategy specific for the Week 16 time point, it also applies to all other analysis time points, eg, 'Week 2' would be substituted for 'Week 16' in the above table, to describe the Week 2 multiple imputation strategy.

The responder endpoints will be summarized, using both BOCF and LOCF 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, 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, patient's not meeting protocol-specified pain criteria, 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 (Chronic Low Back Pain Responder Index) 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 2.1 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 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 48 hours of the Baseline visit (so part of day -2, and 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. Imputation of weekly diary data after week 56 will not be performed.

The Baseline mean will be calculated using potential five values of the Initial Pain Assessment Period (IPAP). 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.

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.

For the other efficacy and safety analyses, the observed data (no imputation for missing data) will be summarized if not otherwise specified. Additional analyses will be performed in order to explore the sensitivity of the efficacy analysis results to missing data as described in this section.

6. ANALYSES AND SUMMARIES

6.1. Primary Endpoints

The primary objective of this study is to evaluate the long term safety of tanezumab 10 mg and 5 mg SC administered every 8 weeks (7 administrations). In this section, the safety measures are summarized.

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

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.

Adverse events (AEs), concomitant medications, laboratory safety tests, physical and neurological examinations (NIS), vital signs, electrocardiogram (ECG), and the Anti-Drug Antibody (ADA) test will be collected for each subject during the study according to the schedule of assessments.

6.1.1. Safety Measures: Adverse Events

Pfizer standard safety data presentations will be made for adverse event data.

The details of Tier-1, Tier-2 and Tier-3 AEs are described in Section 3.5.1.

Treatment-emergent adverse events within Tier-1 will be summarized. The combined Tier-1 adverse events, but not each adverse event will be summarized using Risk Differences between each tanezumab group and celecoxib, together with 95% CI and p-value, using exact methods. Treatment-emergent adverse events within Tier-2 will be summarized using Risk Differences between each tanezumab group and celecoxib, together with 95% CI, using exact methods. There will be no multiplicity adjustment for these significance tests. These tables will be produced for the comparisons of tanezumab 5mg versus celecoxib and tanezumab 10mg versus celecoxib.

The following footnote will be used in the Tier-1 AE tables: “P-values and confidence interval are not adjusted for multiplicity and should be used for screening purpose only. 95% CIs 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 sign data and ECG data.

The non-standard safety tables as described in Section 3.5.1 will also be summarized. The details are described below.

Summary of AEs, Incidence of AEs, Incidence of AEs leading to discontinuation and summary of Serious AEs will be shown for the whole study period (including the safety follow-up period) will be created.

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

- 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.
- 'Incidence and severity' tables of treatment-emergent adverse events of Abnormal Peripheral Sensation (APS), and Sympathetic Nervous system, 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.
- Adverse events associated with laboratory abnormalities, vital signs, and ECG measurement may be summarized.

6.1.2. Safety Measures: Standard Safety Assessments (safety laboratory testing [chemistry, and hematology], vital signs)

Pfizer standard safety data presentations will be made for laboratory test data and vital signs data. The following non-standard safety tables will also be included:

- 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.

6.1.3. Safety Measures: Orthostatic (supine/standing) Blood Pressure Assessments

The following non-standard safety tables will be included

- Incidence of subjects with orthostatic hypotension (defined in Section 3.5.3 above), for each visit and any post-baseline 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.

6.1.4. Safety Measures: Survey of Autonomic Symptoms (SAS) Scores

The following non-standard safety tables will be included:

- Summary of the Survey of Autonomic Symptoms (SAS) number of symptoms reported and total symptom impact score as continuous data, at each visit, and for the change from Screening score (scores defined in Section 3.5.4 above).

6.1.5. Safety Measures: Electrocardiogram (ECG, 12-lead) Assessments

Pfizer standard safety data presentations will be made for ECG data such as PR, QRS, QT, QTcF, QTcB, RR intervals, and heart rate [HR].

6.1.6. Safety Measures: Joint Safety Adjudication Outcomes and Total Joint Replacements

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, pathological fracture, normal progression of OA and other joint outcomes. The incidence of subjects with any of the adjudication outcomes of rapidly progressive osteoarthritis (type-1 or type-2), subchondral insufficiency fracture, primary osteonecrosis, or pathological fracture will also be summarized. Total joint replacement (TJR) will also be summarized. Output will show the number and percentage with 95% CI (exact confidence interval) of subjects for each outcome. In addition, subject years of exposure for individual treatment group and exposure adjusted event rates (number of subjects with an adjudicated joint safety event divided by exposure) of each outcome (with 95% CI) will also be shown. For each outcome, association with TJR, affected joint(s), and baseline Kellgren-Lawrence Grade of affected joint will be summarized. Number of TJRs during study, study period of first TJR occurrence, number of only elective TJR, number of joints with TJR prior to study, location of TJR prior to study 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.

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 1063 substudy will be part of the 1063 study CSR as there are <10 total joint replacement events expected in the substudy.

6.1.7. Safety Measures: Neurologic Examination (Neuropathy Impairment Score [NIS])

The change from Baseline in the NIS for Weeks 2, 4, 8, 16, 24, 32, 40, 48, 56, 64 and 80 will be summarized. The number and percentage of subjects whose NIS score worsened (change>0), improved (change<0) or had no change will also be shown. 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 and to the last post baseline value. The last post baseline value will only be used in the summary of change from baseline (not be used in the summary with change categories).

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.

6.1.8. Safety Measures: Outcome of Consultation

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

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

6.1.9. Safety Measures: Anti-drug Antibody Assessments (ADA)

The following assessments of ADA data will be made using all available data:

- 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 efficacy and safety profile.

6.1.10. Safety Measures: Physical Examinations

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

- Incidence of physical examination findings at screening.
- Incidence of musculoskeletal physical examination findings at screening.

6.2. Secondary Endpoints

All analyses in this section use the ITT analysis set, if not otherwise specified. If not otherwise specified, the observed data (no imputation for missing data) will be summarized.

All efficacy assessments during the treatment period are made on the analysis windows defined in Appendix 2.1. 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.

6.2.1. Average Low Back Pain Intensity (LBPI) Score

Change from Baseline to Weeks 1, 2, 4, 8, 12, 16, 24, 32, 40, 48, 56 and 64 in average LBPI score.

For each treatment and each timepoint assessed, the endpoint will be summarized along with 95% CIs. Handling of missing data is described in Section 5.3.

The ANCOVA model for change from baseline of aLBPI up to week 56 described in section 5.2.2 will also be conducted. The plot of LS mean (+/- SE) for change from baseline up to week 56 will also be prepared.

LBPI data obtained after the early termination visit and considered as off-treatment will be shown in the listing and not be included in any summaries and analyses.

Cumulative distribution of percent change from Baseline in average LBPI score to Weeks 16, 24 and 56.

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 observed data, mixed BOCF/LOCF, and also LOCF and BOCF imputation for aLBPI (see Section 5.3).

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.

Plots of cumulative distribution of percent change from Baseline in average LBPI score to Week 16 and Week 56 (mixed BOCF/LOCF) will be created.

Response as defined by a ≥30%, ≥50%, ≥70% and a ≥90% reduction from Baseline in weekly average LBPI score derived from the subject diary at Weeks 16, 24, 40 and 56.

For each treatment and each timepoint assessed, the number and percentage of subjects in each response category endpoint will be summarized along with 95% CIs. Treatment differences of proportion of responder between each tanezumab treatment group versus celecoxib group will be estimated with 95% CI of Agresti-Caffo method. These analyses use observed data, mixed BOCF/LOCF, and 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.

6.2.2. Roland-Morris Disability Questionnaire (RMDQ) total score

Change from Baseline to Weeks 2, 4, 8, 16, 24, 32, 40, 48, 56 and 64 in the RMDQ total score.

For each treatment and each timepoint assessed, the endpoint will be summarized along with 95% CIs. Handling of missing data is described in Section 5.3.

The ANCOVA model for change from baseline of RMDQ up to week 56 described in section 5.2.2 will also be conducted. The plot of LS mean (+/- SE) for change from baseline up to week 56 will also be prepared.

Response as defined by a ≥30%, ≥50%, ≥70% and a ≥90% reduction from Baseline in the RMDQ score at Weeks 16, 24, 40 and 56.

For each treatment and each timepoint assessed, the endpoint will be summarized along with 95% CIs. These analyses use observed data, mixed BOCF/LOCF, and 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.

Cumulative distribution of percent change from Baseline in RMDQ score to Weeks 16 and 56.

The cumulative RMDQ response at Weeks 16 and 56 using response definitions from a reduction of >0% to =100% (in steps of 10%) will be summarized using mixed BOCF/LOCF imputation for RMDQ. Plots of cumulative distribution of percent change from Baseline in RMDQ score to Week 16 and Week 56 (mixed BOCF/LOCF) will also be created.

6.2.3. Patient's Global Assessment of Low Back Pain (PGA-LBP)

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.

For each treatment and each timepoint assessed, the endpoint will be summarized along with 95% CIs. Handling of missing data is described in Section 5.3.

Treatment Response: Improvement of ≥2 points in Patient's Global Assessment of Low Back Pain at Weeks 16, 24, 40 and 56.

For each treatment and each timepoint assessed, the endpoint will be summarized along with 95% CIs. Handling of missing data is same with Section 5.3.

Treatment Response: Any improvement (change<0), no change (change=0), any worsening (change>0) at Weeks 16, 24, 40 and 56.

For each treatment and each timepoint assessed, the number and percentage of subjects in each category will be shown. Handling of missing data is same with Section 5.3.

6.2.4. Brief Pain Inventory-short form (BPI-sf) scores

Change from Baseline to Weeks 2, 4, 8, 16, 24, 40, 56 and 64 in the 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.

For each treatment and each timepoint assessed, the endpoint will be summarized along with 95% CIs. Handling of missing data is described in Section 5.3.

6.2.5. Chronic Low Back Pain Responder Index (CLBP-RI) Analysis

CLBP-RI analysis [composite endpoint of average LBPI score, Patient's Global Assessment of Low Back Pain, and RMDQ total score] at Weeks 16, 24, 40 and 56.

For each treatment and each timepoint assessed, the endpoint will be summarized along with 95% CIs. Handling of missing data is same with Section 5.3.

Plot of CLBP-RI analysis at week 16, 24, 40 and 56 (mixed BOCF/LOCF) will be created.

6.2.6. Euro Quality of Life Health State Profile (EQ-5D-5L™)

EQ-5D-5L™ dimensions and overall health utility score at Baseline, Weeks 16 and 56

The Baseline and Weeks 16 and 56 responses in the five dimensions (mobility; self-care; usual activity; pain/discomfort; anxiety/depression) and overall health utility score from the EuroQol 5 Dimensions (EQ-5D-5L) 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 2.3.

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-5L 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, minimum, maximum, number of subjects) will be shown for the health utility score, and the EQ-VAS measure of health today.

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

WPAI:LBP change from Baseline to Week 16, 56 and 64, 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.

This summary of WPAI:LBP will use observed data only (no imputation for missing data). These 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 2.3. The change from Baseline in the impairment scores of the WPAI:LBP will be summarized by treatment group.

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. In addition, descriptive statistics (mean, standard deviation, median, min, max, number of subjects) for the change from Baseline to each timepoint and the absolute scores at the individual timepoints of Baseline and Weeks 16, 56, and 64 will be calculated.

6.2.8. Discontinuation Due to Lack of Efficacy

Incidence of and time to discontinuation due to lack of efficacy

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 summarized along with 95% CIs 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. The differences with 95% CIs will be calculated for the comparisons of each

tanezumab group versus the celecoxib group. A supplementary analysis for the incidence of discontinuation from treatment due to 'INSUFFICIENT CLINICAL RESPONSE' will also be performed.

The time to discontinuation due to lack of efficacy will be summarized and estimated failure curves will be produced using Kaplan-Meier estimation. The time to selected percentiles will also be shown. These are influenced by the frequency of discontinuation, but are expected to be shown for the 1st, 2nd, 5th, 10th and 25th percentiles, in addition to the minimum and maximum time to discontinuation. Any subject who discontinues for any other reason prior to the planned Week 56 visit will be censored at the time of discontinuation. Subjects who complete the study or who discontinue for any reason after the Week 56 visit (including lack of efficacy) will be censored at the Week 56 visit.

6.2.9. Usage of Rescue Medication

Usage of rescue medication (incidence, and number of days of usage) during Weeks 1, 2, 4, 8, 12, 16, 24, 32, 40, 48, 56 and 64.

Incidence of rescue medication use will be summarized up to Week 64 using observed data and LOCF imputation. Imputation of weekly diary data after week 56 will not be performed. Number and percentage of subjects taking rescue medication will be shown.

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 2.3.

The number of days of rescue medication use per week endpoint will be summarized for the Weeks 1, 2, 4, 8, 12, 16, 24, 32, 40, 48, 56 and 64 using observed data and LOCF imputation. Imputation of weekly diary data after week 56 will not be performed. 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. Number of subjects, mean, median, standard deviation, minimum and maximum will be shown.

Usage of rescue medication (amount taken) during Weeks 1, 2, 4, 8, 12, and 16.

The amount of rescue medication use per week will be summarized for the Weeks 1, 2, 4, 8, 12 and 16 using observed data and LOCF imputation. 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. Number of subjects, mean, median, standard deviation, minimum and maximum will be shown.

6.2.10. Health Care Resource Utilization (HCRU)

HCRU at Baseline, Weeks 64 and 80.

Individual items in HCRU at Baseline, Weeks 64 and 80 will be summarized as described in Appendix 2.3. As for questions obtaining number of visits and number of nights stayed in the hospital, n, mean, median, standard deviation, minimum and maximum will be shown. As for questions obtaining binary or categorical responses, number and percentage of each response category will be shown.

6.2.11. Treatment Satisfaction Questionnaire for Medication v.II (TSQM) Score

TSQM score at Weeks 16 and 56.

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.

Summary tables showing number and percentage of patients by value and treatment group will be shown for TSQM items 1-2 and 4-11, and the four satisfaction parameters.

This summary and analysis will use observed data only (no imputation for missing data). The calculation of these endpoints are described in Appendix 2.3.

6.2.12. Patient Reported Treatment Impact Assessment-Modified (mPRTI)

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’).

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).

6.3. Other Endpoints: Tertiary Endpoints

6.3.1. Plasma Tanezumab Concentrations

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.

6.3.2. Serum NGF Assessments

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.

6.4. Subset Analyses

Analysis of the data relating to Total Joint Replacement surgery will be performed using the TJR subset analysis set. Reporting of total joint replacement events including surgery and recovery 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. This data from Study 1063 will be reported separately.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

The following baseline characteristics will be summarized.

- A summary of baseline characteristics. This summary includes average LBPI, 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 (Appendix 4.3).
- Summaries of Baseline painDetect score in continuous scale (Mean/SD/Min/Max) and by painDetect category (painDetect score <= 12, 13 to 18, >=19).
- Plot of the cumulative distribution function of painDetect score

6.5.2. Study Conduct and Subject Disposition

The following subject dispositions will be summarized.

- Summary of number of patients treated by treatment group.
- Summary table and listing of inclusion and exclusion criteria that are not met by subjects who were screened (but not randomized).
- Summary of time to 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, 17-24, 25-32, 33-40, 41-48, 49-56, >56, total) and for the safety follow-up period (Weeks 1-8, 9-16, 17-24, >24, total).
- A summary of discontinuation up to End of Treatment period, and up to End of Study period.

6.5.3. Study Treatment Exposure

The following study treatment exposure will be summarized.

- Number of SC doses taken.
- Summary of oral study medication compliance. This is calculated for the interval from Baseline to 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. The number and percentage of subjects in each category (>=50%, >=80% and >=100% compliance) will be shown.

In addition, summary of duration in treatment period and post-treatment follow-up period will be summarized. The number and percentage of patients with each duration category will also be shown. The duration categories are 1-2, 3-4, 5-8, 9-12, 13-16, 17-24, 25-32, 33-40,

41-48, 49-56, >56 for duration in treatment period and 1-8, 9-16, 17-24, >24 for duration in post-treatment follow-up period, respectively.

6.5.4. Concomitant Medications, Non-Drug Treatments and Drug Treatments Prior to Start of Study Treatment

The following concomitant medications and non-drug treatment will be summarized.

- Concomitant drug treatments during the treatment phase through week 64 or up to 16 weeks post the last SC dose will be summarized. Summary of concomitant medications for non-NSAID medication and drug treatments for CLBP, concomitant NSAID medications, concomitant analgesics not used for CLBP, and concomitant non-drug treatments will be shown separately.

- Concomitant drug treatments after 16 weeks post last SC dose will be summarized for analgesics and other treatment for CLBP.

- Number of days 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 Week 56 to Week 64. 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 will be used. A summary of average number of days of NSAID use will be displayed by each interval. Also, a summary of the overall number of days of NSAID use from Day 1 to Week 64 will be shown, as well as the number and percentage of subjects who exceeded the limit of 80 days of NSAID use during this interval.

The treatment prior to start of study treatment will also be summarized. Summary of analgesics and drug treatments for CLBP, analgesics not used for CLBP, non-analgesics and non-CLBP medication, and non drug treatments prior to start of study treatment will be shown separately.

7. INTERIM ANALYSES

7.1. Introduction

There is no interim analysis for efficacy data 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 external Data Monitoring Committee (E-DMC).

7.2. Interim Analyses and Summaries

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 E-DMC.

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 celecoxib 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.

CCI CCI *and the RR is* C CC *and the difference in the number of subjects with adjudicated events joint safety event 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 E-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 E-DMC.*

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.

8. REFERENCES

1. 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.

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

3. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727-1736.

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

5. 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).

9. APPENDICES

Appendix 1. SUMMARY OF EFFICACY ANALYSES

Note: BL=Baseline

Endpoint	Analysis Set	Statistical Method	Model/Covariates	Missing Data	Objective
Change from Baseline to Weeks 1, 2, 4, 8, 12, 16, 24, 32, 40, 48, 56 and 64 in aLBPI	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in aLBPI	ITT	None (summary)	NA	Multiple imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in aLBPI	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in aLBPI	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 1, 2, 4, 8, 12, 16, 24, 32, 40, 48 and 56 in aLBPI	ITT	ANCOVA	BL score, Treatment group (Study site as a random effect)	Multiple imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 2, 4, 8, 16, 24, 32, 40, 48, 56 and 64 in RMDQ total score	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in RMDQ total score	ITT	None (summary)	NA	Multiple imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in RMDQ total score	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in RMDQ total score	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis

Change from Baseline to Weeks 2, 4, 8, 16, 24, 32, 40, 48 and 56 in RMDQ total score	ITT	ANCOVA	BL score, BL LBPI score, Treatment group (Study site as a random effect)	Multiple imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 2, 4, 8, 16, 24, 32, 40, 48, 56 and 64 in PGA-LBP	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in PGA-LBP	ITT	None (summary)	NA	Multiple imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in PGA-LBP	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in PGA-LBP	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Reduction of >0%, ≥10%, to ≥90% (in steps of 10%) and =100% from Baseline to Week 16, 24 and 56 in the aLBPI	ITT	None (summary)	NA	Observed	Secondary Endpoint Analysis
Reduction of >0%, ≥10%, to ≥90% (in steps of 10%) and =100% from Baseline to Week 16, 24 and 56 in the aLBPI	ITT	None (summary)	NA	Mixed BOCF/LOCF	Secondary Endpoint Analysis
Reduction of >0%, ≥10%, to ≥90% (in steps of 10%) and =100% from Baseline to Week 16, 24 and 56 in the aLBPI	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Reduction of >0%, ≥10%, to ≥90% (in steps of 10%) and =100% from Baseline to Week 16, 24 and 56 in the aLBPI	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 16, 24, 40 and 56 in the aLBPI	ITT	None (summary)	NA	Observed	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 16, 24, 40 and 56 in the aLBPI	ITT	None (summary)	NA	Mixed BOCF/LOCF	Secondary Endpoint Analysis

Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 16, 24, 40 and 56 in the aLBPI	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 16, 24, 40 and 56 in the aLBPI	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 16, 24, 40 and 56 in RMDQ total score	ITT	None (summary)	NA	Observed	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 16, 24, 40 and 56 in RMDQ total score	ITT	None (summary)	NA	Mixed BOCF/LOCF	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 16, 24, 40 and 56 in RMDQ total score	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 16, 24, 40 and 56 in RMDQ total score	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Reduction of >0%, ≥10%, to ≥90% (in steps of 10%) and =100% from Baseline to Week 16 and 56 in RMDQ total score	ITT	None (summary)	NA	Mixed BOCF/LOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 2, 4, 8, 16, 24, 40, 56 and 64 in the BPI-sf Worst Pain	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Worst Pain	ITT	None (summary)	NA	Multiple imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Worst Pain	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Worst Pain	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis

Change from Baseline to Weeks 2, 4, 8, 16, 24, 40, 56 and 64 in the BPI-sf Average Pain	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Average Pain	ITT	None (summary)	NA	Multiple imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Average Pain	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Average Pain	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 2, 4, 8, 16, 24, 40, 56 and 64 in the BPI-sf Pain Interference Index	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference Index	ITT	None (summary)	NA	Multiple imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference Index	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference Index	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 2, 4, 8, 16, 24, 40, 56 and 64 in the BPI-sf Pain Interference with General Activity	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with General Activity	ITT	None (summary)	NA	Multiple imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with General Activity	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with General Activity	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis

Change from Baseline to Weeks 2, 4, 8, 16, 24, 40, 56 and 64 in the BPI-sf Pain Interference with Walking Ability	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with Walking Ability	ITT	None (summary)	NA	Multiple imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with Walking Ability	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with Walking Ability	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 2, 4, 8, 16, 24, 40, 56 and 64 in the BPI-sf Pain Interference with Sleep	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with Sleep	ITT	None (summary)	NA	Multiple imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with Sleep	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with Sleep	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 2, 4, 8, 16, 24, 40, 56 and 64 in the BPI-sf Pain Interference with Normal Work	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with Normal Work	ITT	None (summary)	NA	Multiple imputation	Secondary Endpoint Analysis

Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with Normal Work	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 16 and 56 in the BPI-sf Pain Interference with Normal Work	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
The CLBP-RI response at Weeks 16, 24, 40 and 56	ITT	None (summary)	NA	Observed	Secondary Endpoint Analysis
The CLBP-RI response at Weeks 16, 24, 40 and 56	ITT	None (summary)	NA	Mixed BOCF/LOCF	Secondary Endpoint Analysis
The CLBP-RI response at Weeks 16, 24, 40 and 56	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
The CLBP-RI response at Weeks 16, 24, 40 and 56	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Percentage of subjects with an improvement of ≥2 points from Baseline to Weeks 16, 24, 40 56 in the PGA-LBP	ITT	None (summary)	NA	Observed	Secondary Endpoint Analysis
Percentage of subjects with an improvement of ≥2 points from Baseline to Weeks 16, 24, 40 56 in the PGA-LBP	ITT	None (summary)	NA	Mixed BOCF/LOCF	Secondary Endpoint Analysis
Percentage of subjects with an improvement of ≥2 points from Baseline to Weeks 16, 24, 40 56 in the PGA-LBP	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Percentage of subjects with an improvement of ≥2 points from Baseline to Weeks 16, 24, 40 56 in the PGA-LBP	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Percentage of subjects with any improvement (change<0), no change (change=0), any worsening (change>0) in the PGA-LBP at Weeks 16, 24, 40 and 56	ITT	None (summary)	NA	Observed	Secondary Endpoint Analysis

Percentage of subjects with any improvement (change<0), no change (change=0), any worsening (change>0) in the PGA-LBP at Weeks 16, 24, 40 and 56	ITT	None (summary)	NA	Mixed BOCF/LOCF	Secondary Endpoint Analysis
Percentage of subjects with any improvement (change<0), no change (change=0), any worsening (change>0) in the PGA-LBP at Weeks 16, 24, 40 and 56	ITT	None (summary)	NA	BOCF	Secondary Endpoint Analysis
Percentage of subjects with any improvement (change<0), no change (change=0), any worsening (change>0) in the PGA-LBP at Weeks 16, 24, 40 and 56	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Time to discontinuation due to lack of efficacy (up to Week 56/End of Treatment)	ITT	KM estimates	NA	Observed Data	Secondary Endpoint Analysis
Incidence of rescue medication use during Weeks 1, 2, 4, 8, 12, 16, 24, 32, 40, 48, 56 and 64	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Number of days of rescue medication use during Weeks 1, 2, 4, 8, 12, 16, 24, 32, 40, 48, 56 and 64	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Amount (mg) of rescue medication taken during Weeks 1, 2, 4, 8, 12, 16	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
Incidence of rescue medication use during Weeks 1, 2, 4, 8, 12, 16, 24, 32, 40, 48 and 56	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Number of days of rescue medication use during Weeks 1, 2, 4, 8, 12, 16, 24, 32, 40, 48 and 56	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis

Amount (mg) of rescue medication taken during Weeks 1, 2, 4, 8, 12, 16	ITT	None (summary)	NA	LOCF	Secondary Endpoint Analysis
Change from Baseline to Weeks 16, 56 and 64 in WPAI Impairment scores (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)	ITT	None (summary)	NA	Observed Data	Secondary Endpoint Analysis
EQ-5D dimensions (Mobility; Self-care; Usual activity; Pain/Discomfort; Anxiety/Depression) and Overall Health Utility at Baseline and Weeks 16 and 56	ITT	Summary	NA	Observed Data	Secondary Endpoint Analysis
EQ-VAS at Baseline and Weeks 16 and 56	ITT	Summary	NA	Observed Data	Secondary Endpoint Analysis
mPRTI endpoints (willingness to re-use; patient preference; patient satisfaction) at Weeks 16 and 56	ITT	Summary	NA	Observed	Secondary Endpoint Analysis
TSQM endpoints (satisfaction with effectiveness, side effects and convenience, and overall satisfaction) at Weeks 16 and 56	ITT	Summary	NA	Observed	Secondary Endpoint Analysis
Healthcare Resource Utilization at baseline, week 64 and 80	ITT	Descriptive Summary	NA	Observed	Secondary Endpoint Analysis

Appendix 2. DATA DERIVATION DETAILS

Appendix 2.1. 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, an 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]

As for BPI-sf, the visit window is below.

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	[142,225]
Week 40	281	[226, 309]
Week 48	337	[310, 365]
Week 56	393	[366, 421]
Week 64	449	[422, 477]

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 16 and 56, and ET1. WPAI:LBP is collected at Baseline, Weeks 16, 56, 64, and ET1 and ET2. HCRU is collected at Baseline, Weeks 64 and 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, TSQM, mPRTI

Visit	Target Study Day	Window
Baseline	1 (defined as initial day of study drug administration)	[-7, 1]
Week 16	113	[2, 253]
Week 56	393	[254, no upper limit]]

HCRU

Visit	Target Study Day	Window
Baseline	1 (defined as initial day of study drug administration)	[no lower limit, 1]
Week 64	449	[2,505]
Week 80	561	[506, no upper limit]

WPAI: OA,

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, 421]
Week 64	449	[422, 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 weekly diary average Low Back Pain Intensity scores, the data will be allocated to study weeks based on calendar weeks relative to the baseline visit (study day 1) and not based on the post-baseline visit schedule. The table below describes the visit days for each week. The mean of available weekly data in each of the weekly intervals will be used to calculate the weekely diary average Low Back Pain Intensity scores.

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

At the Screening visit only, the Low Back Pain Intensity (Numeric Rating Scale) will be administered with a patient worksheet for screening purposes.Also, the daily LBPI scores collected during the Initial Pain Assessment Period (Days-5 to -1), will be averaged to assess versus study inclusion criteria.

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 date. 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 Week 16 score, and the available scores from Days 99-101 will be used to calculate the Week 15 score.

After the Week 16 visit, 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 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 2.2. Definition of Protocol Deviations that Relate to Statistical Analyses/Populations

Not applicable.

Appendix 2.3. 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 Japan.

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 impairment while working due to Low Back Pain	Q5*10
Percent overall work impairment 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 this study)

Question	Response	Scoring
During the last 3 months, what services did you receive directly related to your low back pain? • 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 last 3 months, have you visited the emergency room due to your low back pain?	Yes, No	No = 0 Yes = 1
How many times?	Number of visits	0-999
During the last 3 months, have you been hospitalized due to your low back pain?	Yes, No	No = 0 Yes = 1
How many nights in total did you stay in hospital due to your low	Number of Nights	0-999 (max should be 92)

back pain in the last 3 months?		
Did you use these aids or devices to help you in doing things because of your low back pain 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 low back pain?	Yes, No	No = 0 Yes = 1 Not applicable = 2
How long ago did you quit your job because of your low back pain?	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 2.1 for the 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 5.3 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 from 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 3. DATA SET DESCRIPTIONS

Not applicable.

Appendix 4. STATISTICAL METHODOLOGY DETAILS

Appendix 4.1. 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 4.2. Further Details of 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),[4] 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

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

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

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

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

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

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

Appendix 4.3. painDETECT Questionnaire[5]

PainDETECT score will be calculated based on the below table.

Item	Score
Gradation of pain*	
Do you suffer from a burning sensation (eg, stinging nettles) in the marked areas?	0–5
Do you have a tingling or prickling sensation in the area of your pain (like crawling ants or electrical tingling)?	0–5
Is light touching (clothing, a blanket) in this area painful?	0–5
Do you have sudden pain attacks in the area of your pain, like electric shocks?	0–5
Is cold or heat (bath water) in this area occasionally painful?	0–5
Do you suffer from a sensation of numbness in the areas that you marked?	0–5
Does slight pressure in this area, eg, with a finger, trigger pain?	0–5

Pain course pattern	
 Persistent pain with slight fluctuations	0
 Persistent pain with pain attacks	-1
 Pain attacks without pain between them	+1
 Pain attacks with pain between them	+1
Radiating pain	
Does your pain radiate to other regions of your body? Yes/No	+2/0

* For each question: never, 0; hardly noticed, 1; slightly, 2; moderately, 3; strongly, 4; very strongly, 5.