Statistical Analysis Plan (SAP)

Protocol Title:	A Phase 2b, Multicentre, Multinational, Double-blind, Dose-finding Study, incorporating an open label substudy, in Adult Patients with Type I, III or IV Osteogenesis Imperfecta Treated with setrusumab (BPS804)
Protocol Number:	MBPS205
Protocol Version, Date	Version 7, 19 July 2019
ICON ID:	3082/0003
Document Version, Date:	1.0, 22 October 2019

Prepared by:

ICON Clinical Research Services

On behalf of:

Mereo BioPharma 3 Ltd. 1 Cavendish Place, London, W1G 0QF

SIGNATURE PAGE

Prepared at ICON Clinical Research by:

[redacted] 25 Oct 2019 03:12:004+0000
REASON: I approve this document as author.
412c0b51-c5cb-4161-82a6-5debf620771c

[redacted] Senior Statistician I — Date (DD Mmm YYYY)

Reviewed at ICON Clinical Research by:

[redacted] 24 Oct 2019 14:44:028+0000
REASON: I approve this document
70e4f952-b655-4aad-9112-4a377520c5d4

[redacted], Senior Programming Manager
Biostatistics and Programming — Date (DD Mmm YYYY)

Approved at Mereo BioPharma 3 Ltd by:

[redacted] PhD 24 Oct 2019 14:58:046+0000
REASON: I approve this document
81466017-e24d-4e49-9d5d-ce762a2ad2d8

[redacted] Head of Clinical Operations
Mereo BioPharma 3 Ltd — Date (DD Mmm YYYY)

REVISION HISTORY

Version/Date	Version name	Section	Changes implemented
1.0/ 22OCT2019	Initial approved version	N/A	N/A

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

The following abbreviations will be used within this SAP.

Abbreviation or special term	Explanation
ADA	Anti-drug antibody
AE	Adverse Event
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
BMD	Bone mineral density
BMI	Body Mass Index
BV/TV	Bone volume fraction
CI	Confidence Interval
CRF	Case Report Form
CS	Clinically Significant
CSR	Clinical Study Report
CTMS	Clinical Trial Management System
DXA	Dual-energy x-ray absorptiometry
ED_{50}	Median Effective Dose
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EQ-5D-5L	Euroqol 5-dimension 5-level descriptive system
FAS	Full analysis set
FEA	Finite Element Analysis
HR	Heart Rate
HRpQCT	High Resolution peripheral Quantitative Computed Tomography
ICH	International Conference on Harmonisation
IWRS	Interactive Web Response System
IV	Intravenous
LSMeans	Least Square Means

MCP-Mod	Multiple Comparison Procedure Modelling
MedDRA	Medical Dictionary for Regulatory Activities
OI	Osteogenesis Imperfecta
P1CP	Carboxy-Terminal Propeptide of type 1 Procollagen
P1NP	Procollagen 1 N-terminal propeptide
PD	Pharmacodynamic
PK	Pharmacokinetic
PP	Per-Protocol
PRO	Patient Reported Outcome
PT	Preferred Term
QoL	Quality of Life
QTc	Corrected QT interval of the electrocardiogram
RR	Respiratory Rate
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SE	Standard Error
SF-12	Short Form 12 Health Survey
SI	Standard International
SOC	System Organ Class
TbN	Trabecular number
TbTh	Trabecular thickness
TEAE	Treatment-Emergent Adverse Event
TESAE	Treatment-Emergent Serious Adverse Event
TFLs	Tables, Figures and Listings
Tr. vBMD	Trabecular volumetric bone mineral density
VAS	Visual analogue scale
WBC	White Blood Cell
WHO	World Health Organisation

1 INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to provide detailed descriptions of the statistical methods, data derivations and data displays for study protocol MBPS205 Amendment 5 "A Phase 2b, Multicentre, Multinational, Double-blind, Dose-finding Study, incorporating an open label substudy, in Adult Patients with Type I, III or IV Osteogenesis Imperfecta Treated with setrusumab (BPS804)" dated 12 December 2018.The table of contents and templates for the Tables, Figures and Listings (TFLs) will be produced in a separate document.

Any deviations from this SAP will be described and justified in the Clinical Study Report (CSR).

The preparation of this SAP has been based on International Conference on Harmonisation (ICH) E9[1].

All data analyses and generation of TFLs will be performed using SAS 9.4® or higher.

Note: This plan does not address the validation analyses to the HRpQCT measurements or the pharmacokinetic (PK) analyses. These analyses will be described in a separate analysis plan.

2 STUDY OBJECTIVES

The study is designed to investigate the dose-response to setrusumab in a large cohort of Type I, III and IV OI patients.

2.1 Primary objective

To demonstrate that setrusumab increases radial trabecular volumetric bone mineral density (Tr.vBMD) on high resolution peripheral quantitative computed tomography (HRpQCT) and bone strength on finite element analysis (FEA) in patients with OI Type I, III or IV.

2.2 Secondary objectives

The secondary objectives are:

1. To determine the dose-response relationship of Tr. vBMD to setrusumab after 12 months of treatment
2. To evaluate the onset of treatment effect
3. To evaluate the effect of setrusumab on fracture rate (peripheral, major [long-bone], and vertebral fractures)
4. To evaluate the effect of setrusumab on vertebral fractures and vertebral height
5. To evaluate the effect of setrusumab on bone mineral density (BMD)
6. To evaluate the effect of setrusumab on bone quality
7. To evaluate changes in radial Tr. vBMD on HRpQCT and bone strength FEA during the 12 months post setrusumab treatment period
8. To evaluate the effect of setrusumab on HRpQCT parameters
9. To evaluate the effect of setrusumab on body composition
10. To evaluate the effect of setrusumab on markers of bone composition
11. To evaluate the effect of setrusumab on Patient-Reported Outcomes (PROs) and Quality of Life (QoL)
12. To evaluate the pharmacokinetics (PK) of setrusumab
13. To evaluate potential induction of antidrug antibodies (ADAs) by setrusumab and their effect on safety and PK
14. To evaluate safety and tolerability of setrusumab

2.3 Exploratory objectives

The exploratory objectives are:

1. To evaluate the effect of setrusumab on bone quality
2. To evaluate the effect of setrusumab on changes in use of physical aids
3. To evaluate the effect of setrusumab on hearing

4. To evaluate the effect of setrusumab on bone formation and resorption parameters measured by HRpQCT

3 STUDY DESIGN

3.1 General study design

MBPS205 is a phase 2, multicentre, multinational, double-blind, dose range-finding study comparing 3 blinded doses of setrusumab (20 mg/kg, 8 mg/kg and 2 mg/kg) and an open-label treatment arm (20 mg/kg).

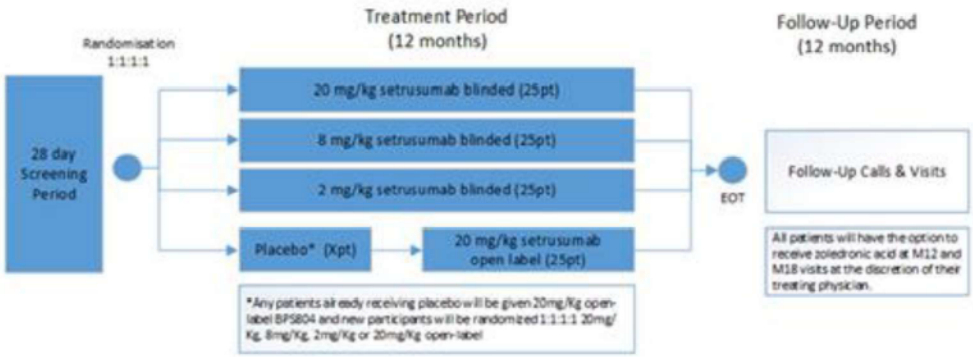
Following a screening period lasting up to 28 days, eligible participants will be randomised 1:1:1:1 into the 4 treatment arms stratified by OI type (Type I or III/IV) and will receive study treatment (setrusumab) for 12 months. Any participants assigned to placebo according to the previous MBPS205 protocol design will be transferred to open-label 20 mg/kg treatment for 12 months.

Participants will have HRpQCT scans of their non-dominant distal radius at baseline, 3-month (for open-label participants only), 6-month and 12-month time points to determine the effect of treatment on Tr. vBMD. The primary analysis will be conducted once all data up to month 12 has been collected for the blinded treatment groups.

Following completion of the study treatment participants can receive an optional single dose of zoledronic acid and will enter 12 months follow-up period and will have visits at Months 14, 18 and 24 to establish the rate of offset of treatment effect. Participants will also receive follow-up telephone calls at Months 16 and 21. Participants can receive an optional further dose of zoledronic acid at Month 18 at the discretion of their treating physician. During the follow-up period participants will have HRpQCT scans of their non-dominant distal radius and tibia at Month 18 and Month 24.

The Study Flow Chart is presented in Figure 1.

Figure 1: Study Flow Chart



An Independent Data Monitoring Committee will monitor safety throughout the study and make recommendations on the selected dose.

3.2 Randomisation and blinding

3.2.1 Randomisation

All participants will be centrally randomised using an Interactive Web Response System (IWRS) in a 1:1:1:1 ratio to receive either 20mg/kg, 8mg/kg, 2mg/kg or open-label 20 mg/kg stratified by OI type (Type I or III/IV). Subject to informed consent being given by the participant, all participants assigned to placebo prior to protocol amendment 4.0 will be reassigned to begin 12 months of open-label 20 mg/kg setrusumab treatment. ICON Biostatistics will produce a randomisation list that links participant numbers to randomisation numbers. These randomisation numbers will be linked to different treatment regimens and the randomisation scheme will be stratified according to the Sillence Clinical Classification of OI type into 2 groups (Type I or III/IV).

Following confirmation from the investigator, the pharmacist will contact the IWRS and be issued with a treatment regimen. Should the participant be allocated to a setrusumab regimen, then the pharmacist will input the patient's weight into the IWRS and receive a calculated dose per participant.

3.2.2 Blinding

Study Blinding is conducted to ensure that treatment assignment is unbiased and concealed from patients and investigator staff. With the exception of the open-label treatment arm, investigators and participants will remain blinded to each participant's assigned study treatment throughout the course of the study. In order to maintain this blind, an otherwise uninvolved 3rd party (pharmacist) will be responsible for the reconstitution and dispensation of all study treatment. A separate unblinded Clinical Research Associate (CRA) will monitor the pharmacy.

3.2.3 Unblinding

Upon completion of the treatment period (final Month 12 or EOT data collected for the final patient) the blind will be broken and the primary efficacy analysis completed. Analyses for the follow-up period will be carried out separately to the primary analysis once all patients have completed the 12 months follow-up period, or have withdrawn from it.

In the event of a Quality Assurance audit, the auditor(s) will be allowed access to unblinded study treatment records at the site(s) to verify that randomisation/dispensing has been performed accurately.

The IWRS will be programmed with blind-breaking instructions. The study blind may be broken if, in the opinion of the investigator, it is in the participant's best interest to know the study treatment assignment. Mereo must be notified before the blind is broken unless identification of the study treatment is required for a medical emergency in which the knowledge of the specific blinded study treatment will affect the immediate management of the participant's condition (eg, antidote is available). In this case, Mereo must be notified within 24 hours after breaking the blind. The date

and reason that the blind was broken must be recorded in the source documentation and electronic case report form (eCRF), as applicable.

3.3 Study treatments and assessments

During the treatment period study participants will receive 2 mg/kg, 8 mg/kg or 20 mg/kg setrusumab IV, administered monthly for 12 months via 60 minute infusions of 250mL 5% dextrose with setrusumab. Patients randomised prior to Protocol Amendment 4.0 may have received a matching placebo during the treatment period prior to transitioning to open label 20 mg/kg.

Dose changes are not permitted during the treatment period.

The maximum study duration from randomisation to end of the safety follow-up period is 24 months.

A detailed description of procedures and assessments to be conducted during this study can be found in the protocol.

4 STUDY ENDPOINTS

4.1 Primary efficacy endpoint

The primary efficacy endpoint of this study is the change from baseline in Tr. vBMD (radius) on HRpQCT and change from baseline in bone strength on FEA at 12 months.

4.2 Secondary endpoints

The secondary efficacy endpoints of this study are:

1. Tr. vBMD (radius) on HRpQCT at 12 months.
2. Tr. vBMD (tibia & radius) on HRpQCT and bone strength on FEA at 6 months
3. Tr. vBMD (tibia & radius) on HRpQCT and bone strength on FEA at 3 & 6 months in the open-label treatment arm
4. Total fracture rate at 12 months
5. Peripheral fracture rate at 12 months
6. Vertebral fracture rate at 12 months
7. Long-bone fracture rate at 12 months
8. Changes in vertebral fractures and vertebral height with Genant's semi-quantitative method and 6-point quantitative morphometry from baseline at 6 and 12 months
9. Changes in lumbar, whole body, and proximal femur DXA BMD (dual-energy x-ray absorptiometry bone mineral density) (absolute and T-score) from baseline at Month 6 and Month 12
10. Changes in bone histomorphometry
11. Tr. vBMD (tibia & radius) on HRpQCT and bone strength on FEA at 18 and 24 months
12. Changes (tibial and radial) in Total vBMD, cortical vBMD, bone volume fraction (BV/TV), peripheral to medullary trabecular bone density ratio (Met/Inn), trabecular thickness (TbTh), trabecular number (TbN), inhomogeneity, cortical thickness, and cortical porosity from baseline at 6, 12, 18, and 24 months
13. Changes in Tr. vBMD (tibia) from baseline at 12, 18, and 24 months
14. Changes (tibial and radial) in Total vBMD, cortical vBMD, bone volume fraction (BV/TV), peripheral to medullary trabecular bone density ratio (Met/Inn), trabecular thickness (TbTh), trabecular number (TbN), Inhomogeneity, cortical thickness, and cortical porosity on HRpQCT and bone strength on FEA at 3 months in the open-label treatment arm.
15. Changes in body height, weight and body mass index (BMI) from baseline at 6 and 12 months
16. Changes in lean and fat body mass from whole body DXA

17. Changes from baseline at each visit in bone turnover markers and metabolic biomarkers associated with bone including:
    - Parathyroid hormone [PTH]
    - Aminoterminal propeptide of type 1 procollagen [P1NP]
    - Carboxy-terminal propeptide of type 1 procollagen [P1CP]
    - Osteocalcin [OC],
    - Bone-specific alkaline phosphatase [BSAP]
    - Carboxy-terminal telo-peptide [CTX-1]
    - Amino-terminal telo-peptide [NTX-1]
    - Receptor activator of nuclear factor kappa-B ligand [RANKL]
    - Osteoprotegerin
    - Transforming growth factor beta [TGF-β]
    - Sclerostin
    - Released C-terminal pro-peptide of Type V collagen [Pro-C5]
    - Neo-epitope of MMP-2
    - 9 mediated degradation of Type V collagen [C5M])
18. Change in total scores from baseline to Month 6 and 12 on Short Form 12 Health Survey (SF-12), EuroQol 5-dimension 5-level descriptive system (EQ-5D-5L) and Osteogenesis Imperfecta specific Quality of Life Questionnaire for Adults (OIQoLA)
19. Change in OIQoL-A pain and activity subscale scores from baseline to Month 6 and 12
20. Serum concentrations of setrusumab
21. Serum concentrations of anti-setrusumab antibodies
22. Serum concentrations of setrusumab neutralising antibodies
23. Safety and Tolerability endpoints including:
    - Treatment-emergent adverse events (TEAEs)/Treatment-emergent serious adverse events (TESAEs)
    - Infusion site reactions
    - Vital signs
    - Physical examinations
    - ECG
    - Clinical laboratory tests

4.3 Exploratory endpoints

The exploratory endpoints of this study are:

1. Changes in Trabecular Bone Score (TBS) from baseline at 6 and 12 months
2. Changes in usage or need for physical aids from baseline at 12 months
3. Change in auditory function from baseline at 12 months in a subset of patients
4. Change in bone formation from baseline at 6 and 12 months, including:

- Normalised newly mineralised bone volume (MV/BV)
- Normalised newly mineralised surface area (MS/BS)
- Mineralised thickness (MTh)
- Normalised formation patch number density (N.F. Patch/BV)
- Formation patch volume (F.PatchVol.)
- 3D bone formation rate (3D BFR/BS)
- 3D mineral apposition rate (3D MAR)

5. Change in bone resorption from baseline at 6 and 12 months, including:
   - Normalised newly eroded bone volume (EV/BV)
   - Normalised newly eroded surface area (ES/BS)
   - Erosion depth (ED)
   - Normalised resorption cavity number density (N.R.Cav./BV)
   - Resorption cavity volume (R.Cav.Vol.)
   - 3D bone resorption rate (3D BRR/BS)
   - 3D mineral resorption rates (3D MRR)

5 SAMPLE SIZE AND POWER

The number of participants is based on power considerations for the trabecular vBMD (radius) derived from HRpQCT at Month 12. The primary analysis method for this endpoint will be on the change from baseline within each treatment group, so the primary hypothesis to be tested is

$$H0: \mu T \leq 0 \text{ vs. } H1: \mu T > 0$$

where μT denotes the change from baseline for each dose of Setrusumab. Under the assumptions of a maximum change from baseline of [redacted], a standard deviation at both timepoints of [redacted] and a correlation between both measurements of [redacted] a sample size of 25 patients per group yields approximately 80% power for a one-sided test with significance level 0.025, if the data are analysed on the log scale using a t-test.

For a change from baseline of [redacted], with standard deviation as above the one-sided test would have a probability of 80% to yield a p-value below 5%, being considered indicative of a trend.

Following a review of the trabecular vBMD (radius) derived from HRpQCT at Month 6 of those patients within the 20mg/kg Setrusumab open label treatment group, a change from baseline standard deviation of [redacted] was observed. This yields to a correlation between baseline and month 6 visits of [redacted] (much higher than the [redacted] that was originally assumed). If such a correlation is observed at month 12, given 25 subjcts per arm, the one-sided test would have a probability of 80% to yield a p-value below 5% for a change in baseline of [redacted].

As a secondary analysis, the MCPMod approach will be applied to the data in order to assess a non-flat dose-response curve and to obtain information about the underlying dose-response curve. Due to the missing placebo group, the power of this approach depends heavily on the underlying dose-response curve.

The following candidate models will be used:

- EMax (ED50 of 1.5 mg)
- Exponential (δ = 2)

Under the assumptions of each of the two candidate models being the true dose-response curve, and in addition an Sigmoid EMax model with parameters ED50 = 10 mg and h = 2.4, a contrast test at a one-sided significance level of 2.5% has the following power for each of the three models:

Model	EMax	Exponential	Sigmoid EMax
Power	13.7%	73.2%	60.7%

The total sample size required is therefore determined as 100:

- 25 participants on open-label setrusumab (20 mg/kg)
- 25 participants on each of the active arms (2 mg/kg, 8 mg/kg, 20 mg/kg)

Approximately 100 participants will be randomised.

6 ANALYSIS POPULATIONS

Analyses for the primary efficacy endpoints and selected secondary endpoints will be analysed on the FAS, the Non-restricted FAS and the PP samples, while analyses for the remaining secondary efficacy endpoints will be analysed on the FAS sample only. Data collected for patients while taking placebo prior to protocol amendment 4.0 will not be included in any efficacy analysis (i.e treatment comparisons), however will be included in efficacy descriptive statistic summaries.

Safety and tolerability endpoints will be analysed using the safety population. Data collected for patients while taking placebo prior to protocol amendment 4.0 will be included in all safety summaries.

Pharmacokinetic data will be analysed using the PK population.

6.1 Full Analysis Set (FAS)

The Full Analysis set consists of all participants who are randomised to one of the blinded treatment arms and take at least 1 dose of study treatment. This does not include patients randomised to placebo prior to protocol amendment 4.0.

Participants will be analysed according to the randomised treatment.

6.2 Modified Full Analysis Set (mFAS)

The Modified Full Analysis Set consists of all participants from the FAS who have sufficient cross-sectional overlap across HRpQCT scans. Subjects must have at least 70% overlap across all performed scans (either Radius or Tibia).

For all analysis utilising the mFAS population only scans with at least 70% overlap will be included for each associated location (Radius or Tibia).

Participants will be analysed according to the randomised treatment.

6.3 Non-restricted FAS

The Non-restricted FAS consists of all participants who are randomised and take at least 1 dose of study treatment. This includes data for patients taking placebo prior to protocol amendment 4.0 and those randomised to open label 20mg/kg. Data from the two 20 mg/kg dose group maybe pooled for the follow-up analysis.

Participants will be analysed according to the randomised treatment.

6.4 Per-Protocol population (PP)

The Per-protocol population consists of all participants from the FAS who have been treated according to the protocol and fulfill the following criteria (to be further described in the classification meeting plan):

1) Specific inclusion/exclusion criteria satisfied
2) Absence of relevant protocol violations with respect to factors likely to affect the efficacy of treatment where the nature of protocol violation will be defined before breaking the blind
3) Adequate study medication compliance which will be determined before breaking the blind.

Participants will be analysed according to the treatment to which they were randomised.

6.5 Safety population (Safety)

The Safety Population consists of all participants who are administered at least 1 dose of study treatment. This includes data for patients taking placebo prior to protocol amendment 4.0 and those randomised to open label 20mg/kg. Data from the two 20 mg/kg dose group maybe pooled for the follow-up analysis.

Participants will be analysed according to the treatment they actually received.

6.6 PK population

The PK population consists of all participants in the Safety Population who have at least 1 quantifiable setrusumab serum concentration.

Patients will be analysed according to their randomised treatment.

6.7 Other Populations Defined for Tables and Listings

For the purposes of tables and listings a further two populations are defined:

- All enrolled patients (i.e patients who have signed informed consent)
- Randomised population (all randomised patients).

6.8 Protocol deviations/violations and exclusions from analysis sets

All violations and exclusions of subjects from analysis populations will be identified and documented at the Classification Meeting prior to study unblinding. The review of each subject's data will be conducted using (but not limited to) the following sources of information:

- Supportive subject listings, provided by the ICON ahead of the Classification Meeting, based on data recorded on the eCRF.
- Protocol Deviation Logs, retrieved from Clinical Trial Management System (CTMS).

7 STATISTICAL CONSIDERATIONS AND ANALYSIS

7.1 Derived Variables

7.1.1 General Variables

Study Day

The derived study day will be relative to the date of first active (i.e non-placebo) study medication (Derived Study Day 1) and calculated as:

- (assessment date – date of first dose of study medication) + 1, for assessments on or after the first dose of study medication
- (assessment date – date of first dose of study medication), for assessments prior to date of first dose of study medication.

7.1.2 Definitions relative to demographic and other baseline characteristics

Age

Age at randomisation will be calculated as:
Age (years) = (date of randomisation - date of birth + 1) / 365.25

Weight, Height and BMI

Weight, recorded in pounds on the eCRF, will be converted in kilograms (1 pound = 0.45359 kg). Height, recorded in inches on the eCRF, will be converted in centimetres (1 inch = 2.54 cm) (International System of Units).

Body mass index (BMI) will be calculated in kg/m^2 as: weight (kg)/(height (m))^2.

Temperature

Temperature, recorded in Fahrenheit degrees on the eCRF, will be converted in Celsius degrees:
Celsius degrees = (Fahrenheit degrees - 32) x (5/9)

Time since last Fracture

Time since last Fracture in months will be calculated as:

(Date of first dose of study medication – date of last fracture + 1) / 30.4375

7.1.3 Definitions relative to efficacy parameters

7.1.3.1 HRpQCT Assessments

HRpQCT scans are overlapped with all previous visit scans within a patient so that a common area is assessed across all visit timepoints. As such for all summaries and analyses only the assessments derived from the common area will be used. Such assessments can be identified where the "analysis timepoint" is populated with the patient's final visit timepoint. For example, for a patient that has there final scan at month 12, all assessments where analysis timepoint is populated with "Month 12" will be used.

7.1.3.2 Vertebral fracture assessments

The lateral spine radiographs will be assessed for vertebral fractures using the Genant Semiquantitative Scoring Method assigning each vertebra a grade between 0 and 3 (SQ score):

- Grade 0: normal
- Grade 1: mild fracture
- Grade 2: moderate fracture
- Grade 3: severe fracture

A new vertebral fracture is defined as an SQ score >=1 at a post baseline visit where the baseline SQ score = 0 for the associated vertebrae. A subject is defined as having a new fracture at a visit if they have at least 1 vertebra with a new fracture.

A worsening vertebral fracture is defined as an increase in SQ score >=1 from the previous visit (where the baseline SQ score >0) for the associated vertebrae. A subject is defined as having a worsening fracture at a visit if they have at least 1 vertebra with a worsening fracture.

The sum of all grades at a visit is defined as sum of all SQ scores across all vertebrae assessed at a visit by the Genant Semiquantitative Scoring Method.

The below vertebral height parameters (for each vertebra) will be derived from the 6-point vertebral morphometry assessments:

- Anterior /Posterior height ratio (HA/HP)
- Mid / Posterior height ratio (HM/HP)

7.1.3.3 Time to first fracture

The time to first fracture for each fracture category (Total, Peripheral, Vertebral, Non-vertebral and Long-bone) is defined as the time (in days) from the date of first dose of active study medication until the date of date of diagnosis (defined as first occurrence of symptoms of fracture) of each specific fracture type.

Patients who do not have a facture will be censored at the date of completion/discontinuation of the treatment period for each category.

7.1.3.4 Short Form 12 Health Survey (SF-12)

SF-12

The SF-12 is a generic, 12-item survey that measures 8 domains of health: physical functioning, role limitations due to physical health, bodily pain, general health perceptions, vitality, social functioning, role limitations due to emotional problems, and mental health. It yields scale scores for each of these 8 domains and 2 summary measures of physical and mental health: The Physical Component Summary and the Mental Component Summary. Please see Appendix A **Error! Reference source not found.**for detailed scoring of each domain and component scores.

7.1.3.5 5-level descriptive system (EQ-5D-5L)

EQ-5D-5L

The EQ-5D-5L is a standardised measure of health status comprised of a descriptive system of 5 health-related quality of life states (i.e., mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and a Visual Analogue Scale (VAS) of overall health.

Each dimension is rated on a 5-point response scale indicating severity of problems, where 1 is "no problems" and 5 is "extreme problems". The 5 questions are scored and together contribute to the EQ-5D index (utility) score between 0 and 1 (1 being perfect health), which will be calculated using the developers' algorithm based on country-specific reference score sets. Please see Appendix B for detailed scoring of the EQ-5D index score.

The EQ-5D VAS is a measure of overall self-rated health status, used and analysed separately from the index score. The VAS ranges from 0 to 100, with higher scores indicative of better overall health.

7.1.3.6 Osteogenesis Imperfecta specific Quality of Life Questionnaire for Adults (OIQoL-A)

OIQoL-A

The OIQoL-A is a PRO instrument for use in adults with OI Type I, III, or IV. The OIQoL-A currently measures 5 areas of quality of life related to OI (Physical Function, Pain, Hearing Loss, taking care/Concerns, Social and Family Life and Activities). For each visit the OIQOL-A total score and pain and activity subscale scores will be derived. Please see Appendix C for detailed scoring of the total and subscale scores.

7.1.4 Definitions relative to safety parameters

7.1.4.1 Adverse Event (AE)

Treatment Emergent Adverse Event (TEAE)

A Treatment Emergent AE (TEAE) is defined as an AE occurring or worsening on or after the first dose of study medication. AEs that occurred for subjects while taking placebo prior to protocol amendment 4.0 will not be considered treatment emergent for the active treatment groups. TEAEs will also be split between AEs that started during the Patient's treatment period and those that start on or after the first day of the follow-up period.

Duration of AEs

The duration of an AE will be calculated as the resolution date minus the start date plus 1.

7.1.4.2 Treatment Compliance

Duration of Exposure

The duration of Exposure to study medication (number of days) will be defined as:

(date of last dose of study medication – date of first dose of study medication) + 28 days

7.1.4.3 Prior, Concomitant and Follow-up Medications/Procedures

Medications and Procedures will be assigned as being prior to study treatment, concomitant with study treatment or taken during the follow-up phase based on the start and stop dates of the medication and dosing dates.

If the medication/procedure stop date is before the date of the first dose of study medication, the medication/procedure will be assigned as being prior to study treatment. Otherwise, the medication/procedure will be assigned as being concomitant with study treatment unless the start date of the medication/procedure is after the date of entry into the follow-up period, when it will then be classified as occurring in the follow-up phase.

Medications/procedures that occurred for patients while taking placebo (prior to protocol amendment 4.0) will be considered as being prior to study medication as well as concomitant for placebo.

7.1.4.4 Potential Hy's Law Criteria

Patients will be deemed as meeting potential Hy's law Criteria if their ALT or AST > 3x ULN and bilirubin > 2 x ULN at the same visit.

Further assessment to determine that no other reason can be found to explain the combination of increased aminotransferase and serum total bilirubin, such as viral hepatitis, alcohol abuse,

ischemia, pre-existing liver disease, or another drug capable of causing the observed injury will be conducted.

7.1.4.5 Duration of Hospitalisation

The duration of hospitalization (number of days) will be defined as:

(date of discharge – date of admission) + 1

7.2 Handling of missing data and outliers

7.2.1 Missing data analysis methods

Unless specified, data in summary tables will be presented using Observed Case (OC) data and therefore no missing data will be imputed.

For data listings, unless specified, all data will be presented as they have been recorded (e.g. missing and partial dates will not be replaced).

7.2.2 Handling of missing or incomplete data

7.2.2.1 Partial dates of last fracture

In order to calculate the time since last fracture, the following rules will be applied for partial dates:

- if the day of the month is missing it is imputed to be the 15th
- if both the day and month are missing, they are imputed to be June 30th
- missing years will be left as missing.

The above will be flagged and described as a footnote in the appropriate listings.

7.2.2.2 Missing items within PRO data

The method of handling missing data for each questionnaire will be based upon the author's recommendation. Details can be found in Section 7.1.3.

7.2.2.3 Concomitant medications definitions and handling of missing or incomplete dates

Should the start date for a medication/procedure be missing or incomplete to the extent that it could be before or after the time of start of study medication, then it will be assumed that the medication/procedure began after the start of study medication (i.e. reported as concomitant medication/procedure). Similarly, if it is not clear whether the medication/procedure start date was on or before the date of entry into the follow-up period, then it will be assumed the medication/procedure began on or before the entry date (i.e. reported as a concomitant medication/procedure) (worst case approach).

For listings of medications, dates will be presented as they have been recorded (missing and partial dates will not be replaced).

7.2.2.4 Definition of treatment-emergent AEs and handling of missing or incomplete dates

In the event of an incomplete onset date, the event will be considered to be treatment-emergent unless the partial onset date information or complete or partial end date confirms onset or end prior to the first dose of study medication date.

For AEs listings, dates will be presented as they have been recorded (missing and partial dates will not be replaced).

8 STATISTICAL METHODS

8.1 General statistical conventions

All statistical procedures will be completed using SAS version 9.4 or higher and ADDPLAN DF 4.0 or later.

8.1.1 Populations for analysis

Demographic and baseline characteristics will be summarised by the Safety population, unless otherwise stated.

Analyses of the primary efficacy endpoint and selected secondary endpoints will be performed on the FAS population and also on the mFAS, PP population and Non-restricted FAS as a sensitivity analysis.

The remaining secondary and exploratory efficacy endpoints will be analysed using the FAS and Non-restricted FAS only.

All safety and tolerability variables will be analysed using the Safety population.

PK data will be analysed using the PK population.

8.1.2 Treatment groups

Statistics will be displayed for the following treatment groups:

- Setrusumab 20 mg/Kg
- Setrusumab 8 mg/Kg
- Setrusumab 2 mg/Kg
- Open Label Setrusumab 20 mg/Kg
- Pooled Setrusumub 20 mg/Kg (where applicable)
- Placebo
- Total: overall (where applicable).

Pooled Setrusumub 20 mg/Kg may only be used for the follow-up analysis.

8.1.3 Descriptive statistics

Continuous variables will be summarised using descriptive statistics including number of non-missing observations (n), means (Mean), standard deviations (SD), minimum (Min), median, and maximum (Max). One additional decimal point for mean and median and 2 additional decimal points for SD will be used.

For categorical variables, summaries will include the number of non-missing observations (n) or the number of patients in the population (N) as applicable, the counts of patients and percentages.

Percentages will be rounded to one decimal place. The number of missing values will be presented as a separate category with no percentage, but only if any missing value is recorded in the data for that summary.

Data in summary tables will be presented on an Observed Cases (OC) basis and summary statistics will only be presented at each visit for which the parameter is scheduled to be collected.

8.1.4 Statistical significance

Unless otherwise stated, all statistical testing will be two-sided and conducted at the significance (alpha) level of 0.05. Two-sided 95% confidence intervals (CIs) will be provided when relevant.

For the primary analysis the LSMeans for each blinded treatment group will be calculated and tested against the null hypothesis of 0 in an a priori hierarchical approach, starting with the highest dose.

For secondary and exploratory endpoints treatment groups will be assessed by pairwise comparisons for secondary efficacy data, and no adjustment for multiplicity will be made.

8.1.5 Definition of Baseline

Unless otherwise stated, Baseline will be the last assessment value before first study dose of study medication. For patients that were randomised to placebo prior to protocol amendment 4.0, assessments taken while on placebo may be utilized for the baseline derivation of each parameter.

For HRpQCT scan parameters baseline assessments maybe collected up to 10 days post the first dose of study medication.

Change from Baseline is defined as (value at assessment date – baseline value).

8.1.6 Visit dates

For each visit, the date recorded by the Investigator in the eCRF will be considered as the visit date in all the algorithms and the listings.

Outputs will include all data collected from screening through to the end of the follow-up period.

8.1.7 Data re-allocation

The following general rules to handle repeated assessments will be considered:

- 12-lead ECG:
    - in case of multiple measurements associated to the same timepoint (e.g. triplicate ECGs), the average value will be considered for HR, PR interval, QRS duration, QT interval and QTcF interval.

8.1.8 Data listings

All the relevant patient data, including those derived and those assessments prior to protocol

amendment 4.0 on the placebo arm, will be presented in individual patient data listings. All listings will be sorted by treatment group, investigational site, patient number, date/time and visit. The patient's sex and age will be stated on each listing. Unless otherwise stated, data listings will be based on all patients randomised.

Unscheduled visit results will be included in date/time chronological order within patient listings, but will not be tabulated.

8.2 Patient disposition

All patients who provided informed consent will be included in a summary of patient disposition. The number of patients screened, the number of screen failures, the frequency of patients randomised (by treatment group and overall), the frequency and percentage of patients in the Safety population, in the FAS population, in the mFAS population , in the Non-restricted FAS population, in the PP population and in the PK population (by treatment group and overall) will be summarised.

Patient disposition information will be summarised by treatment group and overall. The number of patients completing and withdrawing the treatment period and follow-up period will be tabulated in the study disposition table which will also include the reasons for treatment discontinuation and early withdrawal of the patient as reported on the eCRF.

Finally, the reasons for screen failure as reported on the eCRF will be tabulated.

A listing will include the randomisation/kit numbers, randomisation date/time and study disposition data (using the enrolled population).

8.3 Protocol deviations

The number of patients excluded from FAS, mFAS, Non-restricted FAS, Safety, PP and PK populations and reasons for exclusion will be summarised by treatment group and overall.

All key protocol deviations identified will be summarised by treatment group and overall.

In addition all protocol deviations will be listed based on data recorded on the eCRF and/or protocol deviation Logs (from CTMS).

8.4 Demographics and baseline characteristics

No formal comparison between treatment regimens on demographics and baseline characteristics will be conducted.

8.4.1 Demographics

Demographics including age, sex, race, ethnicity, weight, height, BMI and child-bearing potential will be listed and summarised using descriptive statistics for continuous variables and tabulated for categorical variables. Summary will be produced using the Safety Population.

8.4.2 Fracture/OI history

Fracture/ OI history including the number patients experiencing any fracture within the last 2 years (5 years from protocol amendment 4.0), time since last fracture (months), description of fracture (location), fracture anatomical location, percentage compression (vertebral only), evidence of interval healing, fracture history patient reported (Y/N), OI Type and COL1A1/COL1A2 testing (Y/N) will be summarised by treatment group and overall using the safety population.

8.4.3 Medical and Physical Aids History

A summary of Medical and Physical Aids history will be presented by medical history code, system organ class (SOC) and preferred term (PT), by treatment group and overall using Medical Dictionary for Regulatory Affairs® (MedDRA) using the Safety population.

8.4.4 Prior, Concomitant and Follow-up medications

Prior and concomitant medications will be coded using the World Health Organization (WHO) Drug Dictionary.

Prior, concomitant and follow-up medications will be summarised by Anatomical Therapeutic Chemical (ATC) classification level 4 and Preferred Name by treatment group and overall for the Safety population.

8.4.5 Concomitant Procedures/Physical Aids

All concomitant Procedures and Physical Aids will be coded using the World Health Organization (WHO) Drug Dictionary and will be presented by medication class, treatment group and overall.

8.4.6 Concomitant Therapy

The number of patients taking Vitamin D and Calcium Therapy during the treatment and follow-up periods will be summarised by treatment and overall.

In addition the number of patients taking Zoledronic Acid Therapy at Month 12, Month 18 and not at all during the follow-up period will be summarised by site, treatment and overall.

8.4.7 Extent of exposure

8.4.7.1 Treatment duration

Duration of treatment (in months) will be categorised in intervals (0-1, 1-3, 3-5, 5-7, 7-9, 9-11 >11) and summarised descriptively by treatment regimen on the Safety population.

Study drug exposure (in months) will be summarised by treatment group on the Safety population using descriptive statistics.

8.4.7.2 Treatment compliance

The number of subjects with infusion interruptions and incomplete infusions will be summarised at each month and overall.

8.5 Efficacy analyses

This section addresses separately the analyses to be conducted on the primary, secondary and exploratory efficacy variables. Analyses for the primary efficacy endpoints and selected secondary endpoints will be analysed on the FAS, mFAS, the Non-restricted FAS and the per-protocol populations, while analyses for the remaining secondary and exploratory efficacy endpoints will be analysed on the FAS population only.

All definitions relative to efficacy and exploratory endpoints are detailed in Section 7.1.3.

8.5.1 Analysis methods

For all continuous outcomes, where appropriate, an ANCOVA model analysis will be performed. Analyses will include the fixed, categorical effects of treatment and randomisation stratum as well as the associated baseline value as a covariate. Significance tests for treatment differences (between Setrusumab doses) will be based on least-squares means (LSMeans).

ANCOVA analyses dealing with nested observations within a patient (e.g., vertebral assessments) will consider this relationship in the modelling by specifying patient as a random factor.

If the input data is log-transformed then LSMeans and LSMean differences between treatment groups will be back-transformed to their original scale for descriptive purposes. After substracting the geometric mean of the log baseline values, the back-transformed LSMeans will correspond to the percentage change from baseline, while the back-transformed LSMeans differences will correspond to the ratio of percentage changes from baseline.

Time to event variables (e.g., time to first fracture) will be analysed using Kaplan-Meier methods, and quartiles for the median time to event will be presented together with their 95% confidence intervals.

Count variables (e.g number of new fractures) will be analysed using Poisson regression. Analyses will include treatment and randomisation as fixed effects and the logarithm of the duration of exposure to study medication will be included as an offset variable in the model. In case over dispersion is a concern, the negative-binomial regression model will be applied instead.

For binary outcomes the proportion of patients with a response will be summarised and compared between treatment groups using Chi squared test.

8.5.1.1 Multiplicity

All treatment groups will be assessed by pairwise comparisons. No adjustment will be made for multiple comparisons except for primary analysis.

8.5.1.2 Treatment by center interaction analysis (multi-center study)

No analysis will be made to assess the treatment-by-center interaction.

8.5.2 Analysis of primary efficacy endpoint

The primary efficacy endpoint for this study is the change in baseline in Tr. vBMD (radius) on HRpQCT and change from baseline in bone strength on FEA at 12 months. The primary endpoint and dose response analysis for the primary endpoint will be performed on the FAS population and repeated on the mFAS, Non-restricted FAS and the PP populations

8.5.2.1 Primary analysis

The HRpQCT results for the Trabecular vBMD (radius) will be log transformed and an ANCOVA model will be used to calculate the LSMeans for the change from baseline as described in Section 8.5.1. The change from baseline LSMeans for each treatment group will be tested against the null hypothesis of 0 using a t-test, in an a priori hierarchical approach, starting with the highest dose.

The LSMeans and standard errors will be displayed at each visit aswell as the corresponding P-value from each t-test.

Differences between LSMeans, their confidence intervals and P-values will also be displayed but will be analysed in an exploratory fashion only. The type I error rate will be 5% two-sided. P-values above 5% but below 10% (two-sided) will be considered as indicative for a trend.

LSMeans and LSMean differences between treatment groups will be back-transformed to their original scale for descriptive purposes. After subtracting the geometric mean of the log baseline values, the back transformed LSMeans will correspond to the percentage change from baseline, while the back-transformed LSMeans differences will correspond to the ratio of percentage changes from baseline.

Line graphs of the LSmean HRpQCT results (+/- SE) and change from baseline in HRpQCT results (+/- SE) will be presented.

A descriptive summary will also be presented for the non-log transformed HRpQCT results for the Trabecular vBMD (radius) values,change from baseline and percentage change from baseline in HRpQCT by treatment group and OI type.

For supportive evidence, the above analysis will be repeated on the mFAS, Non-restricted FAS in both versions and the PP population for HRpQCT results for the Trabecular vBMD (radius) and repeated for the bone strength on FEA for the FAS, Non-restricted FAS and PP population.

8.5.2.2 Dose Response Analysis

The change from baseline LSMeans for the HRpQCT results for the Trabecular vBMD (radius) will further be analysed using the MCPMod methodology, using the following candidate models:

- EMax (ED_{50} of 1.5 mg)
- Exponential ($\delta = 2$)
- Sigmoid EMax (ED50 of 1.5 mg and h=2.4)

Parameter estimates and predicted values for each dose will be calculated for all models. This analysis will be repeated for the bone strength on FEA for supportive evidence.

8.5.3 Analysis of secondary efficacy endpoint(s)

The analysis of selected secondary efficacy endpoints will be performed on the FAS, mFAS, Non-restricted FAS and PP populations. The remaining secondary endpoints will be analysed on the FAS Population only.

8.5.3.1 Tr. vBMD (tibia) on HRpQCT

Descriptive summary tables will be presented for HRpQCT results for the Tr. vBMD (tibia) ,change from baseline and percentage change from baseline in Tr.vBMD (tibia) on HRpQCT results by treatment group and visit.

The log transformed change from baseline in Tr.vBMD (tibia) on HRpQCT will be analysed via ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1). This analysis will be conducted on the FAS, Non-restricted FAS, mFAS and PP populations.

8.5.3.2 DXA BMD

Descriptive summary tables will be presented for DXA Lumbar, Whole Body and Proximal femur BMD values and percentage change from baseline values by treatment group and visit.

The percentage change from baseline in for DXA Lumbar, Whole Body and Proximal femur BMD will also be analysed via ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1). This analysis will be conducted on the FAS, Non-restricted FAS and PP populations. A similar analysis will be conducted on the associated DXA Lumbar, Whole Body and Proximal femur BMD change from baseline T-scores.

8.5.3.3 Vertebral Assessments

All vertebral endpoints will be analysed on the FAS Population and PP Population only.

8.5.3.3.1 Vertebral Fractures

For all vertebral assessments the X-ray scans will be used as source.

The number and proportion of patients with a new vertebral fracture and new or worsening vertebral fracture will be summarised by treatment group throughout the treatment period (at each visit). In addition, a P-value for the comparison between each group will be provided using a Chi squared test.

The total number of new vertebral fractures and new or worsening vertebral fractures per patient during the treatment period will be compared between each treatment group using a poisson regression model as described in section 8.5.1. An estimate of the ratio of new vertebral fractures and new or worsening vertebral fractures between treatment groups will be presented together with a two-sided 95% confidence interval and P-value.

For patients that have a vertebral fracture a shift table of the change in fracture grade will be presented. The number and percentage of patients with each level of change (from worst baseline grade to worst post-baseline grade) will be displayed.

Finally the total sum of vertebral fracture grades will also be analyses via ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1).

8.5.3.3.2 Vertebral Height

Vertebral height derived from 6-point quantitative morphometry will be summarized at 6 and 12 months. The mean (across all vertebrae) Anterior, Mid and Posterior heights, HA/HP ratio and HM/HP ratio will be summarised descriptively.

The maximum change from baseline in the HA/HP and the HM/HP ratio for fractured vertebrae will be analysed using ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1).

8.5.3.4 Bone Histomorphometry

Changes in bone histomorphometry parameters will be analysed using ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1) using the FAS population only.

8.5.3.5 High Resolution peripheral Quantitative Computed Tomography (HRpQCT)

Changes in baseline in the below HRpQCT parameters at each visit will be analysed using ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1) across visits during the treatment period and follow-up period:

- Total vBMD
- Cortical vBMD
- Bone volume fraction (BV/TV)

- Peripheral to medullary trabecular bone density ratio (Met/Inn)
- Trabecular thickness (TbTh)
- Trabecular number (TbN)
- Inhomogeneity
- Cortical thickness
- Cortical porosity
- Cortical area (Ct.Ar)
- Total Area (Tt.Ar)
- Trabecular separation (Tb.Sp)
- Trabecular area (Tb.Ar)

Only patients in the open-label treatment arm will have assessments at Month 3 and as such at this visit no treatment comparison will be performed.

Analyses will be performed on the FAS, mFAS and PP populations.

8.5.3.6 Body composition

Changes in baseline in the below body composition parameters at each visit will be analysed using ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1):

- Body height
- Body weight
- Body Mass Index (BMI)
- Lean body mass
- Fat body mass

8.5.3.7 Bone Turnover Markers and Metabolic Biomarkers

Changes in baseline in the below bone turnover markers and metabolic biomarkers (and their log transformed values) at each visit will be analysed using ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1):

- Parathyroid hormone [PTH]
- Aminoterminal propeptide of type 1 procollagen [P1NP]
- Carboxy-terminal propeptide of type 1 procollagen [P1CP]
- Osteocalcin [OC]

- Bone-specific alkaline phosphatase [BSAP]
- Carboxy-terminal telo-peptide [CTX-1]
- Amino-terminal telo-peptide [NTX-1]
- Receptor activator of nuclear factor kappa-B ligand [RANKL]
- Osteoprotegerin
- Transforming growth factor beta [TGF-β]
- Sclerostin
- Released C-terminal pro-peptide of Type V collagen [Pro-C5]
- neo-epitope of MMP-2
- 9 mediated degradation of Type V collagen [C5M])

8.5.3.8 PRO's and QoL

All PRO and QoL analyses below will be performed on the FAS and Non-restricted FAS Populations.

8.5.3.8.1 SF-12

Changes from baseline in the norm-based 8 domains scores and 2 component score of the SF-12 at each visit during the treatment period will be analysed using ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1).

8.5.3.8.2 EQ-5D-5L

Changes from baseline in the EQ-5D-5L index score and VAS score at each visit during the treatment period will be analysed using ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1).

8.5.3.8.3 OIQoL-A

Changes from baseline in the OIQoL-A total score, pain subscale score and activity subscale score at each visit during the treatment period will be analysed using ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1).

8.5.3.9 Fractures

All fracture analyses below will be performed on the FAS, Non-restricted FAS and PP populations utilising the X-day data.

The number and proportion of patients with at least one new Peripheral, Vertebral, Non-Vertebral, Long-bone or any fracture will be summarised by treatment group and OI type at Month 12 and presented as the annualised fracture rate. In addition, each quarters fracture rate and the cumulative fracture rate by month will be presented. For each rate, a P-value for the comparison between each group will be provided using a Chi squared test.

Graphs of the cumulative fracture rate by month will be presented.

The time to first fracture will be analysed for each fracture category using Kaplan-Meier methods as per Section 8.5.1. As per Section 7.1.3 patients who do not have a facture will be censored at the date of completion/discontinuation of the treatment period for each category.

Quartiles for the median time to event will be presented together with their 95% confidence intervals. A Kaplan-Meier plot will also be presented. Comparisons between treatment regimens will be performed by means of the log-rank test.

8.5.4 Analysis of exploratory endpoint(s)

All exploratory analyses below will be performed on the FAS only.

8.5.4.1 Trabecular Bone Score (TBS)

Changes from baseline in the TBS score at each visit during the treatment period will be analysed using ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1).

8.5.4.2 Auditory Function

The audiometry score (Undetectable hearing loss, mild, moderate or severe) at each visit will be summarised descriptively by ear and treatment group.

8.5.4.3 Bone formation

Changes in baseline in the below bone formation parameters at each visit will be analysed using ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1):

- Normalised newly mineralised bone volume (MV/BV)
- Normalised newly mineralised surface area (MS/BS)
- Mineralised thickness (MTh)
- Normalised formation patch number density (N.F. Patch/BV)
- Formation patch volume (F.PatchVol.)
- 3D bone formation rate (3D BFR/BS)
- 3D mineral apposition rate (3D MAR)

8.5.4.4 Bone Resorption

Changes in baseline in the below bone resorption parameters at each visit will be analysed using ANCOVA in the same manner as the change from baseline in Tr.vBMD (radius) on HRpQCT (see Section 8.5.2.1):

- Normalised newly eroded bone volume (EV/BV)
- Normalised newly eroded surface area (ES/BS)
- Erosion depth (ED)

- Normalised resorption cavity number density (N.R.Cav./BV)
- Resorption cavity volume (R.Cav.Vol.)
- 3D bone resorption rate (3D BRR/BS)
- 3D mineral apposition rates (3D MRR)

8.5.4.5 Follow-up Period Subgroup Analysis

A subgroup analysis (descriptive statistics and associated ANCOVA analyses) will be performed for HRpQCT, DXA, vertebral radiograph, CTX-1, and P1NP parameters at all follow-up visits by the administration of zoledronic acid during the follow-up period (none; at Month 12; at Month 18).

8.6 Safety analyses

All definitions relative to safety endpoints are detailed in Section 7.1.4.

All the safety analyses will be based on the Safety population and will be performed for all safety variables specified below.

No statistical testing will be performed.

8.6.1 Adverse events

All AEs will be classified by SOC and PT according to the MedDRA Version 21.1 or higher.

Details for imputing missing or partial start dates of adverse events are described in Section **Error! Reference source not found.**.

All AE summaries will be generated by treatment period, follow-up period and overall (i.e including those that started in both the treatment period and follow-up period).

Notes:

- Two AEs with the same PT and classified in the same category (pre-treatment AE or TEAE) will be considered as two different events when calculating the "number of events" in the tables.
- Where a patient has the same AE, based on preferred terminology, reported multiple times in the same category (pre-treatment AE or TEAE), the patient will only be counted once at the preferred terminology level in AE frequency tables.
- Where a patient has multiple AEs within the same SOC in the same category (pre-treatment AE or TEAE), the patient will only be counted once at the SOC level in AE frequency tables.
- AEs where the intensity is missing will be assumed to be "Severe"
- AEs where the causality is missing will be assumed to have "Reasonable possibility of relatedness

An overall summary of AEs will be provided. The total number of events and number and proportion of patients experiencing any AEs, TEAEs, AEs related to study medication, SAEs, SAEs related to study medication, severe AEs, AEs leading to permanent discontinuation of study treatment and AEs leading to death will be tabulated for each treatment group and overall.

Additionally, TEAEs, related TEAEs, serious TEAEs, related SAEs, severe TEAEs, AEs leading to permanent discontinuation of study treatment and AEs leading to death will be summarised by SOC and PT for each treatment group and overall (number and percentage of patients experiencing at least one AE per PT as well as the number of observed events per PT). All TEAEs will also be summarised separately by maximum severity grade for each SOC and PT and by relationship to study medication for each SOC and PT.

The frequency and incidence of TEAEs will be presented by system organ class and preferred term for each treatment group (number and percentage of participants experiencing at least 1 AE per preferred term as well as the number of observed events per preferred term).

Infusion site reactions will be tabulated and summarised by treatment group and overall.

All AE tables will be presented on both the pooled and the non-pooled version of the Safety population.

All AEs will be presented in full in a comprehensive listing including patient number, treatment regimen, intensity, seriousness, actions taken, outcome, causality, onset/stop and duration. Details of all TEAEs of special interest, SAEs, AEs leading to permanent discontinuation of study treatment and AEs leading to death will be listed separately.

8.6.1.1 Adverse Events of Special Interest

The AESIs for setrusumab are:

- Fractures
- Cardiovascular events

The frequency and incidence of all treatment emergent AESIs will be presented by system organ class and preferred term for each treatment group.

8.6.2 Clinical laboratory evaluations

Haematology, blood chemistry and urinalysis assessment will be conducted at Screening, Baseline, Months 1, 2, 3, 6, 9 and 12/EOT during the treatment period and months 14,18 and 24 during the follow-up period.

Clinical laboratory data will be reported in Standard International (SI) units. Descriptive statistics will be presented for quantitative laboratory parameters for each treatment group and time-point. Similarly, changes from Baseline will be summarised.

Values outside the normal range will be categorised as H (above the normal range) or L (below the normal range) based on the laboratory's reference range and these will be flagged in the individual data listings.

Line graphs of mean calcium values (+/- SE) and change from baseline in calcium (+/- SE) will be presented.

Serum pregnancy test results and Follicle-stimulating hormone and estradiol assessments recorded at screening will be listed only.

8.6.2.1 Liver Function

The number and proportion of patients with laboratory values for liver function and enzymes of clinical concern as follows will be presented descriptively by treatment group:

Parameter	Level for clinical concern
ALT	> 1.5 times the upper limit of the normal range (ULN) > 2xULN > 3xULN > 5xULN > 10xULN
AST	> 1.5xULN > 2xULN > 3xULN > 5xULN > 10xULN
Bilirubin Total	> 2xULN > 3xULN
ALT/AST and Bilirubin Total	ALT/AST> 3xULN and Bilirubin Total> 2xULN (assessments to occur at the same visit)

8.6.3 Vital signs

Vital signs will be performed at each planned visit during the treatment period and at months 14, 18 and 24 during the follow-up period.

Descriptive statistics (observed values and changes from baseline) will be presented for each treatment group and time-point for vital sign measurements (body temperature, respiratory rate (RR), pulse rate, systolic blood pressure, diastolic blood pressure, weight and BMI).

A data listing of vital signs for patients will be provided.

8.6.4 Physical examinations

All physical examination data and abnormalities will be listed by patient and body system.

8.6.5 Electrocardiograms

Single 12-lead ECG will be obtained at Screening, Months 6 and 12/EOT during the treatment period and Month 14 during the follow-up using an ECG machine that automatically calculates the heart rate (HR) and measures PR, QRS, QT and QTc intervals.

Descriptive statistics will be presented for the 12-lead ECG measurements for each treatment group and time-point. Similarly, changes from baseline will be summarised.

In addition, the overall ECG interpretation will be summarised by presenting the number and percentage of patients with "Normal", "Abnormal, NCS" and "Abnormal, CS" for each treatment group and time-point.

A data listing of all ECG measurements will be provided.

8.6.6 Heath Economics

The number of patients that required hospitalisation during the treatment period and follow-up period will be summarised descriptively. In addition the duration of hospitalisation, the unit to which the patient was admitted and the reason for hospitalisation during will be summarised descriptively by treatment group.

8.7 Other analysis

8.7.1 PK

Setrusumab concentrations will be summarised on the PK population by descriptive statistics by treatment group and time-point, and including the geometric mean and coefficient of variation.

A graphical representation of the setrusumab concentrations distribution using mean plots (mean and SD) will be provided over time per setrusumab dose-level on the PK population. All 3 dose-levels will be plotted on the same graph.

8.7.2 DMC

An independent, unblinded external DMC will periodically review accumulating safety data. This will include data evaluation of accumulating unblinded safety data of Setrusumab. Full details of composition, operational aspects, and data to be reviewed and recommendation of the DMC is provided in a separate DMC charter and DMC SAP.

8.7.3 Interim Analysis

There is no formal interim analysis planned, however the primary efficacy analysis and study unblinding will be performed when all data up to Month 12 has been collected. Given the study has

been powered at Month 12 and there will not be multiple comparisons, there is no requirement to adjust the Type I error rate.

Due to hierarchical nature of the primary end-point, the complexity of the data and timelines involved, Mereo has implemented a two-stage data transfer into the study database for the primary analysis. The initial transfer will include the majority of study data analysis including a lock and transfer of the clinical database and transfer of the primary end-point Tb. vBMD, the second transfer will include finite analysis (FEA) data from the HRpQCT, audited PK data and ADA data.

An analysis and unblinding of Sponsor will be completed on the first dataset – at this point the database will be locked and no further data manipulation will be completed. The vendors completing the data analysis for the second transfer will remain blinded to patient treatment allocation by nature that this information is retained at ICON Plc– additionally the FEA is a computed model process and the PK and ADA a fully validated assay which cannot be influenced by any full or partial unblinding. The data transfers will be completed into the clinical database with no further data manipulation for the existing data.

The Clinical Study report (CSR) will be prepared based upon the results of the primary analysis.

8.7.4 Follow-up Analysis

The data collected during the off treatment follow-up will be separately analysed and published in an addendum to the CSR.

9 CHANGES TO PLANNED ANALYSIS FROM PROTOCOL

Not applicable.

10 REFERENCES

1. ICH Topic E9: Statistical Principles for Clinical Trials (CPMP/ICH/363/96 – adopted March 1998).

11 APPENDICES

Appendix A – SF12

The calculation of the SF12 V2.0 norm-based domain and component score is a 3 step process:

1) Reverse score / recalibrate item scores
2) Compute transformed raw scale domain scores (0-100)
3) Convert transformed raw scale domain scores to norm-based domain/component scores

Step 1: Reverse Score / Recalibrate Item Scores

Question 1: In General, Would you say your health is:

Response to Question 1	Scoring
Excellent	5.0
Very Good	4.4
Good	3.4
Fair	2.0
Poor	1.0

Question 2: Does you health limit you in:

a) Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling or playing golf
b) Climbing several flights of stairs

Response to Question 2a/2b	Scoring
No, not limited at all	3.0
Yes, limited a little	2.0
Yes, limite da lot	1.0

Question 3: How much of the time have you had any of the following result of your physical health:

a) Accomplished less than you would like
b) Were limited in the kind of work or other activities

Response to Question 3a/3b	Scoring
None of the time	5.0
A little of the time	4.0
Some of the time	3.0

Most of the time	2.0
All of the time	1.0

Question 4: How much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems:

a) Accomplished less than you would like
b) Were limited in the kind of work or other activities

Response to Question 4a/4b	Scoring
None of the time	5.0
A little of the time	4.0
Some of the time	3.0
Most of the time	2.0
All of the time	1.0

Question 5: How much did pain interfere with your normal work:

Response to Question 5	Scoring*
Not at all	5.0
A little bit	4.0
Moderately	3.0
Quite a bit	2.0
Extremely	1.0

* Scores are reversed from reported numeric values

Question 6: How much of the time during the past 4 weeks:

a) Have you felt calm and peaceful
b) Did you have a lot of energy
c) Have you felt downhearted and depressed

Response to Question	6a/6b Scoring*	6c Scoring
None of the time	1.0	5.0
A little of the time	2.0	4.0

Some of the time	3.0	3.0
Most of the time	4.0	2/0
All of the time	5.0	1.0

* Scores are reversed from reported numeric values

Question 7: How much of the time has your physical health or emotional problems interfered with your social activities:

Response to Question 7	**Scoring***
None of the time	5.0
A little of the time	4.0
Some of the time	3.0
Most of the time	2.0
All of the .time	1.0

Step 2: Compute transformed raw scale scores

The 8 domain raw scale scores are calculated by first summing the responses for all items in that particular domain (see below table), then converting the score to a transformed 0-100 scale using the below formula:

SF-12v2 Scale	**Sum items**	**Lowest, highest possible raw scores**	**Possible raw score range**
Physical Functioning (PF)	2a,2b	2,6	4
Role Physical (RP)	3a,3b	2,10	8
Bodily Pain (BP)	5	1,5	4
General Health (GH)	1	1,5	4
Vitality (VT)	6b	1,5	4
Social Functioning (SF)	7	1,5	4
Role Emotional (RE)	4a,4b	2,10	8
Mental Health (MH)	6a,6c	2,10	8

Transformed Scale = $\frac{\text{Actual raw summed score - lowest possible raw score}}{\text{Possible raw score range}} * 100$

Step 3: Convert transformed raw scale domain scores to norm-based domain scores

To convert the transformed raw scale domain scores (as per step 2) to norm domain scores:

1) Firstly Z-scores for each domain must be calculated using the formulas below (based upon 1998 general US population):

   PF_Z = (PF – 81.18122) / 29.10558

   RP_Z = (RP – 80.52856) / 27.13526

   BP_Z = (BP – 81.74015) / 24.53019

   GH_Z = (GH – 72.19795) / 23.19041

   VT_Z = (VT – 55.59090) / 24.84380

   SF_Z = (SF – 83.73973) / 24.75775

   RE_Z = (RE – 86.41051) / 22.35543

   MH_Z = (MH – 70.18217) / 20.50597

   PCS_Z= (PF_Z * 0.42402) + (RP_Z * 0.35119) + (BP_Z * 0.31754) + (GH_Z * 0.24954) + (VT_Z * 0.02877) + (SF_Z * -0.00753) + (RE-Z * -0.19206) + (MH_Z * -0.22069)

   MCS_Z= (PF_Z * -0.22999) + (RP_Z * -0.12329) + (BP_Z * -0.09731) + (GH_Z * -0.1571) + (VT_Z * 0.23534) + (SF_Z * 0.26876) + (RE_Z *0.43407) + (MH_Z * 0.48581)

2) Secondly norm-based scores are calculated by multiplying each Z-score by 10 and adding the resulting product to 50
   e.g for Norm-Based PF = 50 + (PF_Z *10)

Appendix B – EQ-5D-5L

To calculate the EQ-5D-5L Index score, firstly each response will be converted to a score of 1-5:

Mobility Response	Scoring
I have no problems in walking about	1
I have slight problems in walking about	2
I have moderate problems in walking about	3
I have severe problems in walking about	4
I am unable to walk about	5
Self-Care Response	
I have no problems washing or dressing myself	1
I have slight problems washing or dressing myself	2
I have moderate problems washing or dressing myself	3
I have severe problems washing or dressing myself	4
I am unable to washing or dressing myself	5
Usual Activities	
I have no problems doing my usual activities	1
I have slight problems doing my usual activities	2
I have moderate problems doing my usual activities	3
I have severe problems doing my usual activities	4
I am unable to doing my usual activities	5
Pain / Discomfort	
I have no pain or discomfort	1
I have slight pain or discomfort	2
I have moderate pain or discomfort	3
I have severe pain or discomfort	4
I have extreme pain or discomfort	5
Anxiety / Depression	
I am not anxious or depressed	1
I am slightly anxious or depressed	2
I am moderately anxious or depressed	3
I am severely anxious or depressed	4
I am extremely anxious or depressed	5

The score from each domain will then be converted to a deduction from full health (using EQ-5D-5l value set for England) and aggregated to form the index score:

Full health (11111) =	1.000
Mobility level 1 (No problems):	0
Mobility level 2 (Slight):	0.058
Mobility level 3(Moderate):	0.076
Mobility level 4 (Severe):	0.207
Mobility level 5 (Unable)	0.274
Self-care level 1 (No problems):	0
Self-care level 2 (Slight):	0.050
Self-care level 3 (Moderate):	0.080
Self-care level 4 (Severe):	0.164
Self-care level 5 (Unable)	0.203
Usual activities 1 (No problems):	0
Usual activities 2 (Slight):	0.050
Usual activities 3 (Moderate):	0.063
Usual activities 4 (Severe):	0.162
Usual activities 5 (Unable):	0.184
Pain/discomfort 1 (No pain):	0
Pain/discomfort 2 (Slight):	0.063
Pain/discomfort 3 (Moderate):	0.084
Pain/discomfort 4 (Severe):	0.276
Pain/discomfort 5 (Extreme):	0.335
Anxiety/depression 1 (No Anxiety)	0
Anxiety/depression 2 (Slight):	0.078
Anxiety depression 3 (Moderate):	0.104
Anxiety/depression 4 (Severe):	0.285
Anxiety/depression 5 (Extreme):	0.289

For example, the index score for a patient with health state 23245 is calculated as follows:

Full Health =1

Mobility deduction = 0.058

Self-care deduction= 0.080

Usual Activity deduction = 0.050

Pain/discomfort deduction = 0.276

Anxiety / depression deduction = 0.289

Index Score = 1 – (0.058 + 0.080 + 0.050 + 0.276 + 0.289) = 0.247

Appendix C – OIQoL-A

Introduction and Item Scoring

The OIQoL-A consists of 37 items; however, the first 4 items are informational only and are not included in the scoring. Therefore, up to 33 items (Q5-Q37) are included in the total score. Thirteen items are numerical rating scales (NRS) scored between 0 and 10, and the remainder are categorical items scored between 0 and 4. One item, Q26, has a "not applicable" answer option, and therefore may not be included in the total score. The scoring of individual items is summarized in the table below.

Items	Response scale	Scoring
Q 5, 6, 7 Q 9, 10, 11 Q 14 Q 15, 16, 17	11-point NRS (0-10)	0-10
Q 8, 12, 18	11 points NRS (0-100%)	0-10
Q13	5-point Likert scale: *"Not at all"* to "*Every night"*	0-4
Q 19-25 Q 34-37	5-point Likert scale: "*No difficulty*" to "*Unable to do"*	0-4
Q 26	"*Not applicable*" option included	0-4 or Missing
Q27-33	5-point Likert scale: "*Never*" to "*All the time"*	0-4

Total Score Calculation (0-100)

The total score is calculated on a 0-100 scale, where higher scores indicate a greater (negative) impact on quality of life. If all items are completed and Q26 is not marked as "not applicable," the total score calculation is based on 33 items as follows:

[((Q5/10)*(1/33))+…((Q18/10)*(1/33))+((Q13/4)*(1/33))+ ((Q19/4)*(1/33))+… + ((Q37/4)*(1/33))]*100

If Q26 is marked as "not applicable," or any additional items are missing as a result of incomplete data forms, the denominator weighting for each item should be the total number of completed, scored items. However, if more than 20% of the items on the instrument have missing data, the total score should not be calculated and should be considered missing.

Pain Subscale

The "worst" pain item (Q6) will be used to evaluate pain severity.

Impact on Activities of Daily Living (ADL)

Items #19-26 comprise the ADL domain, so that the subscale score is based on a total of 7 or 8 items. For example, if Q26 is not marked as "not applicable" and no items are missing:

[(Q19/4)*(1/8))+… ((Q26/4)*(1/8))]*100

If Q26 is marked as "not applicable" and no items are missing:

[(Q19/4)*(1/7))+… ((Q25/4)*(1/7))]*100

Similarly if more than 20% of the items on the instrument have missing data, the ADL subscale score should not be calculated and should be considered missing.