PROTOCOL A4091061

A PHASE 3 RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTICENTER STUDY OF THE ANALGESIC EFFICACY AND SAFETY OF THE SUBCUTANEOUS ADMINISTRATION OF TANEZUMAB (PF-04383119) IN SUBJECTS WITH CANCER PAIN PREDOMINANTLY DUE TO BONE METASTASIS RECEIVING BACKGROUND OPIOID THERAPY

STATISTICAL ANALYSIS PLAN (SAP)

Version: 5.0

Date: 08 January 2021

[redacted]

[redacted]

TABLE OF CONTENTS

[redacted]

[redacted]

LIST OF TABLES

LIST OF FIGURES

APPENDICES

[redacted]

[redacted]

[redacted]

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for study A4091061 is based on the Protocol Amendment 4 dated 14Jun2018.

Table 1 Summary of Major Changes in SAP Amendments

SAP Version	Change	Rationale
1	Not Applicable	Not Applicable
2	Removed analyses involving 10 mg subcutaneous (SC)	To be aligned with Protocol Amendment 3
2	Updated a few analyses	To be aligned with program-level decisions.
3	Revised futility stop criterion and added efficacy stop criterion in the interim analysis; reduced the final sample size	To be aligned with Protocol Amendment 4
3	Updated a few analyses and texts	To be aligned with program-level decisions; Additionally, to implement clarifications, removal of redundant text, and correction of typos
4	Update analysis covariates related to randomization strata	revise analysis model covariates to use the strata as indicated in the project database, due to a high rate of mis-match between randomization strata and project database data
5	Updated analyses	CCI [redacted] The definition of full analysis set is clarified to include all randomized subjects who were randomized to either tanezumab 20 mg or placebo SC and received at least one dose of SC study medication. Subjects randomized to tanezumab 10 mg are excluded.

		The analysis approach to address potentially important protocol deviations is clarified. Per-protocol analysis set will be used when needed. Evaluations of around the clock opioid medications and rescue medications are clarified based on the analysis models specified in the protocol. Imputation of missing data in these medications are updated to support these evaluations. The analyses of treatment interaction are clarified. It is clarified that geographical difference will be examined by region. Definition of region is updated to combine Asia with the single Australian site that randomized one subject. A sensitivity analysis is added to evaluate the potential impact of COVID-19. It is clarified that if the number of subjects with non-index visceral cancer pain sites is < 10, then the ANCOVA will not be applied to this subgroup. CCI [redacted] It is clarified that no early unblinding for efficacy data at Week 24 will be performed. The summary of efficacy analyses in Appendix 1 is updated to align with the updated text.

2. INTRODUCTION

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

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study A4091061. 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

2.1.1. Primary Objective

Demonstrate superior analgesic efficacy of tanezumab 20 mg SC versus matching placebo SC at Week 8 in subjects, with cancer pain predominantly due to bone metastasis, receiving background opioid therapy.

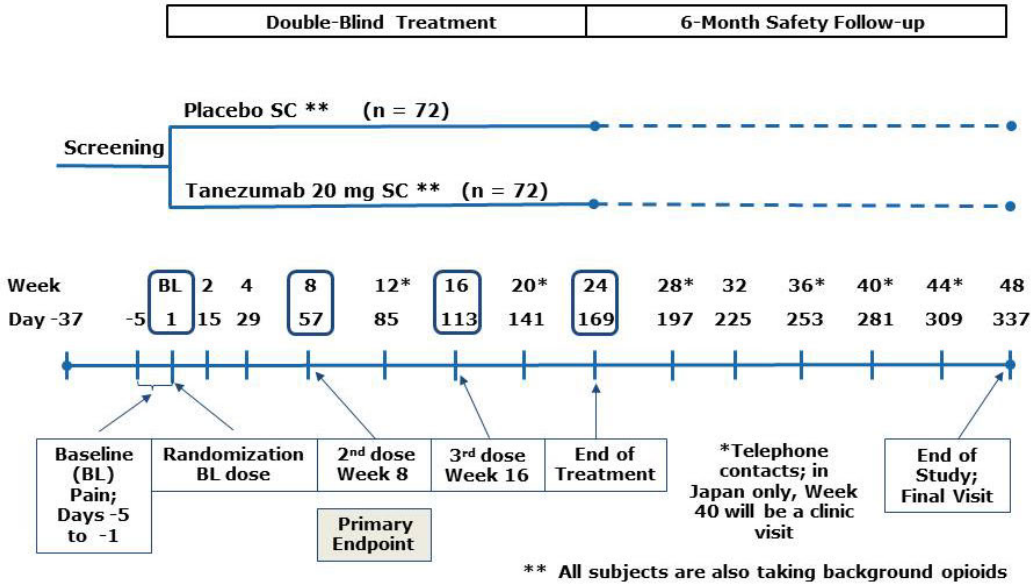
2.1.2. Secondary Objective

Evaluate the safety of tanezumab 20 mg SC versus matching placebo SC in subjects, with cancer pain predominantly due to bone metastasis, receiving background opioid therapy.

2.2. Study Design

The study design is summarized in the diagram below.

Figure 1. Study Design



This is a randomized, double blind, placebo controlled, multicenter, parallel-group Phase 3 study in cancer subjects requiring treatment with background opioids for pain due to bone metastasis.

The protocol was initially designed to include 3 treatment groups (tanezumab 20 mg SC, tanezumab 10 mg SC, and placebo), and was amended (Amendment 3) after study start to discontinue the tanezumab 10 mg dose arm. It is estimated that approximately 11 subjects in

total were randomized to receive tanezumab 10 mg SC prior to implementation of Amendment 3.

Following implementation of Amendment 3, subjects were randomized in a 1:1 ratio (planned 85 subjects/arm) to one of two treatment arms: tanezumab 20 mg SC or matching placebo SC, each administered in addition to background opioids. Subjects who had been randomized to the 10 mg dose treatment arm and who were in the double-blind treatment period at the time of implementation of Amendment 3 were administered 20 mg for any remaining doses.

In protocol Amendment 4, the sample size has been reduced. A total of 155 subjects will be randomized. This comprises 144 subjects (72/treatment arm) randomized to receive tanezumab 20 mg SC or matching placebo SC plus an estimated 11 subjects previously randomized to receive tanezumab 10 mg SC.

Subjects will receive a total of 3 SC injections, separated by 8 weeks in addition to background opioids administered throughout the study. Treatment groups will include:

1. *Placebo SC (matching tanezumab SC) in addition to background opioid therapy.*

2. *Tanezumab 20 mg SC in addition to background opioid therapy.*

The study is designed with a post-randomization duration of 48 weeks and will consist of three periods: Pre-Treatment (up to 37 days), Double-Blind Treatment (24 weeks) and 6-month Safety Follow-up (24 weeks). The Pre-Treatment Period will include a Screening Period (lasting up to 32 days) with washout of prohibited study medication and stabilization of background opioid regimen prior to a 5-day Baseline Assessment Period (BAP). Confirmation of radiographic eligibility by a central radiologist based on protocol-defined x-rays will take place during the Pre-Treatment period. The study is designed such that post randomization, contacts with subjects are made approximately every 4 weeks through the end of the Safety Follow-up period. The Double-Blind Treatment Period consists of 6 in-clinic visits (including 3 dosing visits) and 2 phone contact visits. Because of the long half-life of tanezumab (approximately 21 days), the End of Double-Blind Treatment visit takes place 8 weeks after the last dose of SC medication is administered. The Safety follow-up period begins with the completion of the End of Treatment visit and includes 4 phone contacts and 2 additional in-clinic visits, with the exception of sites in Japan where 3 phone contacts and 3 additional in-clinic visits will occur.

Stratification variables are (i) tumor aggressiveness (assessed by Eastern Cooperative Group [ECOG] performance status and (ii) presence/absence of concomitant anticancer treatment (eg, chemotherapy or hormonal therapy or anti hormonal therapy).

The end of treatment period is at Week 24, with the safety follow-up period up to Week 48. The primary time point for efficacy is Week 8. The period of interest for most study drug associated safety results is the treatment period. Selected safety results will also be provided

separately for the safety follow-up period, and some results will be provided for the combined overall study period comprising the treatment and safety follow-up periods.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint(s)

- *Change from Baseline to Week 8 in the daily average pain intensity in the index bone metastasis cancer pain site.*

  Baseline is defined as the mean average daily pain NRS score during the Baseline assessment period prior to randomization (expected to be 5 days). The Week 8 pain intensity value is the mean of the daily average pain intensity scores for the 7 days prior to the Week 8 visit. See Appendix 2.1 for calculation details if the Week 8 visit is not on the scheduled Day 57. *If any of the seven Week 8 daily scores are missing then the Week 8 value will be calculated over the remaining observations.*

3.2. Secondary Endpoint(s)

3.2.1. Efficacy Measures

- *Change from Baseline to Weeks 1, 2, 4, 6, 12, 16 and 24 in the daily average pain intensity NRS score in the index bone metastasis cancer pain site.*
- *Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16 and 24 in the daily worst pain intensity NRS score in the index bone metastasis cancer pain site.*
- *Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16 and 24 in the weekly average pain intensity NRS score in non-index cancer pain sites.*
- *Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16 and 24 in the weekly worst pain intensity NRS score in non-index cancer pain sites.*
- *Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16 and 24 in the daily average pain intensity NRS score in the non-index visceral cancer pain sites.*
- *Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16 and 24 in the daily worst pain intensity NRS score in the non-index visceral cancer pain site.*

[redacted]

[redacted]

[redacted]

- *Response as defined by a ≥30%, ≥50%, ≥70%, and ≥90% reduction from Baseline in the daily average and daily worst pain intensity NRS score in the index bone metastasis cancer pain site at Weeks 1, 2, 4, 6, 8, 12, 16 and 24.*

- *Change from Baseline (collected at Randomization visit) in Patient's Global Assessment of Cancer Pain at Weeks 2, 4, 8, 16 and 24.*

- *Response defined as an improvement of ≥2 points in Patient's Global Assessment of Cancer Pain at Weeks 2, 4, 8, 16 and 24.*

CCI

CCI

CCI

CCI

3.2.2. Opioid Use and Opioid Adverse Effects Measures

- *Average daily total opioid consumption (in mg of morphine equivalent doses) at Weeks 1, 2, 4, 6, 8, 12, 16 and 24.*

- *Average number of doses of rescue medication required per week at Weeks 1, 2, 4, 6, 8, 12, 16 and 24.*

- *Change from Baseline in the weekly Opioid-Related Symptom Distress Scale at Weeks 2, 4, 8, 16, and 24.*

The Opioid-Related Symptom Distress Scale (OR-SDS) is a questionnaire on the frequency, severity and level of bother of 10 symptoms. For each symptom the mean of

the frequency, severity and bother is calculated to become the Multi-Domain Average (MDA). These are the four dimensions for each symptom. The mean of each dimension over all symptoms is calculated to become the frequency, severity, bother and MDA composite scores.

CCI [redacted]

3.4. Baseline Variables

Baseline is generally defined as the last observation prior to first receipt of study drug, within the baseline window as defined in Appendix 2.1.

For pain intensity scores, *baseline is defined as the mean average daily Pain NRS score during the Baseline Assessment Period prior to Randomization (expected to be 5 days).* It will be calculated as the mean of the non-missing pain scores over study days -5 to -1. If fewer than 5 are available between study days -5 and -1, the baseline will be the mean of the available scores.

Baseline opioid use (daily amount) will be defined as the use during the 5-day Baseline Assessment Period. It will be calculated as the mean of the non-missing opioid use values over study days -5 to -1. If fewer than 5 are available between study days -5 and -1, the baseline will be the mean of the available values.

Region will be used in statistical models. Based on planned study countries, regions are anticipated to be Europe, Middle East, Latin America, Asia (including the single site in Australia).

The stratification variables are (i) tumor aggressiveness (assessed by Eastern Cooperative Group [ECOG] performance status and (ii) presence/absence of concomitant anticancer treatment (eg, chemotherapy or hormonal therapy or anti hormonal therapy).

3.4.1. Covariates

For all models analyzing the continuous primary and secondary efficacy endpoints, the corresponding Baseline value will be used as a covariate, together with the Baseline average pain intensity of the index pain site and baseline total opioid use (using morphine equivalent amount in mg). Region will be fitted as a fixed effect in the ANCOVA models. The randomization stratification variables of tumor aggressiveness and presence/absence of concomitant anticancer treatment will also be included as fixed effects.

- A listing of subjects with mis-matches between the stratification variables entered at randomization and the case report form data will be provided. In analysis models and descriptive summarization the strata as indicated in the project database data will be used.

For the models analyzing the categorical/binary response efficacy endpoints, the model will include terms for Baseline average pain intensity of the index pain site, in addition to the stratification factors.

For response endpoints relating to the Patients Global Assessment of Cancer Pain (PGA-CP), the Baseline PGA-CP will also be used as a covariate, in addition to Baseline average pain intensity of the index pain site and the fixed effect of the stratification factors

For the models analyzing the amount of opioid consumption and number of doses of rescue medication use, the model will include terms for Baseline average pain intensity of the index pain site, stratification factors, and region.

3.5. Safety Endpoints

- *Adverse events.*
- *Standard safety assessments (safety laboratory testing [chemistry, hematology], sitting vital signs, ECG [12-lead]).*
- *Orthostatic (supine/standing) blood pressure assessment.*
- *Weight measurements.*
- *Physical examinations.*
- *Joint safety adjudication outcomes.*
- *Total joint replacements.*

CCI [redacted]

CCI [redacted]

- *Anti-drug antibody (ADA) assessments.*

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 (for example, during the screening or baseline assessment period), or

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

The effective duration of treatment is determined by the lag time. Any event occurring within the lag time, whether this occurs during a break in treatment or at the end of treatment, is attributed to the corresponding treatment period. An infinite lag will be used for the study, meaning any treatment-emergent AE reported in the database will be included in tables of AEs up to end of study.

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

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

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

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

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

Anhidrosis Bradycardia Hypohidrosis	Orthostatic hypotension Syncope

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.

A 3-tier approach will be used to summarize AEs. Under this approach, AEs are classified into 1 of 3 tiers. Different analyses will be performed for different tiers. A description of the three tiers and analyses are given in Section 6.6.1.

All summaries of adverse events will be shown for adverse events that begin or worsen from the first SC dose (treatment-emergent) up to the end of the treatment period. In addition a selection of adverse event tables will be produced for the safety follow-up period and the whole period up to the end of the study, including the treatment period and safety follow-up period.

3.5.2. Vital Signs

The incidence of orthostatic hypotension at each visit, at any treatment period visit (including unscheduled visits) and at any safety follow-up period visit (including unscheduled visits) will be summarized. The definition of orthostatic hypotension is:

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

An additional summary will be provided for 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.

3.5.3. Total Joint Replacement and Surgical Endpoints

A summary of adjudication outcomes (including outcomes of rapidly progressive osteoarthritis (type-1 only), rapidly progressive osteoarthritis (type-2 only), rapidly progressive osteoarthritis (type-1 or type-2 combined), subchondral insufficiency fracture, primary osteonecrosis, and pathological fracture) and total joint replacements will be provided.

Reporting of total joint replacement events including surgery and recovery will be described in a separate Statistical Analysis Plan that will cover patients undergoing total joint replacement from Studies 1059, 1061 and 1063.

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.

If a subject was:

- Randomized but not treated, then that subject will be excluded from all efficacy and safety analyses.
- Treated but not randomized, then by definition that subject will be excluded from the efficacy analyses, but will be reported under the treatment they actually received for all safety analyses.
- Randomized but received incorrect treatment, then that subject will be reported under their randomized treatment group for all efficacy analyses, but will be evaluated on a case-by-case basis for presentation for safety analyses. Decisions will be made before unblinding.

4.1. Full Analysis Set

The modified intent to treat (mITT) analysis set is the primary analysis set for efficacy analyses. It consists of all subjects who were randomized to either tanezumab 20 mg or placebo SC, and received at least one dose of SC study medication. Subjects randomized to tanezumab 10 mg prior to Amendment 3 are excluded from this set. This analysis set is used in the presentations of all efficacy data, and all data listings, and is labeled as the ‘mITT Analysis Set’ or ‘mITT Population’.

4.2. Potentially Important Protocol Deviations

The number of patients with potentially important protocol deviations will be examined to assess their impact on the interpretation of the study results. When necessary, the effect of protocol deviations may be examined by repeating the primary analysis using the per-protocol (PP) analysis set, defined as all subjects in the mITT analysis set who are not major protocol deviators (which would potentially affect efficacy). Examples of major protocol deviators are described below in Section 4.2.1 and 4.2.2.

4.2.1. Major Deviations Assessed Prior to Randomization

- Inclusion criteria: #5-7, 9-11.
- Exclusion criteria: #1 and 4-7
- Randomization criteria: #1, 2, and 4

4.2.2. Major Deviations Assessed Post-Randomization

- Prohibited medications that could affect pain assessments (protocol section 5.8.1) for the Week 8 endpoint, for example:
    - dosing of opioids greater than protocol-specified criteria
    - new use or change to dose of non-opioid or adjuvant analgesics
    - new use or change to dose of anti-neoplastic or bone metastasis therapies outside protocol-specified criteria
    - inadequate washout period of pain medications
    - dosing of analgesia within 48 hours of diary completion

4.3. Safety Analysis Set

The safety analysis set is defined as all subjects treated with tanezumab or placebo SC, including subjects who received tanezumab 10 mg prior to Amendment 3. This analysis set will be labeled as the 'Safety Analysis Set' or 'Safety Population' in the corresponding data analyses and summary presentations.

4.4. Other Analysis Sets

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

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

5.1.1. Statistical Hypotheses

The treatment comparison being made in this study is tanezumab 20 mg versus placebo. For the treatment comparison, the null and alternative hypotheses are shown below (note $\mu_{TREATMENT}$ relates to the mean change from Baseline for the specified treatment group). All tests will be 2-sided.

Comparison of tanezumab versus placebo will be made for data up to and including Week 24.

Null Hypothesis	H0: $\mu_{\text{TANEZUMAB 20mg}} - \mu_{PLACEBO} = 0$
Alternative Hypothesis	H1: $\mu_{\text{TANEZUMAB 20mg}} - \mu_{PLACEBO} \neq 0$

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

5.1.2. Statistical Decision Rules

The Type I error rate (α-level) used in the assessment of treatment comparison for the primary efficacy endpoint is two-sided 5%.

Control of the type I error rate will only apply to the primary treatment comparison versus placebo for the average pain intensity of the index pain site at Week 8 (model with the primary imputation analysis). The overall type I error rate will be allocated across the planned interim analysis and, if any stopping criteria are not met, the final analysis. At each of the interim and the final analysis, the efficacy boundary will be defined using the Lan-DeMets alpha spending function with the O'Brien-Fleming style boundary.

Regardless of the outcome of the analysis of the primary endpoint, other efficacy endpoints will be tested. No adjustment for multiple comparisons will be made for these other efficacy endpoints, and for the safety endpoints. The α-level for each hypothesis test for the secondary and exploratory analyses will be two-sided 5%.

The planned interim analysis will also include the assessment of futility. T*he non-binding futility stopping boundary will be defined also using EAST software (Cytel inc.), using the conditional power of 10% (based on estimated delta/sigma) boundary.*

5.2. General Methods

Subjects, including those randomized before Amendment 3, are randomized at Baseline to one of three treatment groups: placebo SC, tanezumab 10 mg SC, or tanezumab 20 mg SC. These will be labeled as placebo, tanezumab 10 mg, and tanezumab 20 mg for the three

treatment groups respectively. Subjects who were randomized to tanezumab 10 mg SC at baseline and later received tanezumab 20 mg SC after Amendment 3 will be labeled as tanezumab 10/20 mg group. Data from subjects who are randomized to tanezumab 10 mg SC will only be listed or summarized, but not analyzed.

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

The general study design for efficacy, as depicted below, includes a planned treatment period through the Week 24 visit, and a planned 24-week post-treatment safety follow-up period. Efficacy data planned to be collected during this post-treatment safety follow-up period (diary data for subjects who discontinue early) are intended to have efficacy measures contemporaneous to safety observations during this period. They are not intended to assess treatment effects or compare treatment groups. *Unless otherwise stated, all data up to Week 24 will be summarized, and data for the specified time points up to Week 24 will be analyzed.*

1061 Study Visits / Analysis Windows

Week	B	2	4	8	16	24	32	40	48
Day	1	15	29	57	113	169	225	281	337
completer	x	x	x	x	x	x			
drop after Week 8 dose	x	x	x	x	x ET1	CP ET2			
drop after Day 1 dose	x	x	x	x ET1	CP ET2				

B = Baseline; x = collection of most efficacy endpoints; CP = cancer pain diary assessments; ET = Early Termination Visit

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

Efficacy assessments at study visits 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 up to 8 weeks from this last SC dose window is 'on-treatment', and any data in a window 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.

For example, the table below shows on-treatment and off-treatment windows for the planned collection visits for the BPI-sf data during the treatment period:

[redacted]

[redacted]

Last SC Dose Analysis Window	On-treatment Analysis Window Data	Off-treatment Analysis Window Data
Baseline	Weeks 2, 4, 8	Weeks 16, 24
Week 2	Weeks 2, 4, 8	Weeks 16, 24
Week 4	Weeks 2, 4, 8	Weeks 16, 24
Week 8	Weeks 2, 4, 8, 16	Week 24
Week 16	Weeks 2, 4, 8, 16, 24	None
Week 24	Weeks 2, 4, 8, 16, 24	None

Efficacy data collected via subject diary (NRS pain scores and opioid medication use) are collected daily or weekly, not at study visits. Opioid medications collected at study visits that were not recorded in the diary will be evaluated to determine if they were used as rescue medications. Diary efficacy data will be considered on-treatment if it is collected up to 12 weeks (84 days) after the last SC dose. Diary efficacy data collected more than 12 weeks (84 days) after the last SC dose will be considered off-treatment and excluded from summaries and analyses of treatment period efficacy data, ie, for presentations up to Week 24.

5.2.1. A summary of all efficacy analyses is given in Appendix 1. The treatment group ordering in these analyses outputs will be: Placebo, Tanezumab 20mg. In outputs based on safety analysis set, the treatment ordering will be: Placebo, Tanezumab 10 mg, Tanezumab 10/20 mg, Tanezumab 20 mg. The Tanezumab 10/20 mg group includes subjects who have received both tanezumab 10 mg SC prior to Amendment 3 and tanezumab 20 mg SC after Amendment 3.Analyses for Binary Data

Binary response parameters will be analyzed using logistic regression for binary data, with covariates described in Section 5.2.1. Output will show the number and percentage of subjects in each response category, and odds ratios (with 95% CIs) for the treatment comparison shown in Section 5.1.1.

For the daily worst and average pain (in the index site) response efficacy endpoints (defined by a ≥30%, ≥50%, ≥70% and ≥90% change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16 and 24), will be summarized and analyzed using logistic regression for binary data, with model terms for Baseline average/worst pain subscale score, stratification variables, and treatment group. Imputation for missing data will use both LOCF and BOCF, where imputation with BOCF will lead to the subject being assessed as a non-responder for the response endpoint at a particular time point. Also, in order to closely match the primary imputation analysis, a mixed BOCF/LOCF imputation for response endpoints will be used. In this analysis BOCF imputation (ie, a subject would be a non-responder) would be used for missing data due to discontinuation for the reasons of lack of efficacy ('Insufficient Clinical Response' on the End of Treatment Subject Summary Case Report Form)*, adverse event, or death up to the time point of interest, and LOCF imputation would be used for missing data for any other reason.*

The response parameter of an improvement of ≥2 in the Patient Global Assessment of Cancer Pain will be analyzed as described above using mixed BOCF/LOCF as described above for

missing data (using the covariates of Baseline Patient Global Assessment of Cancer Pain and Baseline daily average pain).

The response endpoints of improvement in PGA-CP ≥2 and average and worst pain intensity ≥30, 50, 70 and 90% improvements are analyzed using logistic regression with covariates as defined in Section 3.4.1, at Weeks 1, 2, 4, 6, 8, 12 (pain intensity responses only), 16, and 24. These analyses use mixed BOCF/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.

5.2.2. Analyses for Continuous Data

Primary Analysis

The primary efficacy endpoint will be analyzed using an ANCOVA model, with model terms for Baseline score, the stratification variables, Baseline opioid use (ie, morphine equivalent amount in mg)*, region and treatment group. The stratification variables are (i) tumor aggressiveness (assessed by Eastern Cooperative Group [ECOG] performance status and (ii) presence/absence of concomitant anticancer treatment (eg, chemotherapy or hormonal therapy or anti hormonal therapy).*

The primary analysis of the primary endpoint will use multiple imputation for missing data, to account for uncertainty around the unobserved subject response. The basis for imputing missing values will be dependent on the reasons for missing data. For subjects with missing data due to discontinuation prior to Week 8 for lack of efficacy, or for an adverse event or death, imputation will be based on sampling from a normal distribution using a mean value equal to the subject's Baseline efficacy value and the standard deviation (over the placebo and tanezumab 20 mg groups) of the observed efficacy data at Week 8. For subjects with missing data for any other reason, imputation will be based on sampling from a normal distribution using a mean value equal to subject's last observed efficacy value and standard deviation (over the placebo and tanezumab 20 mg groups) of the observed efficacy data at Week 8. Imputed values will be truncated at 0 and 10, but not rounded. One hundred imputation samples will be used, and the ANCOVA model described above will be used for each imputation dataset. The final results will be calculated using the combined sets of results from each imputation dataset analysis, based on the standard method (Little & Rubin, 2002), which is described in Appendix 3.

Primary Endpoint Sensitivity Analyses

A number of sensitivity analyses will be performed on the primary efficacy endpoint in order to assess the robustness of the conclusions for the primary objective. These relate to the analyses for missing data and the analysis population, and the homogeneity of the results across factors that may influence efficacy. The analyses described below will not be subject

to any multiplicity adjustment. The mITT analysis set is used in the analyses numbered 2 to 5 below, and Per-Protocol analysis set used in analysis number 1 below.

(1) Per-Protocol Analysis Set

When needed (see Section 4.2), the primary analysis described above will be repeated, but using the Per-Protocol analysis set in place of the mITT analysis set. This analysis will assess the robustness of the efficacy conclusions to subjects who have more strictly adhered to protocol inclusion and exclusion criteria, and to protocol defined study procedures.

(2) Alternative Missing Data Analyses

There are four additional analyses that will assess the robustness of the efficacy conclusions to the choice of multiple imputation as the primary method for accounting for missing data.

In the first and second analyses, the primary ANCOVA analysis model described above will be repeated, but using BOCF and LOCF respectively for missing data (note these are single imputation analyses). Note that this analysis along with the same analysis but at Week 16 is a part of secondary analyses; See the section "Secondary Endpoint Analyses" below.

The third sensitivity analysis for the primary endpoint at Week 8 will use a mixed model repeated measures analysis using the observed and imputed data up to Week 8 from the primary multiple imputation analysis, with covariate terms for Time (study week, treated as a categorical variable), Treatment Group, and Time-by-Treatment interaction, and other covariates such as Baseline score, the stratification variables, Baseline opioid use (ie, morphine equivalent amount in mg) and region. The unstructured covariance will be used in the modeling of the within-subject errors in the analysis. See Appendix 2.1 for details on windows.

The fourth sensitivity analysis for the primary endpoint at Week 8 will use a mixed model repeated measures analysis using all observed data up to Week 8, with covariate terms for Time (study week, treated as a categorical variable), Treatment Group, and Time-by-Treatment interaction, and other covariates such as Baseline score, the stratification variables, Baseline opioid use (ie, morphine equivalent amount in mg) and region. The unstructured covariance will be used in the modeling of the within-subject errors in the analysis.

A summary of the missing data pattern will be shown for the average pain intensity for Baseline and Weeks 1, 2, 4, 6, and 8. This summary will show the incidence of subjects with each pattern of observed and missing data over these visits. This summary will be shown overall, and split by treatment group.

(3) Interaction Analyses

Interaction analyses will be performed for the primary endpoint, exploring how these factors affected the treatment effect, one at a time: region, baseline pain score, type of primary

cancer, baseline opioid use, tumor aggressiveness, or concomitant anticancer treatment. These analyses will fit the covariate terms including Baseline score, the stratification variables, Baseline opioid use (ie, morphine equivalent amount in mg), factor being explored and treatment group in addition to the interaction term of treatment group by factor. All factors will be made into categorical variables and included as a fixed effect in the model. Baseline pain score and baseline opioid use will be categorized as ≤ median and > median. It is assumed that all regions will have 10 or more subjects. In the case where regions have <10 subjects, these regions will be collapsed into one category in the interaction analysis. Similar approach will be used for other factors such as the primary cancer type. To aid the interpretation of the treatment by factor interactions, a subgroup analysis within each factor level will be conducted with a point estimate and 95% CI to show how treatment effect varied by the factor.

(4) Analysis only including subjects who are randomized under protocol Amendment 3 and 4

In this study, the majority of subjects are randomized under Amendment 3 and 4. It is estimated that only approximately 11 subjects had been randomized, prior to the time of Amendment 3, to receive tanezumab 10 mg SC. To assess the impact of subjects who are randomized before Amendment 3 on the primary analysis result, the primary analysis will be repeated with inclusion of only subjects who are randomized under protocol Amendent 3 and 4.

(5) Analysis only including data collected prior to the COVID-19 pandemic

To assess the potential impact of COVID-19 on the study, the primary analysis will be repeated while only including data collected before March 11, 2020, the day when the World Health Organization designated COVID-19 as a global pandemic. For sites in China, the analysis will only include data collected before January 9, 2020, the day when COVID-19 was identified as the causative agent of outbreak in Wuhan by the China Center for Disease Control and Prevention.

Secondary Endpoint Analyses

The change from Baseline for the daily average and worst pain intensity in the bone metastasis index cancer pain site will be summarized for each week from 1 to 24. The change from Baseline to Days 1, 2, 3, 4, 5, 6, and 7, and to *Weeks 1, 2, 4, 6, 8 (Worst Pain), 12, 16 and 24 will be analyzed using analysis of covariance (ANCOVA) as described above, using multiple imputation.*

A secondary analysis for the change from Baseline in the daily average pain scores will use a repeated measures mixed effects model, on the available data over Weeks 1 to 24. Estimates for treatment groups and treatment difference for Weeks 1, 2, 4, 6, 8, 12, 16 and 24 will be shown. Additional secondary analysis for the change from Baseline to Week 8 and 16 in the

daily average pain will use single imputation Last Observation Carried Forward (LOCF) and Baseline Observation Carried Forward (BOCF) for missing data.

The mean of the subject's average and worst pain in the non-index cancer sites, over all non-index sites (for up to 2 sites per subject) and for visceral non-index cancer sites will be calculated for Baseline and for each week, and for the change from Baseline. The change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16 and 24 in average and worst pain in the non-index cancer pain sites will be analyzed using analysis of covariance (ANCOVA) as described above, using multiple imputation. For non-index visceral cancer sites, the ANCOVA will not be conducted if the number of available subjects is < 10. *An additional analysis will utilize only those sites where Baseline average/worst pain is at least 5. If a subject has no nominated non-index sites (first analysis) or no nominated non-index sites where Baseline is at least 5 (second analysis) then that subject would be excluded from the respective analysis.* Since some subjects have non-index sites without associated baseline pain scores, a sensitivity analysis will be conducted to examine the change from baseline or screening in weekly average pain intensity in the non-index cancer pain sites.

Additional ANCOVA analyses with multiple imputation for the primary endpoint will be used to examine the interaction of treatment group with region*, Baseline pain score, type of primary cancer (eg, breast, lung, etc), Baseline opioid use and the stratification parameters.*

CCI

The change from Baseline to Weeks 2, 4, 8, 16 and 24 in the Patient Global Assessment of Cancer Pain will be summarized by treatment group, and analyzed using ANCOVA as described above with multiple imputation. A second analysis of this parameter will use Cochran-Mantel-Haenszel test for the change from Baseline to Weeks 2, 4, 8, 16 and 24.

CCI

Daily Opioid Consumption and Doses of Rescue Medication Use

Average daily opioid consumption (mg of morphine equivalent dosage) and average number of doses of rescue opioid consumption per week will be summarized for each week up to Week 24. Percent change from Baseline in average daily opioid consumption will be analyzed using ANCOVA on the rank scores with treatment and the stratification variables as factors. Missing data will be imputed using LOCF. *The average number of doses of rescue opioid consumption per week will be analyzed using a negative binomial model taking into account Baseline daily average pain and Baseline opioid use.*

The number of doses of rescue medication per week will be summarized by treatment group for each week up to Week 24. The total number of doses of rescue medication per week *in Weeks 1, 2, 4, 8, 16, and 24 will be analyzed using a negative-binomial regression model using the log-total number of days of data collection* within the analysis week *as the subject offset variable. The resulting analysis will show the estimated rate of opioids taken as rescue medication for each week. This estimated rate will be shown by treatment group with standard error and 95% CI. The ratio of the opioid usage rate between tanezumab + opioid versus opioid alone will be shown (with standard error and 95% CI).*

In these models the error term is defined with a negative binomial distribution, and 'log' is used as the link function. Output from these analyses will be the estimated amount of opioid use per day, and number of doses of rescue medication per week, in each treatment group, and (following the exponential back transformation) the ratio of medication use for the treatment comparison shown in Section 5.1. The 95% CIs will be given for the estimates of both the individual treatment groups and the treatment group ratio.

The Opioid-Related Symptom Distress Scale (OR-SDS) is a questionnaire on the frequency, severity and level of bother of 10 symptoms. For each symptom the mean of the frequency, severity and bother is calculated to become the Multi-Domain Average (MDA). These are the four dimensions for each symptom. The mean of each dimension over all symptoms is calculated to become the frequency, severity, bother and MDA composite scores. Each of the four dimensions will be summarized by treatment and treatment difference for the 10 symptoms and the overall composite, a total of 44 sets of summary measures. The MDA for each symptom and the four dimensions for the composite score will be analyzed for each time point. The analysis of this data will use a mixed effects repeated measures model and show analysis results for the change from Baseline to Weeks 2, 4, 8, 16, and *24.*

5.2.3. Analyses for Categorical Data

The proportion of subjects who have a reduction from Baseline to Weeks 8, 16 and 24 in average and worst pain of >0%, ≥10% to ≥90% (in steps of 10%), and =100% (ie, average/worst pain is 0) will be tabulated. This will be shown using observed cases and mixed BOCF/LOCF as described above for missing data. The data on reduction from baseline to week 8 using mixed BOCF/LOCF will also be plotted. 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.

The Cochran-Mantel-Haenszel (CMH) test will be performed for the PGA-CP, with additional summaries, for the change from Baseline to Weeks 2, 4, 8, 16, and 24. Changes by each level of improvement will be summarized, as well as any improvement (change<0), any worsening (change>0). This analysis will provide a sensitivity analysis for the ANCOVA analysis of the PGA. The missing data imputation used for this analysis will be mixed BOCF/LOCF.

For any analysis using the CMH test, if there are too few subjects in any stratification combination group (defined as <15 subjects in any of the combinations of stratification factors) then an unstratified test will be performed.

CCI

CCI

5.3. Methods to Manage Missing Data

The primary efficacy endpoint is the change from Baseline to Week 8 in the daily average pain intensity in the index bone metastasis cancer pain site. The primary analysis of the primary endpoint will use multiple imputation for missing data at Week 8 (where the method for imputation will be dependent on the reason for missing data) followed by the ANCOVA analysis with the model described below for the multiple imputed datasets. The imputation strategies are described in the following table.

While the table describes the multiple imputation strategy specifically for the Week 8 time point, multiple imputation analysis at other time points will use the same strategy but with the appropriate time point, eg, 'Week 2' substituted for 'Week 8' in the table.

Type of Missing Data	Imputation Method
Missing data resulting from discontinuation due to Death, Adverse Events (AEs) or Insufficient Clinical Response (Lack of Efficacy, LoE) prior to or during the Week 8 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 8 over all mITT subjects. This is a multiple imputation version of BOCF single imputation method. [Seed 1 below]
Missing data for other reasons, ie, • Subject did not discontinue on or before Week 8 (includes discontinuation for	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

any reason after the end of the Week 8 visit reporting window*) • Subject discontinued for a different reason prior to or during the Week 8 visit reporting window*.	observed efficacy data at Week 8 over all mITT subjects. For example if last observation for a subject is at Week 5, then the imputation sample for that subject is created using the subject's Week 5 observation and the standard deviation of the Week 8 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. [Seed 2 below]

* 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: [1] 1001-1100; [2] 2001-2100. Imputed Week 8 data for the PGA-CP will be rounded to integer scores in the range 1 to 5. Imputed Week 8 data for the average and worst pain intensity and CCI [redacted] that are <0 and >10 will be truncated to 0 and 10, respectively. Imputed Week 8 data for the CCI [redacted] will be rounded to integer scores in the range 0 to 10. The ANCOVA analysis described in Section 5.2.2 (with covariates in Section 3.4.1) will be used for each imputation dataset, and the overall results will be calculated to take account of the variability both within and between imputation datasets using standard methods (Little & Rubin, 2002), which are described in Appendix 9.3.2.

This analysis will be used for the primary efficacy endpoint at Week 8, plus secondary analyses at other time points, and also for a range of secondary efficacy endpoints at all time points. When using the multiple imputation method described above for time points earlier or later than Week 8, then the reason for missing data is assessed up to the end of the window for that particular time point (see Appendix 9.2.1).

Four additional methods will explore the sensitivity of the effect of missing data. The first method of Baseline Observation Carried Forward (BOCF) for missing data at the primary time point of Week 8 will impute the subject's Baseline value for the Week 8 time point, and therefore a zero change from baseline. If a subject's baseline data is also missing then that subject's data remain missing for the post-baseline time point. The second method of Last Observation Carried Forward (LOCF) for missing data at the primary time point of Week 8 will impute the subject's last observed data value for the efficacy endpoint. With LOCF, if a subject is missing all post-baseline efficacy data for a given efficacy endpoint, then

baseline will be carried forward (if baseline is missing then the subject would have no contributing data to be included in the analysis). In both the BOCF and LOCF imputation analyses, the same main effects ANCOVA model as described below will be used. The third method will use Mixed Model for Repeated Measurements (MMRM) utilizing the datasets created by the multiple imputation process up to and including Week 8 (see Appendix 2.1 for details on windows; if multiple observations are within a window, only the single observation selected for analysis by the windowing algorithm will be used in the MMRM analysis). The fourth method will use MMRM utilizing all available data (including off-treatment data) up to and including Week 8.

Analyses of the primary endpoint at secondary time points will use the BOCF and LOCF imputation methods for missing data, and use the same (main effects) ANCOVA model as described for the primary analyses.

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

Note, if Baseline is missing then the subject data for the change from Baseline will be set to missing for all efficacy analyses for that parameter. A subject who has a missing Baseline score will be missing for the response criteria for endpoints where the response is based on one parameter.

CCI [redacted]

For the calculation of the total Around The Clock (ATC) opioid consumption while subjects are still in the study, any missing ATC opioids data will be imputed by carrying forward the last recorded daily data up to Week 8 (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 average ATC opioid consumption can then be calculated for each week the subject is in the study. To derive the endpoint of weekly average total daily opioid consumption, add the respective weekly average daily rescue opioid consumption, which is calculated by dividing the observed total rescue opioid consumption by the number of days during the observation period. Percent change from baseline in this endpoint will be analyzed using ANCOVA on the rank scores. The last weekly score for the average total daily opioid

consumption will be used for weeks (LOCF weekly data) 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, e.g., if Week 8 is missing then Week 7 data can be used).

For rescue opioids, the average daily number of doses at a week will be computed as the total number of doses observed divided by the number of days under observation for that week. The endpoint of average number of doses of rescue medication required per week will be analyzed with a negative binomial regression model, in which the response variable is the total number of doses during the week, and the offset variable is log of the number of days under observation.

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 3-8. Imputation of weekly diary data (LOCF weekly data) after Week 8 will use the last available weekly diary data score available.

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

For the non-diary secondary endpoints based on CCI PGA-CP CCI if data were not collected during the 5-day BAP period, the closest measurements prior to randomization can be used as baseline.

CCI

6. ANALYSES AND SUMMARIES

A summary of the details of the efficacy analyses is presented in tabular format in Appendix 1.

6.1. Primary Endpoint(s)

See Appendix 1.

6.2. Secondary Endpoint(s)

See Appendix 1.

CCI [redacted]

[redacted]

CCI [redacted]

[redacted]

6.5. Baseline and Other Summaries and Analyses

The following non-standard baseline tables will be included:

- A summary of baseline characteristics. This summary includes BPI-sf average pain and worst pain, PGA-CP, metastatic disease characteristics (eg, location of primary tumor, time since histopathological diagnosis, time since bone metastasis diagnosis of metastatic disease), diabetes status (from medical history and/or pre-treatment HbA1c $\geq$ 6.5%), and stratification parameters. This summary will also include a summary of the number of subjects who are $\geq$75 years old.

6.5.1. Concomitant Medications and Non-Drug Treatments

Summaries of various classes of concomitant medications based on Case Report Form classifications will be provided, eg, treatments for bone metastasis, non-NSAID and NSAID medications (shown separately).

6.6. Safety Summaries and Analyses

Adverse events, concomitant medications, laboratory safety tests, physical and CCI *examinations, vital signs, ECGs, the anti-drug antibody test will be collected for each subject during the study according to the Schedule of Assessments. Standard safety reporting tables will summarize and list the safety data.*

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.

The following non-standard safety tables will also be included

- Summary of number of patients treated by region, country, and treatment group.
- Incidence and severity of Adverse Events leading to discontinuation.
- Summary of AEs, Incidence of AEs, Incidence of AEs leading to discontinuation and summary of Serious AEs will be shown for the whole study period (including the safety follow-up period).

CCI

CCI

- Summary of the Incidence of sympathetic neuropathy based on investigator assessment and, if performed, expert consultant assessment.
- 'Incidence and severity' tables of treatment-emergent adverse events of Abnormal Peripheral Sensation (APS) and Sympathetic Nervous Function, as defined above. Other adverse events may be added to these groupings if they are observed in this study or other studies in the tanezumab program.
- Summary of inclusion and exclusion criteria that are not met by subjects who were screened (but not randomized).
- Summary of discontinuation by treatment group and reason, and study week of discontinuation for the treatment period (Weeks 1-2, 3-4, 5-8, 9-12, 13-16, 17-24, >24), and for the safety follow-up period (Weeks 1-8, 9-16, 17-24, >24 posttreatment).
- 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 to 20, >20 to 30, >30, and (diastolic BP) only decreases or no change, >0 to 10, >10 to 20, >20.

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

- A summary of incidence of subjects with confirmed orthostatic hypotension, for each visit and any post-baseline incidence of orthostatic hypotension.

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

- Incidence of musculoskeletal physical examination at screening.

CCI [redacted]

- Summary of concomitant medications for cancer pain for non-NSAID and NSAID medications (shown separately).

- Summary of number of days of NSAID use per dosing interval (eg, Day 1 to Week 8, Week 8 to Week 16, Week 16 to Week 24) and for the first 8-week interval in the safety follow-up period. This will show the number and percentage of subjects in an interval who exceeded the limit of 10 days of NSAID use. If an 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 interval. Also, a summary of the overall number of days of NSAID use from Day 1 to Week 32 will be shown, as well as the number and percentage of subjects who exceeded the limit of 36 days of NSAID use during this interval.

- Summary of radiation or radiopharmaceutical therapy.

6.6.1. Adverse Events

Adverse Events of Abnormal Peripheral Sensation will be summarized. CCI [redacted]

CCI [redacted]

[redacted]

[redacted]

Separate adverse event summaries by treatment group for adverse events of decreased sympathetic function will be conducted. More specifically, adverse events with the following preferred terms will be considered to represent adverse events of decreased sympathetic function: Blood pressure orthostatic decreased, bradycardia, dizziness postural, heart rate decreased, orthostatic hypotension, presyncope, sinus bradycardia, syncope, anhidrosis, hypohidrosis, abdominal discomfort, diarrhea, early satiety, fecal incontinence, nausea, vomiting, ejaculation delay, ejaculation disorder, ejaculation failure, hypertonic bladder, micturition urgency, nocturia, urinary frequency, urinary hesitation, urinary incontinence, respiratory distress and respiratory failure. If necessary, this list of preferred terms may be adjusted for updates made to the MEDICAL DICTIONARY FOR DRUG REGULATORY AFFAIRS (MedDRA) dictionary versions used for reporting.

In addition to summaries of adverse events considered to represent adverse events of decreased sympathetic function noted above, adverse events of syncope, bradycardia, orthostatic hypotension, anhidrosis, or hypohidrosis are designated as adverse events of interest that will be reviewed by the unblinded E-DMC.

Selected adverse events of interest and common adverse events will be summarized using Risk Differences (with 95% confidence intervals) between the tanezumab 20 mg group and placebo. In addition, significance testing will be performed for adverse events of interest between the tanezumab 20 mg group and placebo. There will be no multiplicity adjustment for these significance tests.

For the 3-tier adverse event reporting, tier 1 adverse events are defined in the tanezumab Safety Review Plan, and this definition of tier-1 adverse events for the report of study 1061 tables will be finalized prior to the unblinding of this study.

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

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

Adverse events within tier 1 and 2 will be summarized using Risk Differences between tanezumab group and placebo, together with 95% confidence interval, using exact methods. Significance tests will be performed for the tier 1 adverse events. There will be no multiplicity adjustment for these significance tests. These summaries and analyses will also be provided for subjects who are randomized only after Amendment 3, along with incidence and severity summary of treatment-emergent adverse events of Abnormal Peripheral Sensation (APS) and Sympathetic Nervous Function.

The following footnote will be used in the 3-tier AE tables: “P-values and confidence intervals are not adjusted for multiplicity and should be used for screening purposes only. 95% Confidence intervals are provided to help gauge the precision of the estimates for Risk Difference. Risk Difference is computed as ‘Tanezumab 20 mg versus placebo’. Exact methods are used for 95% confidence intervals and significance tests.”

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.

All summaries of adverse events will be shown for adverse events that begin or worsen after the first dose of study drug (treatment-emergent) up to the end of the treatment period. In addition, a selection of adverse event tables will be produced for the off-study medication safety follow-up period, and some will be produced for the whole period up to the end of the study, including the treatment period and safety follow-up period.

6.6.2. Vital Signs

Incidence of orthostatic hypotension using postural changes in blood pressure will be summarized.

6.6.4. Immunogenicity

The following assessments of ADA data will be made:

- A listing of individual serum ADA results sorted by treatment group, subject ID and planned visit. The listing will also include the actual test date/times.

- The proportion of subjects who test positive (ie, develop anti-tanezumab antibodies) and negative will be summarized by treatment group and planned visit. The summary will also include the proportion of subjects who test positive and negative overall in the study.

- Subjects who develop anti-tanezumab antibodies after treatment will be evaluated for the presence of anti-tanezumab neutralizing antibodies, and individual results will be listed.

- Individual subjects with positive ADA results will be evaluated for potential ADA impact on the individual's CCI efficacy and safety profile.

CCI

CCI

6.7.3. Joint Safety Events

The incidence of subjects with any of the joint safety adjudication outcomes of rapidly progressive osteoarthritis (type 1 and type 2), subchondral insufficiency fracture, primary osteonecrosis, or pathological fracture, and for occurrence of total joint replacement will be shown by number of subjects treated and subject years of exposure (treatment plus follow up periods), for individual treatment groups.

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.

7. INTERIM ANALYSES

7.1. Introduction

The following sections describe regular safety data monitoring as well as the single interim analysis to assess futility and the evidence of efficacy in the primary efficacy endpoint.

7.2. Interim Analyses and Summaries

Safety data monitoring

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

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

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

The protocol (or treatment group) stopping rule will be based on the assessment of the number of subjects with adjudicated events of interest (rapidly progressive osteoarthritis type 2, subchondral insufficiency fractures, primary osteonecrosis, or pathological fracture) during the course of the study. Assuming the rate of adjudicated events in the placebo group is no more than 6 per 1000 patient-years, if adjudicated events of interest are reported in 3 or more subjects in any tanezumab treatment group than for the placebo treatment group, a treatment-group or protocol-based stopping rule will be triggered. If the rate of events in the placebo treatment group is higher than 6 per 1000 patient-yrs the appropriate threshold number of events for the stopping rule will be reassessed. 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 are described in Section 9.7.1.1 of the protocol.

Programming and review of unblinded outputs will be performed by individuals independent of the study team.

A single interim analysis for futility or efficacy

The study is designed as a Group Sequential Design using a single interim analysis. The purpose of the interim analysis is to assess the non-binding futility and the evidence of efficacy of the primary efficacy parameter.

The interim analysis will be performed when at least 50% (36 from each treatment group) of subjects have completed or discontinued prior to Week 8. Only subjects in countries with approval of protocol amendment 4 will be included in the interim analysis. *Based on current enrollment, it is expected that the interim analysis will include approximately 36 to 45 subjects per treatment group* who have completed or discontinued prior to Week 8.

The non-binding futility stopping boundary will be defined using EAST (Cytel Inc.), using the conditional power of 10% (based on estimated delta/sigma) boundary, for a one-sided assessment of futility. The efficacy stopping boundary will be defined using the same software, using the Lan-DeMets alpha spending function with the O'Brien-Fleming style boundary, for a one-sided assessment of efficacy. The resulting stopping boundary z scores both for futility and efficacy will be determined by the proportion of subjects included in the interim analysis (over the 72 planned for the final sample size per treatment), i.e., information fraction.

The analysis will be conducted by a statistician outside of Pfizer, with all details of the treatment allocation, analysis and results unknown to all within Pfizer or the contract research organization (CRO) performing the final analysis programming. In the situation where the futility or efficacy stopping rule is met, then the statistician will convey the results of the interim analysis to a senior statistician within Pfizer (but outside the study team) to repeat and ratify the analysis. The decision to stop the study will be made by Sponsor Management as defined in the Interim Analysis Charter. *Results of the interim analysis will not be known to anybody in the study team until the formal unblinding of the study at database lock.*

Before any interim analysis is initiated, the details of the objectives, decision criteria, dissemination plan, and method of maintaining the study blind as per Pfizer's standard operating procedures (SOPs) will be documented and approved in an Interim Analysis charter. In addition, the analysis details must be documented and approved in an interim analysis SAP.

At the time when the SAP amendment 5 is issued, the planned single interim analysis has been performed according to the interim analysis SAP. The interpretation of final analysis of the primary endpoint will take into account the type I error (alpha) spent at the interim analysis.

8. REFERENCES

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

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

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 Week 8 in daily average pain intensity in the index bone metastasis cancer pain site (aPI-IBM)	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Primary Analysis
Change from Baseline to Week 8 in aPI-IBM	PP	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Sensitivity Analysis (Per protocol)
Change from Baseline to Week 8 in aPI-IBM	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use	BOCF	Sensitivity/ Secondary Analysis
Change from Baseline to Week 8 in aPI-IBM	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use	LOCF	Sensitivity/ Secondary Analysis
Change from Baseline to Weeks 1, 2, 4, 6, and 8 in aPI-IBM	mITT	MMRM	BL Score, Time, Treatment Group, stratification factors, region, Time x Treatment Group, BL opioid use	Multiple Imputation Data	Sensitivity Analysis for Week 8 (Secondary Analysis for other time points)
Change from Baseline to Weeks 1, 2, 4, 6, and 8 in aPI-IBM	mITT	MMRM	BL Score, Time, Treatment Group, stratification factors, region, Time x Treatment Group, BL opioid use	Observed Data	Sensitivity Analysis for Week 8 (Secondary Analysis for other time points)
Change from Baseline to Week 8 in aPI-IBM based on data collected prior to January 9th, 2020 in China and prior to March 11th, 2020 everywhere else	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Sensitivity Analysis to assess the potential impact of COVID-19

[redacted]

[redacted]

Endpoint	Analysis Set	Statistical Method	Model/ Covariates	Missing Data	Objective
Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in aPI-IBM	mITT	MMRM	BL Score, Time, Treatment Group, stratification factors, region, Time x Treatment Group, BL opioid use	Observed Data	Secondary Analysis
Missing data pattern for aPI-IBM for Baseline and Weeks 1, 2, 4, 6, and 8	mITT	None (summary)	NA	Observed Data	Supportive summary for missing data
Change from Baseline to Week 8 in aPI-IBM (regions with n ≥ 10)	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use, region x Treatment Group	Multiple Imputation	Additional (Interaction) Analysis
Change from Baseline to Week 8 in aPI-IBM by baseline pain score (<= median, >median)	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, Region, BL opioid use, BL pain score category, BL pain score category x Treatment Group	Multiple Imputation	Additional (Interaction) Analysis
Change from Baseline to Week 8 in aPI-IBM by type of primary cancer (breast, lung, etc.,)	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, Region, BL opioid use, type of cancer, type of cancer x Treatment Group	Multiple Imputation	Additional (Interaction) Analysis
Change from Baseline to Week 8 in aPI-IBM by baseline opioid use (<= median, >median)	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, Region, BL opioid use, BL opioid use category, BL opioid use category x Treatment Group	Multiple Imputation	Additional (Interaction) Analysis
Change from Baseline to Week 8 in aPI-IBM by tumor aggressiveness (ECOG performance status)	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, Region, BL opioid use, tumor aggressiveness, tumor aggressiveness x Treatment Group	Multiple Imputation	Additional (Interaction) Analysis
Change from Baseline to Week 8 in aPI-IBM by concomitant anticancer treatment (presence/absence)	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, Region, BL opioid use, concomitant anticancer treatment, concomitant anticancer treatment x Treatment Group	Multiple Imputation	Additional (Interaction) Analysis

Endpoint	Analysis Set	Statistical Method	Model/ Covariates	Missing Data	Objective
Change from Baseline to Week 8 in aPI-IBM, shown by region (regions with n ≥ 10)	mITT	None (summary)	NA	Multiple Imputation	Supportive summary for interaction analysis
Change from Baseline to Week 8 in aPI-IBM, shown by baseline pain score (<= median, >median)	mITT	None (summary)	NA	Multiple Imputation	Supportive summary for interaction analysis
Change from Baseline to Week 8 in aPI-IBM, shown by type of primary cancer (breast, lung, etc.,)	mITT	None (summary)	NA	Multiple Imputation	Supportive summary for interaction analysis
Change from Baseline to Week 8 in aPI-IBM, shown by baseline opioid use (<= median, >median)	mITT	None (summary)	NA	Multiple Imputation	Supportive summary for interaction analysis
Change from Baseline to Week 8 in aPI-IBM, shown by tumor aggressiveness (ECOG performance status)	mITT	None (summary)	NA	Multiple Imputation	Supportive summary for interaction analysis
Change from Baseline to Week 8 in aPI-IBM, shown by concomitant anticancer treatment (presence/absence)	mITT	None (summary)	NA	Multiple Imputation	Supportive summary for interaction analysis
Change from Baseline to Week 8 in aPI-IBM (including subjects only randomized under Amendment 3)	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Additional Analysis
Change from Baseline to Days 1, 2, 3, 4, 5, 6, 7, and to Weeks 1, 2, 4, 6, 12, 16, and 24 in aPI-IBM	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, , BL opioid use	Multiple Imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in worst PI-IBM	mITT	ANCOVA	BL Score, BL aPI-IBM, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Secondary Endpoint Analysis
Change from Baseline to Week 16 in aPI-IBM	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use	BOCF	Secondary Analysis

Endpoint	Analysis Set	Statistical Method	Model/ Covariates	Missing Data	Objective
Change from Baseline to Week 16 in aPI-IBM	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use	LOCF	Secondary Analysis
Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in weekly average PI non-index cancer pain sites	mITT	ANCOVA	BL Score, BL aPI-IBM, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in weekly worst PI non-index cancer pain sites	mITT	ANCOVA	BL Score, BL aPI-IBM, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in weekly average PI non-index cancer pain sites (only include sites where Baseline average pain is at least 5)	mITT	ANCOVA	BL Score, BL aPI-IBM, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in weekly worst PI non-index cancer pain sites (only include sites where Baseline worst pain is at least 5)	mITT	ANCOVA	BL Score, BL aPI-IBM, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Secondary Endpoint Analysis
Change from Baseline to Week 8 in aPI-IBM	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use, cancer pain site, cancer pain site x Treatment Group	Multiple Imputation	Secondary Analysis –Interaction Analysis
Change from Baseline to Week 8 in aPI-IBM	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use, BL score x Treatment Group	Multiple Imputation	Secondary Analysis –Interaction Analysis
Change from Baseline to Week 8 in aPI-IBM	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use, type of primary cancer, type of primary cancer x Treatment Group	Multiple Imputation	Secondary Analysis –Interaction Analysis

Endpoint	Analysis Set	Statistical Method	Model/ Covariates	Missing Data	Objective
Change from Baseline to Week 8 in aPI-IBM	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use, BL opioid use x Treatment Group	Multiple Imputation	Secondary Analysis –Interaction Analysis
Change from Baseline to Week 8 in aPI-IBM	mITT	ANCOVA	BL Score, Treatment Group, stratification factors, region, BL opioid use, stratification factors x Treatment Group	Multiple Imputation	Secondary Analysis –Interaction Analysis
Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in daily average PI non-index visceral cancer pain sites	mITT	ANCOVA	BL Score, BL aPI-IBM, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in daily worst PI non-index visceral cancer pain sites	mITT	ANCOVA	BL Score, BL aPI-IBM, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Secondary Endpoint Analysis
[redacted]					

Endpoint	Analysis Set	Statistical Method	Model/ Covariates	Missing Data	Objective
[redacted]					
Change from Baseline to Weeks 2, 4, 8, 16, and 24 in PGA-CP	mITT	ANCOVA	BL Score, BL aPI-IBM, Treatment Group, stratification factors, region, BL opioid use	Multiple Imputation	Secondary Endpoint Analysis
Change from Baseline to Weeks 2, 4, 8, 16, and 24 in PGA-CP	mITT	CMH test	Treatment Group [1]	Mixed BOCF/LOCF	Sensitivity Analysis for PGA
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in the aPI -IBM	mITT	Logistic Regression	BL Score, Treatment Group, stratification factors	Mixed BOCF/LOCF	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in the aPI -IBM	mITT	Logistic Regression	BL Score, Treatment Group, stratification factors	BOCF	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in the aPI -IBM	mITT	Logistic Regression	BL Score, Treatment Group, stratification factors	LOCF	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in the worst PI – IBM	mITT	Logistic Regression	BL Score, BL aPI-IBM, Treatment Group, stratification factors	Mixed BOCF/LOCF	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in the worst PI – IBM	mITT	Logistic Regression	BL Score, BL aPI-IBM, Treatment Group, stratification factors	BOCF	Secondary Endpoint Analysis
Percentage of subjects with reduction of ≥30/50/70/90% from Baseline to Weeks 1, 2, 4, 6, 8, 12, 16, and 24 in the worst PI -IBM	mITT	Logistic Regression	BL Score, BL aPI-IBM, Treatment Group, stratification factors	LOCF	Secondary Endpoint Analysis
Percentage of subjects with an improvement of ≥2 points from Baseline	mITT	Logistic Regression	BL Score, BL aPI-IBM, Treatment Group, stratification factors	Mixed BOCF/LOCF	Secondary Endpoint Analysis

Endpoint	Analysis Set	Statistical Method	Model/ Covariates	Missing Data	Objective
to Weeks 2, 4, 8, 16, and 24 in the PGA-CP					
Reduction of >0%, ≥10%, to ≥90% (in steps of 10%) and =100% from Baseline to Weeks 8, 16 and 24 in the aPI-IBM	mITT	None (summary)	NA	Observed	Secondary Endpoint Analysis
Reduction of >0%, ≥10%, to ≥90% (in steps of 10%) and =100% from Baseline to Weeks 8, 16 and 24 in the aPI-IBM	mITT	None (summary, and plot for week 8)	NA	Mixed BOCF/LOCF	Secondary Endpoint Analysis
Reduction of >0%, ≥10%, to ≥90% (in steps of 10%) and =100% from Baseline to Weeks 8, 16 and 24 in the worst PI-IBM	mITT	None (summary)	NA	Observed	Secondary Endpoint Analysis
Reduction of >0%, ≥10%, to ≥90% (in steps of 10%) and =100% from Baseline to Week 8, 16 and 24 in the worst PI-IBM	mITT	None (summary, and plot for week 8)	NA	Mixed BOCF/LOCF	Secondary Endpoint Analysis
Percent change from Baseline in average daily opioid consumption at Weeks 1, 2, 4, 6, 8, 12, 16, and 24 (Include both ATC and rescue opioids)	mITT	ANCOVA on the rank scores	BL aPI-IBM, Treatment Group, stratification factors, region	LOCF	Secondary Endpoint Analysis
Total number of doses of rescue medication at Weeks 1, 2, 4, 6, 8, 12, 16, and 24 (The resulting analysis will show the estimated rate of opioids taken as rescue medication for each week.)	mITT	Negative Binomial model (log total number of days of data collection within the analysis week as offset variable)	BL aPI-IBM, Treatment Group, stratification factors, region, BL opioid use. Use the log-total number of days of data collection as the subject offset variable.	Observed Data within Each Week, LOCF for Weeks That Are Missing	Secondary Endpoint Analysis
Change from Baseline in the weekly Opioid-Related Symptom Distress Scale at Weeks 2, 4, 8, 16, and 24	mITT	MMRM	BL Score, Time, Treatment Group, stratification factors, region, Time x Treatment Group	Observed Data	Secondary Endpoint Analysis

[redacted]

[redacted]

Endpoint	Analysis Set	Statistical Method	Model/ Covariates	Missing Data	Objective
[redacted]					

[1] CMH test will be stratified by the levels of the combined stratification parameters (4 levels). If there are <15 subjects in any combined stratification level then the CMH test will be unstratified

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, and 48. 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 each planned study visit (eg, CCI PGA-CP 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	Variable (typically up to 37 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]

CCI

CCI

CCI

For the average and worst pain intensity NRS scores, the data are collected daily via electronic diary up to the end of Week 8, and thereafter weekly up to Week 24.

The Baseline score is the mean of the non-missing pain scores over study days -5 to -1. If fewer than 5 are available between study days -5 and -1, the baseline will be the mean of the available scores.

The table below describes the visit days for each week (Weeks 1-8). All available diary data in each of the weekly intervals will be used to calculate the mean daily pain score for that study week.

Study Week	Days
1	1-7
2	8-14
3	15-21
4	22-28
5	29-35
6	36-42
7	43-49
8	50-56

However, if a subject receives the Week 8 injection dose prior to Day 57, the Week 8 score will be calculated using the mean of the available scores from the 7 calendar days immediately prior to the Week 8 injection date. Any scores used in this calculation of Week 8 will not also be used in an earlier week calculation, eg, if the Week 8 dose occurs on Day 53, the available scores from Days 46-52 will be used to calculate the average score for Week 8, and the available scores from Days 43-45 will be used to calculate the average score for Week 7.

After the Week 8 visit, pain scores are captured only once a week in the diary. These are grouped in 4-week intervals using visit windows as shown below. If a subject comes in late for a Week 8 visit (or weekly diary is not activated at the visit), and so has daily diary data collected past Day 56, these data will be averaged with any data obtained weekly for any given interval.

Summary Week	Includes Weeks	Days
12	9 - 12	57-84
16	13 - 16	85-112
20	17 - 20	113-140
24	21 - 24	141-168

Appendix 3. STATISTICAL METHODOLOGY DETAILS

Appendix 3.1. Further Details of Safety Data Monitoring

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 3.2. Further Details of the Statistical Methods

A description of the combination of the ANCOVA results from each of the multiple imputed datasets is given below, and taken from Little & Rubin (2002),[2] 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

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

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

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

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

Appendix 3.4. ATC and Rescue Opioid Medication Endpoints

ATC and rescue opioid medication (RM) data are collected daily using an electronic system up to Week 8, and during selected weeks (weekly) after Week 8 and up to Week 24. 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 a similar principle as described above in Appendix 2.1 for the daily and weekly pain data.

In the case of a week based on data from fewer than 7 days due to the implementation of the windowing algorithm, the weekly amount and number of doses calculation will be adjusted to a 7-day period to account for the smaller number of days assessed.

The number of doses of RM use (using daily and weekly data) and the total amount taken over the week will be calculated for the assigned week algorithm described above.

Imputation is described in Section 5.3. Imputation of daily data occurs for ATC opioids up to Week 8 where the patient is in the trial. If the patient discontinues the study before Week

8, then the imputation occurs only up to the end of the week in which patient discontinued the study. Afterwards, weekly data are imputed.

An example of imputation and calculating the two endpoints using the daily diary data is shown below.

Example

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 ATC and rescue medication example data. In this example, the subject's Week 2 visit was on study day 14.

Study Day (Week 2)	Morphine equivalent of ATC Medication taken (mg)	Morphine equivalent of ATC Medication taken with LOCF imputation (mg)	Episodes of Rescue Medication taken in morphine equivalent dose
8	100	100	50 mg, 50 mg
9	Missing	100 [1]	Missing
10	100	100	0
11	100	100	50 mg
12	Missing	100 [1]	Missing
13	100	100	50 mg, 50 mg
14	100	100	50 mg, 50 mg

[1] Using LOCF imputation for missing data

For this subject, the two endpoints on opioid use are calculated as the following:

- Average daily total opioid consumption (in mg of morphine equivalent doses) at Week 2 = 150 mg, in which 100 mg is from the average daily ATC opioid consumption and 50 mg is from the average daily rescue opioid consumption.

- Average number of doses of rescue medication required per week at Week 2 =1.