STATISTICAL ANALYSIS PLAN

RANKL-blockade for the treatment of erosive osteoarthritis (OA) of interphalangeal finger joints. Randomized, double blind, placebo-controlled study to evaluate the efficacy of denosumab 60mg sc every 3 months in patients with erosive osteoarthritis of the interphalangeal finger joints

EudraCT nummer: 2015-003223-53
EC nummer: 2015/0958

Author: Ruth Wittoek
Version 1 dd. 29/03/2020

Content:

1. Study identification
1.1 Study details
1.2 SAP details

2. Background and rationale of the study
2.1 Background
2.2 Objectives and hypotheses
2.3 Study type
2.4 Randomization details and masking
2.5 Sample size calculation

3. Analyses
3.1 Timing of final analysis

4. Study population
4.1 Inclusion and exclusion criteria
    4.1.1 Inclusion criteria
    4.1.2 Exclusion criteria
4.2 Data sets analyzed
    4.2.1 Safety Set
    4.2.2 Full Analysis Set
    4.2.3. Per Protocol Set
4.3 Protocol violations
4.4 Presentation of withdrawal and handling of missing data

5. Outcome measures
5.1 Primary endpoints
    5.1.1 Definition
    5.1.2 Target joints
    5.1.3 Reliability exercise
    5.1.4 Final and consensus scores
5.2 Secondary radiographic endpoints and outcome measures
    5.2.1 Definition
    5.2.2 Target joints
    5.2.3 Reliability exercise
    5.2.4 Final and consensus scores
5.3 Secondary clinical endpoints and outcome measures
    5.3.1 Clinical changes and patient reported outcome measures
    5.3.2 Sonographic changes
5.4. Exploratory endpoints
    5.4.1 DEXA changes
    5.4.2 Post hoc exploratory radiographic endpoints
5.5 Summary of endpoints

6. Statistical analysis

6.1 Summary of baseline data and flow of participants
6.2 Analysis of safety data
6.3 Primary outcome analysis
6.4 Secondary outcome analysis
    6.4.1 Secondary outcome analysis for continuous outcome measures
    6.4.2 Secondary outcome analysis for categorical outcome measures
6.5. Subgroup Analysis
6.6 Descriptive analysis
6.7 Sensitivity Analyses
6.8 Adjustment for multiplicity

7. Safety data

1. Study identification

1.1 Study details

Title: RANKL-blockade for the treatment of erosive osteoarthritis (OA) of interphalangeal finger joints: a Randomized, double blind, placebo-controlled study to evaluate the efficacy of denosumab 60mg sc every 3 months in patients with erosive osteoarthritis of the interphalangeal finger joints

Trial registration number: EudractNr: 2015-003223-53
EC approval number: EC nummer: 2015/0958
Principal Investigators: Dirk Elewaut, Gust Verbruggen, Ruth Wittoek
Amgen Reference Number 20149056

Protocol: final protocol version 1.0 dd 12/08/2015 – Last update version 4.0 17/10/2018

1.2 SAP details

SAP author: Ruth Wittoek
Responsibility in the trial: PI, statistical analysis, blinded sonographer
Senior statistician: Roos Colman

2. Background and rationale of the study

2.1 Background

Erosive osteoarthritis (OA) of the interphalangeal (IP) finger joints is considered an inflammatory subset of osteoarthritis of the hand. Its inflammatory clinical presentation and destructive nature are unmistakable. The accumulation of destructive changes in the IP joints eventually results in considerable disability. Hitherto, no therapeutic interventions are available that act on underlying disease mechanisms and therefore slow down the structural progression in erosive hand OA. The current standard treatment of care in these patients is limited to symptomatic therapy to reduce pain.

In erosive hand OA, bone resorption at the subchondral bone of IP finger joints is a part of the pathogenetic process. This bone resorption is clearly seen on conventional radiographs (CR). Moreover, other features of structural damage can be seen on CR: joint space narrowing, subchondral sclerosis and development of bony proliferation or osteophytes. The osteolytic 'erosive' lesions do characterize erosive hand OA compared to the more common type or non-erosive type of hand OA. This bone resorption induces the collapse of the subchondral plate, normally supporting the overlaying articular cartilage. This is compatible with a pathologic osteoclastic activity supported by the effects of Receptor Activator of Nuclear Factor kappa-β Ligand (RANKL). RANKL is a key driver of maturation and activation of osteoclasts in bone in health and disease. In pathologic conditions, RANKL can be strongly induced in a variety of cell types including stromal cells under the influence of locally produced proinflammatory cytokines such Tumour necrosis factor alpha (TNFα) and Interleukin (IL) 1β.
At the same time, resorption of articular cartilage of the affected IP joints develops. Consequently, the joint space narrowing appears on CR. Key factors in this process are likely to be TNFα and IL-1β which both have important catabolic effects on human chondrocytes. During the course of the disease, inflammatory processes in the synovial

membrane of IP finger joints are clearly seen. Hence, release of cytokines will induce important catabolic effects on the neighbouring chondrocytes.

Thus, similar as observed in other destructive processes noted in inflammatory rheumatic diseases, the TNFα/IL-1β/RANKL-pathway appears to be a key therapeutic target in erosive hand OA.

2.2 Objectives and hypotheses

We hypothesize that treatment with denosumab, RANKL inhibition in erosive IP OA can reduce catabolic osteoclastic activity or structural erosive progression in already affected joints and prevent development of new erosive joints. Moreover, through inhibition of the RANKL pathway, inflammation can be suppressed. By suppressing inflammation and structural progression, beneficial effect on pain experience and functional impairment might follow.

The **primary objective** of the study is to assess the effect of denosumab on the reduction of radiographic erosive progression using GUSS™ (Ghent University Score System) from baseline to week 24.

The **secondary radiographic objective** is to evaluate a reduction in radiographic erosive progression as defined by diminishing the appearance of new erosive IP finger joints from baseline to week 24 and week 48, and between week 24 and week 48.

Secondary clinical objectives are to assess if denosumab provides clinical benefits (improvement of pain, functional limitations (FIHOA and AUSCAN), swollen and tender joint count and patient global assessment of efficacy, and grip strength) compared to placebo at all timepoints, and to assess the effect of denosumab compared to placebo on sonographic features at week 12 and week 48 from baseline. **Exploratory objectives** are to study DEXA changes and mean erosive and remodeling scores by GUSS.

2.3 Study type

A randomized, double blind, placebo-controlled phase 2 clinical trial

2.4 Randomization details and masking

After screening, randomization to placebo or denosumab is randomly (1:1) done by use of a block randomization scheme with a fixed block size of four. The randomization list is generated by a co-worker independent of the study and not being involved in any procedure during the study. Study medication is provided by the pharmacy department. Medication and placebo syringes are identical in appearance and smell. Patients, outcome assessors (GV, RW) and data analyst (RW) retain masked for treatment allocation until lock of the study database.

2.5 Sample size calculation

Sample size calculation was performed based on the estimated difference in GUSS™ over time (Verbruggen G et al. ARD 2010;69(5):862-7). Several assumptions were made, based on data from a previous clinical trial (Verbruggen G et al. ARD 2012;71(6):891-8).

The following assumptions were made:

- the natural progression (mean change) that can be expected over a period of 6 months is

≥ 24 units (data from the placebo treated group),
- the mean difference in GUSS™ change
between the placebo and adalimumab treated group after 6 months was ≥20 units. This was considered as clinically significant since :
1. the smallest detectable difference of GUSS™ was calculated to be 40 units (Verbruggen G. et al. ARD 2010;69(5):862-7) and improved to 10 units after intensive training.
2. the standard deviation of the mean change in GUSS™ is 29,
3. based on the above data, a total change of at least 20 units in GUSS™ in the treatment group is considered to be a clinical relevant effect from a treatment.

The proposed study involves two treatment arms. The level of significance (α) is 0.05.
A sample size of 46 patients in each treatment arm will have 90% power to detect a difference in mean change GUSS™ of 20 units between the placebo and treated group, assuming that the standard deviation is 29 using a t-test with a two-sided 0.05 level of significance.
Taking into account an attrition rate of 8%, a total of 50 patients (46/(1 – 0.08)) should be included in each arm.
From a placebo controlled trial with adalimumab, we learned that, the risk that an individual IP joint evolves from J/S phase to the E phase is 2-3% per year. This risk increases to 15% for joints with a clinical effusion and to 25% for a painful joint with effusion (non-published data). Adalimumab therapy reduced this risk for these inflammatory joints from 25% to 3% (Verbruggen G et al. ARD 2012;71(6):891-8).

Fifty patients in each arm are needed to demonstrate a similar effect of denosumab with a power of 80%.
This sample size analysis took into account the following assumptions :

1) denosumab has a similar effect as adalimumab
2) presence of at least 1 inflamed joint (defined by clinical and sonographic effusion and presence of pain) per patient at baseline.
3) 5% drop-out
4) The proposed study involves two treatment arms. The level of significance (α) is 0.05.
5) a similar background risk for evolution from J/S to E phase.

Taken into consideration both outcome measures, a minimum of 50 patients is required in both treatment arms in order to provide sufficient power for the study.

3. Analyses

3.1 Timing of final analysis

The SAP will be finalized prior to the database lock. The last visit of the last patient was foreseen in June 2019.
The SAP will be finished by 30th of March 2020. Data lock will be done after 1st of April 2020. Statistical analyses will be performed from 2nd of April till approx. end of June 2020. Report to be expected End of September 2020.

4. Study population

4.1 Inclusion and exclusion criteria

Cfr Protocol (cfr final version 4.0 17/10/2018)

4.1.1 Inclusion criteria

A subject will be eligible for study participation if he/she meets the following criteria:

- Males and females ≥ 30 years of age.
- Subjects with hand OA having suffered from transient inflammatory attacks of the interphalangeal finger joints characteristic for what has been termed 'inflammatory' or 'erosive' hand OA.
- Subjects with hand OA showing inflammatory signs, either clinically or ultrasonographically, of the interphalangeal finger joints.
- Subjects with hand OA in which at least 1 interphalangeal finger joint has the typical appearance on the X-rays of a 'J' or 'E' phase joint as defined by the Verbruggen and Veys radiographic scoring system
- Subjects with hand OA where at least 1 interphalangeal finger joint in the 'J' or 'E' phase presents a palpable swelling.
- Able and willing to give written informed consent and to comply with the requirements of the study protocol.

4.1.2 Exclusion criteria

A subject will be excluded from the study if he/she meets any of the following criteria:

- Patients with known hypersensitivities to mammalian-derived drug preparations.
- Patients with clinically significant hypersensitivity to any of the components of Prolia.
- Current and/or Prior treatment with any investigational agent within 90 days, or five half-lives of the product, whichever is longer.
- Previous administration of denosumab from clinical trials or others (e.g. commercial use).
- Vitamin D deficiency [25(OH) vitamin D level < 20 ng/mL (< 49.9 nmol/L)]. Possibility of replenishment and re-screening.
- Subjects with current hypo- or hypercalcemia (normal serum calcium levels: 8.5-10.5 mg/dl or 2.12-2.62 mmol/L).
- Patients currently under bisphosphonate (BP) treatment or any use of oral BPs within 12 months of study enrollment or intravenous BPs or strontium ranelate within 5 years of study enrollment
- Prior use of any chondroprotective drug within 90 days e.g. chondroitin sulfate, glucosamine, avocado-soybean unsaponifiables, tetracyclins, corticosteroids (oral, intramuscular, intra-articular or intralesional).
- Prior use of any immunomodulating drug with possible effects on proinflammatory cytokine metabolism within 90 days a.o. corticosteroids (oral, intramuscular, intra-articular or intralesional), methotrexate, sulfasalazine, leflunomide, D-Penicillin, anti-malarials, cytotoxic drugs, TNF blocking agents.
- History of drug or alcohol abuse in the last year.

- Patients suffering from chronic inflammatory rheumatic disease (e.g. rheumatoid arthritis, spondylarthropathy, psoriatic arthritis, gout, chondrocalcinosis or other auto-immune diseases, e.g. systemic lupus erythematosus).
- History of cancer or lymphoproliferative disease within the past five years, other than a successfully and completely treated squamous cell or basal cell carcinoma of the skin or cervical dysplasia, with no recurrence within the last two years.
- History of any Solid Organ or Bone Marrow Transplant.
- Comorbidities: significant renal function impairment (glomerular filtration < 30 ml/min/1.73m^2 or <50% of normal value), uncontrolled diabetes, unstable ischemic heart disease, congestive heart failure (NYHA III, IV), uncontrolled hypo or hyperparathyroidism, active inflammatory bowel disease, malabsorption, liver failure or chronic hepatic disease (serum AST/ALT levels 3 times above normal), recent stroke (within three months), chronic leg ulcer and any other condition (e.g,. indwelling urinary catheter) which, in the opinion of the investigator, would put the subject at risk by participation in the protocol.
- Subject has any kind of disorder that compromises the ability of the subject to give written informed consent and/or to comply with study procedures .
- Patient who is pregnant or planning pregnancy; if the female subject is of child-bearing age, she must use a valid mean of contraception during the study and for 9 months after last dose of study medication. For males with a partner of childbearing potential: subject refuses to use 1 effective methods of contraception for the duration of the study and for 10 months after the last dose of study medication.
- Female subjects who are breast-feeding.
- History of osteonecrosis of the jaw, and/or recent (within 3 months) tooth extraction or other unhealed dental surgery; or planned invasive dental work during the study.

4.2 Data sets analyzed

4.2.1 Safety Set

The Safety Set includes all patients who receive at least 1 dose of denosumab or placebo. This set will be used to summarize demographics, baseline clinical characteristics and DEXA characteristics, adverse events, laboratory results, vital signs, and safety analysis.

4.2.2 Full Analysis Set

The Full Analysis Set (FAS) will include all patients who are randomly assigned to groups and received at least 1 dose of denosumab or placebo. Intention-to-treat analysis is performed on FAS. This analysis set will be used for efficacy analysis.

4.2.3. Per Protocol Set

The Per-Protocol (PP) Set is a subset of the FAS population, excluding patients with major protocol deviations and who do not complete the study until week 48. This population may be used to summarize efficacy.

The Safety Set, FAS and PP set will be identified prior to database lock.

4.3 Protocol violations
Only major protocol violations thought to affect the valid assessment of the efficacy of the study drug will be considered. A study team, consisting of at least the Principal investigator, study physician and one independent clinician will review the case and make a judgement whether the patient should be excluded from the PP population. Whether a major protocol violation occurred with the consequence of being excluded from the PP population, will be decided before breaking the randomization code.
The following situations will be considered as major protocol violation:

- patients not fulfilling eligibility criteria
- patients having received prohibited concomitant medication
- patients with protocol-required procedure(s) not adhered to
- patients having received less than 50% of study medication according to schedule of study
- patients who withdrew or were withdrawn during the study
- collection of subjective data from patients after they were unblinded (but not withdrawn) during the study period due to medical reasons will not be used in analyses. Objective data will be used in case the study team remained blind to treatment allocation.

4.4 Presentation of withdrawal and handling of missing data
Reasons for early withdrawal from the first year of the study will be summarized.
Missing data will not be imputed for descriptive statistics. For analysis of primary endpoint, data available from all time points will be used in each analysis, accounting for within patient clustering effects by using GEE models. Missing value after the last available visit or questionnaire or assessment due to dropping out the study for any reason will be imputed. The predictors used for the imputation model will be: randomization group, baseline value and values at other time points available (measurements at week 24 will be used to impute measurements at week 48), presence of baseline inflammation (if analyses on 'joint level'), baseline number of affected joints (S, J, E, E/R).

Surgically modified hand joints (arthrodesis or arthroplasty or amputation) at baseline will be considered missing for joint-specified outcomes (tenderness, soft swelling, radiography and ultrasound) at all time points.

5. Outcome measures

5.1 Primary endpoints
5.1.1 Definition
The primary efficacy endpoint is to assess the effect of denosumab on the reduction of radiographic erosive progression from baseline to week 24. The outcome measure being used is GUSS™ (Ghent University Score System). This scoring system is composed of 3 subdomain: subchondral plate, subchondral bone and joint space. Specific features referring to underlying pathology of the disease are being scored on a numerical scale from 0 to 100, with increments of 10. The maximum score refers to a normal joints or completely restored (i.e., non erosive) joint. A total score per joint is made by an equally weighted sum score of all 3 subdomains (min. 0; max. 300). Details of GUSS™ are described elsewhere and an educational atlas is available to consult by the readers (Verbruggen G. et al. ARD

2010;69(5):862-7). For good understanding, the scoring system can change in positive (i.e., more remodeling) or negative direction (i.e., more erosive progression). The smallest detectable difference is 40 units and improved to 10 units after intensive training: this implies that an absolute change over 40 (10) units on a total of 300 in one joint shows significant change over measurement error.

5.1.2 Target joints

Target joints, prone to response to therapy, either erosive progression, stable condition, or remodeling, are selected on baseline CR: upon selection of target joints, the following criteria needs to be fulfilled:

- **A joint in J or E phase, according the Verbruggen and Veys anatomical scoring system, on baseline Xrays**. According to inclusion criteria of the study, a study patient has at least one target joint upon inclusion. If several target joints are available, all are included for efficacy analysis.
- **Presence of inflammatory activity in the joint**, defined by presence of soft tissue swelling upon clinical examination and presence of sonographic swelling (either synovial proliferation or effusion) at baseline

5.1.3 Reliability exercise

Radiographs will be read independently by two trained assessors, blinded to randomization and clinical data. The radiographs will be read with knowledge of time sequence. First reliability analysis will be performed on the readings of the first 20 patients. If ICCs total GUSS™ scores do not exceed 0.80, retraining will be performed upon further reading. Another reliability analysis will be performed on the following 20 readings until ICCs total GUSS™ exceeds 0.80.

Each reader will perform a second reading of a series of images (first 20 patients) to calculate intra-reader reliability, minimal one month after and blinded to the previous reading.

5.1.4 Final and consensus scores

The final scores of radiographic scorings will be the agreement scores amongst the two assessors. In case of disagreement, a consensus score will be made. The consensus score will be made by the two assessors by discussing and reanalyzing the radiograph and joint of interest.

5.2 Secondary radiographic endpoints and outcome measures

5.2.1 Definition

The secondary efficacy endpoint is to assess the effect of denosumab on the development of new erosive joints from baseline to week 24 and week 48, and to assess the effect of denosumab on the erosive progression between week 24 and week 48. The number/ percentage of new erosive (defined as new J or E) joints, according to the Verbruggen and Veys anatomical scoring system amongst the baseline non-erosive or non-remodeled joints (i.e., only baseline N, S, J joints) in a patient in each treatment group at week 24 and 48, and between 24 and 48, will be determined.

5.2.2 Target joints
For analysis of anatomical phase scoring system, all DIP and PIP joints will be assessed and integrated for analysis.

5.2.3 Reliability exercise
Radiographs will be read independently by two trained assessors, blinded to randomization and clinical data. The radiographs will be read with knowledge of time sequence. First reliability analysis will be performed on the readings of the first 20 patients. If unweighted Kappa statistics of anatomical phase scoring system do not exceed 0.80, retraining will be performed upon further reading. Another reliability analysis will be performed on the following 20 readings until unweighted Kappa exceeds 0.80.
Each reader will perform a second reading of a series of images (first 20 patients) to calculate intra-reader reliability, minimal one month after and blinded to the previous reading.

5.2.4 Final and consensus scores
The final scores of radiographic scorings will be the agreement scores amongst the two assessors. In case of disagreement, a consensus score will be made. The consensus score will be made by the two assessors by discussing and reanalyzing the radiograph and joint of interest.

5.3 Secondary clinical endpoints and outcome measures
The secondary objectives of the study are to explore whether denosumab adds clinical benefit in terms of reduction of pain, number of tender and swollen joints, patient assessment of efficacy, grip strength and functional improvement. Also the effects of denosumab on ultrasonographic features will be assessed.

5.3.1 Clinical changes and patient reported outcome measures
Pain is scored on a numeric rating scale (NRS) from 0 to 10. Changes from baseline to week 24 and week 48 will be calculated.
Functional ability is assessed by two questionnaire for functional impairment, AUSCAN and FIHOA at all visits. AUSCAN is a questionnaire consisting of three domains (pain, stiffness, function) with 15 questions in total: questions are responded on a NRS (0-10). Sum scores range from 0 to 150. FIHOA is a questionnaire consisting of ten questions (NRS 0 – 3): total score ranges from 0 to 30. Number of tender and swollen joints upon clinical examination at PIP2-5 and DIP2-5 of both hands (absence (0)/presence (1), 0-16), mean grip strength of 3 attempts of the most affected hand (defined as number of target joints, in case equal between left and right, the dominant hand is considered), and patient global assessment of efficacy (0-10, NRS) are assessed at all visits. All these outcomes are numerical outcomes: changes from baseline to all visits will be calculated and compared between denosumab and placebo group.

5.3.2 Sonographic changes
Presence of 4 variables are scored at baseline, week 12 and week 48: effusion (0-3), synovial proliferation (0-3), PD signal (0-3) and presence of erosions (0/1) at PIP2-5 and DIP2-5. Sum scores on patient level are made for effusion (0-48), synovial proliferation

(0-48), synovitis (effusion plus synovial proliferation)(0-96), PD signal (0-48), and erosions (0-16). Change from baseline will be measured at week 12 and week 48.

5.4. Exploratory endpoints

5.4.1 DEXA changes

T-score and Z-score of bone mineral density (BMD) at femoral neck and lumbar spine are obtained from baseline and Week 48. Percentage of patients with low BMD suggestive of osteoporosis (T-score ≤-2.5), presence of osteopenia (-2.5 < T-score ≤ -1.5) and normal BMD (T-score >-1.5) will be calculated. Changes in percentage of the above categories and changes in absolute T-scores at week 48 from baseline will be calculated for femur and spine and compared between groups.

5.4.2 Post hoc exploratory radiographic endpoints

In order to extrapolate the scoring system on joint level to patient level, a mean progression score is calculated. The mean progression score is the absolute sum score of all negative changes (i.e. erosive progression) of all target joints divided by the number of target joints for that patient:

Mean progression score = $$\frac{\sum |\Delta \text{GUSS erosive progression of target joints}|}{n}$$

With Δ GUSS™ assessed between baseline and Week 24 and n = the number of target joints in a patient

Equally, the mean remodeling score per patient will be calculated and equals the absolute sum score of all positive changes (i.e., remodeling) of all target joints divided by the number of target joints for that patient:

Mean remodeling score = $$\frac{\sum |\Delta \text{GUSS remodeling of target joints}|}{n}$$

With Δ GUSS™ assessed between baseline to Week 24 and n = the number of target joints in a patient. This outcome measure will be an exploratory outcome measure.

Also the mean progression score and mean remodeling score by GUSS™ from baseline to week 48, and week 24 to week 48 will be calculated.

Target joints are defined here by presence of any GUSS changes in the joint throughout the study irrespective of the baseline anatomical score or inflammatory status (clinical swelling or US inflammation).

5.5 Summary of endpoints

Primary endpoint (continuous): GUSS™ (0-300) at week 24

Secondary endpoints:

- Percentage of new erosive joints (J/E) by Verbruggen and Veys amongst the baseline non-erosive or non-remodeled joints (i.e., only baseline N, S, J joints) in a patient
- GUSS™ at week 48 (0-300)
- NRS pain at week 24 (0- 10)
- NRS Patient global assessment of efficacy at week 24 (0-10)
- Functional Index of Hand OA (FIHOA) at week 24 (0-30)
- Australian/Canadian Hand OA index (AUSCAN) at week 24 (0-150)
- Number of tender joints at week 24 (0-16)
- Number of swollen joints at week 24 (0-16)
- Grip strength at week 24 (in kg)
- Ultrasound synovitis score at week 12 (0-96)
- Ultrasound synovial proliferation scores at week 12 (0-48)
- Ultrasound effusion score at week 12 (0 – 48)
- Ultrasound PD score at week at week 12 (0-48)
- Ultrasound erosion score at week 48 (0-16)
- Clinical endpoints (AUSCAN, FIHOA, NRS pain, NRS patient global, swollen joints count, tender joint count, grip strength) and ultrasound parameters at other time points: for clinical endpoints points (week 6, week 12, week 24, week 36), for ultrasound (week 48)
- GUSS™ from week 24 to week 48

Exploratory endpoints:

- DEXA (T- and Z-score at femur total, femoral neck, distal radius and lumbar spine) at week 48
- Percentage in categories according to BMD values (normal, osteopenia, osteoporosis) at week 48
- Mean progression score GUSS™ at week 24 and 48, and from week 24 to 48
- Mean remodeling score GUSS™ at week 24 and 48, and from week 24 to 48

6. Statistical analysis

All statistical calculations will be performed using R version 3.6.1, unless otherwise specified.

6.1 Summary of baseline data and flow of participants

A flow diagram will be produced, in accordance with the consort guidelines (http:/www.consort-statement.org/)

All continuous variables (both primary as secondary endpoints: GUSS™, NRS pain, NRS patient assessment efficacy, AUSCAN score, FIHOA score, grip strength, tender sum scores, swollen sum scores, sonographic sum scores) will be checked for normality. For continuous variables, summary statistics will include sample size (Number of patients, N),

mean and standard deviation (SD) in case of (approximate) normal distribution. Otherwise, median, minimum, maximum values, $25^{th}/75^{th}$ percentile will be presented. Frequency count and percentages will summarized for categorical variables. If not normal distributed variables, non-parametric testing will be performed.

All demographic, baseline data and medical history information will be summarized using Safety Set. Demographic variables will be studied (age, disease duration, sex, etc). Demographic and baseline characteristics will be compared between treatment groups. If unbalanced data, a sensitivity analysis will be performed for the primary analysis by adding the variables for which an unbalance was found as a covariate to the model.

6.2 Analysis of safety data
Analyses of safety data will be performed in the defined Safety Set (cfr. Paragraph 4.2).

6.3 Primary outcome analysis

Primary efficacy analyses will be performed in an intention to treat (ITT) approach on the FAS population.

Changes in GUSS will be analyzed at joint level with generalized estimating equations (GEE), accounting for within-patient clustering. Robust standard errors will be used and the working correlation structure specified exchangeable. Data from all available time points will be used. The independent variables included in the model are treatment group, visit number (categorical), interaction between treatment group and visit number, and the baseline value of the dependent variable (continuous).
If unbalances of demographic variables were found at baseline. A sensitivity analysis will be performed for the primary analysis by adding the variables for which an unbalance was found as a covariate to the model.

All efficacy analyses will be presented by a point estimate of the difference between the treatment groups, with a 95% confidence interval (95% CI) and the two-sided p-value. A p-value below 0.05 ($p<0.05$) will be considered statistically significant.

6.4 Secondary outcome analysis
All secondary outcomes are measured at the level of the patient.

6.4.1 Secondary outcome analysis for continuous outcome measures
Continuous endpoints will be analyzed with generalized estimating equations (GEE) using robust standard errors and the working correlation structure specified exchangeable. Data from all available time points will be used. The independent variables included in the model are treatment group, visit number (categorical), interaction between treatment group and visit number, and the baseline value of the dependent variable (continuous).

$$Y_{ij} = \beta_0 + \beta_1 * \text{Treatment}_i + \beta_2 * \text{Baseline}_i + \beta_3 * \text{Visit}_i + \beta_4 * \text{Visit}_i * \text{Treatment}_i + e_{ij}$$

Where

Y_{ij} = Outcome (continuous) of interest for patient i at visit j
i = patient
j = visit
β_0 = Intercept
β_1 = Regression coefficient for treatment (0 = placebo; 1 = Verum)
β_2 = Regression coefficient for baseline value of Y (continuous)
β_3 = Regression coefficient for visit (1 = week 6; 2 = week 12; 3 = week 24; 4 =week 36; 5 = week 48)
β_4 = regression coefficient for treatment-visit interaction

For estimation of the treatment effect after week 24, the model will include separate estimates of the treatment effect at 6, 12 and 24 weeks, with the coefficient corresponding to treatment*visit interaction term ($\beta_{4)}$ of week 24 being the estimate of interest, with the corresponding p-value indicating statistical significance. Data of week 36 and 48 will not be included in the model for estimation of the treatment effect at week 24. For estimation of the treatment effect after week 36 and 48, the model will additionally include estimate(s) for treatment effect at week 36, and week 36 and 48, respectively.

Secondary efficacy analyses and exploratory endpoint analyses will be done in the PP population.

6.4.2 Secondary outcome analysis for categorical outcome measures
For the binary outcome measures (for example E joint yes/no) a GEE logistic regression for grouped binomial data will be applied. Odds ratios (OR), 95% confidence interval (95% C.I.) and p-value will be calculated by GEE model with treatment and time effect, to assess the odds for difference in new erosive progression from baseline between denosumab treated group and placebo group at week 24 and 48.

6.5. Subgroup Analysis
For the primary outcome measure, a subgroup analysis for inflammatory activity in the joint (Yes/No) is planned. The interaction between the presence of baseline inflammation and treatment effect on change in GUSS scores over 24 weeks will be tested.

6.6 Descriptive analysis
For the primary outcome measure, another descriptive efficacy analysis is performed on an extended group of target joints: all joints showing any progression to J, E or E/R phase throughout the study that were not defined J or E at baseline will be included to study the change in GUSS progression at week 24.

6.7 Sensitivity Analyses

For the primary outcome measure, sensitivity analyses will be performed:
- An adjusted analysis: including the unbalanced variables in the model
- A Per Protocol (PP) analysis.

6.8 Adjustment for multiplicity

As there is only one primary outcome, no adjustments for multiple testing will be performed.

7. Safety data

Safety will be assessed in all patients who randomly assigned to a treatment group and started the allocated intervention (Safety Set). The safety of denosumab will be assessed based on incidence and severity of adverse events (AEs), serious adverse events (SAEs), and changes from baseline through trial completion in routine clinical laboratory tests (Hb, white blood count, calcium, creatinine (GFR), AST, ALT, phosphor), and pregnancy tests for female patients of childbearing potential. The number of AEs, SAEs, withdrawal because of AEs and changes in biochemistry data between baseline and all time points for calcium, and baseline and week 12 and 48 for other lab data will be determined.