STATISTICAL ANALYSIS PLAN

A Phase 3 Randomized, Double-blind, Placebo-controlled, Parallel-group Study to Evaluate the Safety and Efficacy of Denosumab in Pediatric Subjects With Glucocorticoid-induced Osteoporosis

Protocol Number: **20140444**

Version: ***Version 4.0***

Date: **12 NOV 2023**

Authors: [redacted]

NCT Number: NCT03164928
This NCT number has been applied to the document for purposes of posting on Clinicaltrials.gov

Version Number	Date (DDMMMYYYY)	Summary of Changes, including rationale for changes
Original (v1.0)	**21 Sep 2017**	**NA. The first version.**
Amendment 1 (v2.0)	**23 Oct 2018**	**This statistical analysis plan (SAP) was updated based on the protocol amendment 2 for Denosumab Study 20140444 dated 25 May 2018.**
Amendment 2 (v3.0)	**30 Aug 2023**	**This statistical analysis plan (SAP) was updated based on the updates in protocol amendment 4 for Denosumab Study 20140444 dated 10Jul2023. Changes including, but not limited to, the following are incorporated:** • **Updated the fracture endpoints and its analysis.** • **Removed the following safety endpoints and its analysis: Metaphysical Index, Molar eruption, Mandibular shaping.** • **Made administrative and editorial changes throughout the document.**
Amendment 3 (v4.0)	**12 Nov 2023**	**This statistical analysis plan (SAP) was updated to add a sensitivity analysis for the primary efficacy endpoint (based on a FDA request), along with other updated methods.** • **Added the definition of age in months in the Section 6.6.** • **Added sensitivity analysis of the primary efficacy endpoint**

		per the FDA request in Section 10.5.1.2. • **Updated the Sections 6.7.4 10.5.2.2 to provide detail on the nonvertebral and long bone fractures consideration.** • **Updated the Tables A.1.4 - A.1.10 in the Appendix A to include visit window for Screening/Baseline visit.** • **Updated Table C2 in Appendix C for Nonvertebral Fracture Locations** • **Updated Appendix E for the CHQ-PF50.** • **Made administrative and editorial changes throughout the document.**

Table of Contents

List of Tables

List of Appendices

Table of Abbreviations

Term or Abbreviation	Definition
AE	Adverse Event
[redacted]	
BMD	Bone mineral density
BMDS	Biomedical data stewardship standard
BMI	Body mass index
[redacted]	
CHAQ	Childhood Health Assessment Questionnaire
CHQ-PF	Child Health Questionnaire - Parent Form
CT	Computerized tomography
DMC	Data monitoring committee
DoB	Date of birth
DXA	Dual-energy X-ray absorptiometry
EC	Ethics committee
eCRF	Electronic case report form
GiOP	Glucocorticoid (GC)-induced osteoporosis
IC	Informed consent
IP	Investigational product
IVRS	Interactive voice response system
MRI	Magnetic resonance imaging
NTX	Cross-linked N-telopeptides of type 1collagen
PK	Pharmacokinetic
SAP	Statistical analysis plan
[redacted]	
SQ	Semiquantitative
SAS	Statistical Analysis System
WBFPRS	Wong-Baker FACES Pain Rating Scale

1. Introduction

The purpose of this Statistical Analysis Plan (SAP) is to provide details of the statistical analyses that have been outlined within the protocol amendment 4 for denosumab Study 20140444, dated 10Jul2023.The scope of this plan includes the final analysis at 36 months, which will be executed by the Biostatistics department unless otherwise specified.

2. Objectives

2.1 Primary

To evaluate the effect of denosumab on lumbar spine bone mineral density (BMD) Z-score as assessed by dual-energy X-ray absorptiometry (DXA) at 12 months in children 5 to 17 years of age with glucocorticoid (GC)-induced osteoporosis (GiOP).

2.2 Secondary

To evaluate the effect of denosumab in children 5 to 17 years of age with GiOP with respect to:

- Change in lumbar spine BMD Z-score as assessed by DXA from baseline to 6, 18, 24, and 36 months
- Change in proximal femur BMD Z-score as assessed by DXA from baseline to 6, 12, 18, 24, and 36 months
- Incidence of X-ray confirmed long-bone fractures and new and worsening vertebral fractures during 12, 24, and 36 months
- Incidence of improving vertebral fractures at 12, 24, and 36 months compared to baseline
- Incidence of new and worsening vertebral and nonvertebral fractures during 12, 24, and 36 months
- Change in Childhood Health Questionnaire – Parent Form-50 (CHQ-PF-50) Physical Summary Score at 12, 24, and 36 months
- Change in CHQ-PF-50 Psychological Summary Score at 12, 24, and 36 months
- Change in Childhood Health Assessment Questionnaire (CHAQ) Disability Index Score at 12, 24, and 36 months
- Change in Wong-Baker Faces Pain Rating Scale (WBFPRS) at 12, 24, and 36 months
- Change in growth velocity, determined by calculating age-adjusted Z-scores for height, weight, and body mass index (BMI), at 12, 24, and 36 months.
- Serum concentration of denosumab at 1 and 10 days, and 6, 12, and 18 months (additional serum denosumab pharmacokinetics [PK] samples to be collected at day 30 and month 3 in a PK [redacted] substudy of up to 15 subjects)

2.3 Safety

To evaluate the effect of denosumab in children 5 to 17 years of age with GiOP with respect to:

- Adverse events and serious adverse events
- Laboratory parameters
- Vital signs
- Antidenosumab antibodies

2.4 Exploratory

3. Study Overview

3.1 Study Design

This is a phase 3 multicenter, randomized, double-blind, placebo-controlled, parallel group study in pediatric subjects, age 5 to 17 years, with GiOP. Twenty four subjects will be enrolled with at least 16 evaluable subjects for the primary analysis.The study will consist of 2 periods, a 24-month treatment period (including a 12-month placebo-controlled period and a 12-month open-label treatment period) and a 12-month off-treament observation period.

Treatment Period

The treatment period will last for 24 months. Subjects will be randomized in a 2:1 (active:placebo) allocation ratio to receive either denosumab 1 mg/kg body weight (BW) (up to a maximum of 60 mg) subcutaneous (SC) every 6 months (Q6M) or matching placebo SC Q6M for the first 12 months. At the end of the 12-month placebo-controlled treatment period all subjects will receive open-label denosumab 1 mg/kg BW (up to a maximum of 60 mg) SC Q6M for 12 months, with the last dose of investigational product (IP) being administered at month 18.

Observation Period

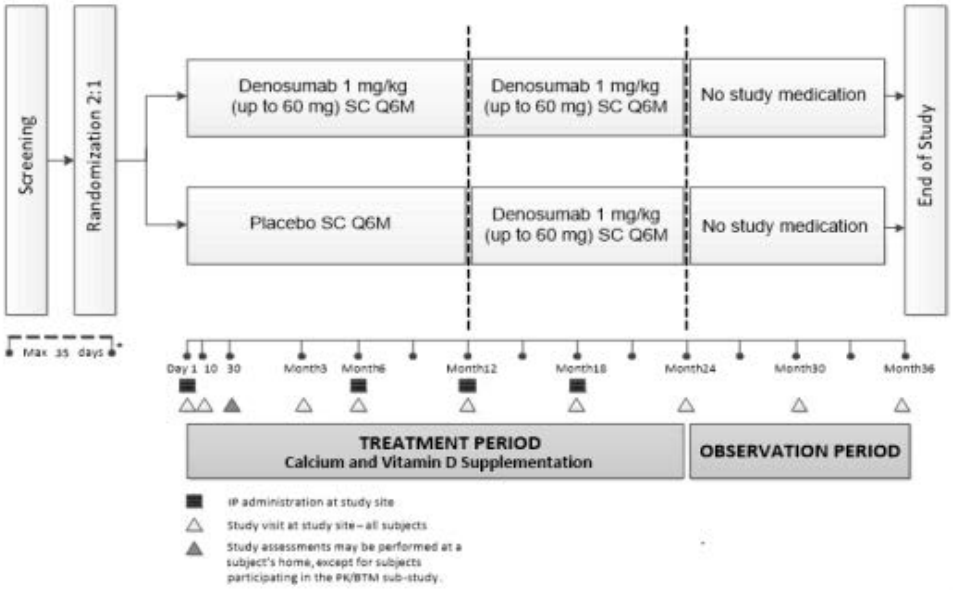
The observation period will start immediately after the 24-month treatment period and last 12 months, during which subjects will receive no IP. Subjects will be followed during the 12-month observation period according to the following guidelines:

- Subjects who are currently on systemic GC for the treatment of the underlying non-malignant condition(s) at the month-24 visit should discontinue IP and transition to another osteoporosis treatment per local standard of care, according to the medical judgment of the investigator, for an additional 12 months or until systemic GC treatment is discontinued
- Subjects who are no longer on systemic GC for the treatment of the underlying non-malignant condition(s) at the month-24 visit should discontinue IP and be followed for an additional 12 months off-treatment; however, subjects who, during the 12-month observation period, resume systemic GC therapy, experience a worsening of their osteoporosis, or require osteoporosis therapy based on the medical judgment of the investigator may resume osteoporosis treatment per local standard of care at the discretion of the investigator

Daily supplements of calcium and vitamin D will be given to all subjects during the 24-month treatment period and if deemed medically warranted by the investigator during the 12-month observation period. The planned length of participation in the study for an individual subject is approximately 3 years, which includes screening (up to 35 days), treatment period (24 months), and off-treatment observation period (12 months). The anticipated dropout rate is approximately 10% in the first year. The overall study design is described by a study schema in Figure 1.

Figure 1. Study Design and Treatment Schema

Study Design and Treatment Schema



3.2 Sample Size

A total of 24 subjects will be randomized into the study (approximately 16 to denosumab and 8 to placebo). Assuming that approximately 10% of subjects will not be evaluable at month 12 for the primary efficacy endpoint due to dropout, the analysis set for the primary efficacy endpoint will include approximately 20 subjects.

4. Study Endpoints and Covariates

4.1 Study Endpoints

4.1.1 Primary Efficacy Endpoint

Change from baseline in lumbar spine BMD Z-score as assessed by DXA at 12 months

4.1.2 Secondary Endpoint(s)

- Change from baseline in lumbar spine BMD Z-score as assessed by DXA at 6, 18, 24, and 36 months
- Change from baseline in proximal femur BMD Z-score as assessed by DXA at 6, 12, 18, 24, and 36 months
- Subject incidence of X-ray confirmed long-bone fractures and new and worsening vertebral fractures during 12, 24, and 36 months
- Subject incidence of improving vertebral fractures at 12, 24, and 36 months compared to baseline
- Subject incidence of new and worsening vertebral and nonvertebral fractures during 12, 24, and 36 months
- Change from baseline in CHQ-PF-50 Physical Summary Score at 12, 24, and 36 months

- Change from baseline in CHQ-PF-50 Psychological Summary Score at 12, 24, and 36 months
- Change from baseline in CHAQ Disability Index Score at 12, 24, and 36 months
- Change from baseline WBFPRS at 12, 24, and 36 months
- Change from baseline in growth velocity, determined by calculating age-adjusted Z-scores for height, weight, and BMI, at 12**,** 24**,** and 36 months
- Serum concentration of denosumab at 1, 10 and 30 days, and 3, 6, 12, and 18 months

4.1.3 Safety Endpoints

- Subject incidence of adverse events and serious adverse events
- Change from baseline in laboratory values
- Change from baseline in vital signs
- Subject incidence of antidenosumab antibodies

4.1.4 Exploratory Endpoint(s)

[redacted]

4.2 Planned Covariates and Subgroups

4.2.1 Covariates

The following covariates are assumed to have prognostic value with respect to the lumbar spine BMD Z-score as assessed by DXA. The association between these covariates with the primary efficacy endpoint might be investigated.

- Sex (Male/Female)
- Race (White, non-White)
- Age (years)
- Baseline lumbar spine BMD Z-score
- Baseline glucocorticoid dose

5. Hypotheses and/or Estimations

The hypothesis of this study is that the change from baseline in lumbar spine BMD Z-score following 12 months of denosumab treatment in children 5 to 17 years of age with GiOP will be greater than placebo.

6. Definitions

6.1 Basic Definitions

Investigational Product

Denosumab 1 mg/kg BW (up to a maximum of 60 mg) SC Q6M or matching placebo SC Q6M

Interactive Voice Response System (IVR System)

The system used to screen, screen failure, randomize and enroll subjects, provide IP box number, and volume administered, as well as to manage denosumab drug supply at the site and to track subject study termination data.

6.2 Study Points of Reference

Baseline

Baseline is the closest recorded measurement before the administration of the first dose of IP. If the measurement is done on the same day as the first dose and the exact measurement time relative to the first dose is unknown, it will be assumed the measurement is done before the administration of the first dose of IP. If a subject does not receive IP, baseline is the closest recorded measurement on or before the enrollment date.

Baseline DXA Assessments

Baseline DXA assessments of lumbar spine and proximal femur are the latest DXA assessments obtained on or before the date of the first IP administration and evaluated by the central imaging vendor. If the baseline DXA assessments need to be repeated, due to poor quality as determined by the central imaging vendor, then the repeat DXA scans obtained within the first 60 Days after the first IP administration may be considered as baseline DXA. Because the baseline DXA will be assessed in duplicate, the baseline BMD value for analysis purpose will be the average of the BMD values from all baseline DXA scans for the same skeletal site.

Baseline Lateral Spine Radiographs

Baseline lateral spine radiograph is the latest radiograph obtained on or before the date of the first IP administration that is evaluated by the central imaging vendor.

End of Study

The end of study date is defined as the date when the last subject is assessed or receives an intervention for evaluation in the study (ie, last subject last visit), following any additional parts in the study (eg, long-term follow-up), as applicable

Study Day 1

The first day of IP administration

Study Day

Pre study day 1: Study Day = (date of interest – date of Study Day 1)

Post study day 1: Study Day = (date of interest – date of Study Day 1) + 1

Analysis Visit Windows

As specified in the protocol, all tests and procedures scheduled to occur on days 10 and 30 should be performed within ± 3 days of the scheduled day, and ± 7 days for visits at Months 3, 6, 12, 18, 24, 30 and 36. However, for analysis purpose the analysis visit windows defined in Appendix A will be used to assign evaluations to the most appropriate nominal visit.

6.3 Study Dates

Informed Consent (IC) Date

The date on which the study informed consent form, and subject assent form if applicable, is signed.

Screening Date

The screening date is the IC date

Enrollment Date

The date on which the investigator confirms that the subject has met all eligibility criteria and the enrollment call is made using the IVR system.

Randomization Date

The date on which the enrollment call is made using IVR system and a randomization number is assigned.

First Dose Date in Double-blind Treatment Period (DBTP)

The date of administration of first dose of IP following randomization (ie, the first date recorded on the IP Administration (IPA) electronic Case Report Form (eCRF) with volume > 0) during the double-blind treatment period. This date is defined as Study Day 1 for subjects who receive at least dose of IP.

First Dose Date in Open-label Treament Period (OLTP)

The date of administration of first dose of open-label IP recorded on the IPA eCRF with volume > 0 during the OLTP following completion of the double-blind treatment period.

Last Dose Date

The date of administration of last dose of IP (ie, the last date recorded on the IPA eCRF with volume > 0)

End of Study Date (i.e End of 36-Month Study Period)

The date recorded on the End of Study eCRF.

End of 12-Month Study Period

The date recorded on the End of Treatment Phase Double Blind eCRF.

End of 24-Month Study Period

The date recorded on the End of Treatment Phase Open Label eCRF.

6.4 Study Time Intervals

Screening Period

The time between the date of informed consent and the enrollment.

Overall Study Period (0-36 Months)

The time period between the first dose date and end-of-study date, inclusive (approximately 3 years for completers).

Double-blind Treatment Period (DBTP)

The time from first dose date to the earlier of 12-month visit date or end-of-study date.

Treatment Period (0-24 Months)

The time from first dose date to the earlier of 24-month visit date or end-of-study date.

Observation Period

The time from end of treatment period (24-month visit) to end of study.

6.5 Subject Disposition

Enrolled

Individuals are considered enrolled when an enrollment call is made using IVRS and a randomization number is assigned. Enrolled individuals are referred to as ‘subjects’.

Exposed to IP

Subjects are defined as exposed to IP if the total amount of IP volume received during the study is greater than zero.

Age at Screening (Informed Consent) (Months)

Number of months as recorded on the subjects demographic eCRF page. A table showing age conversion from months to years is provided in Appendix B.

6.6 Arithmetic Calculations

Age (months) at Visit

Age (months) at visit = Collected age at screening (informed consent) in months + ((visit date – screening date (date of informed consent) +1) / 365.25) * 12

Age at Screening (Informed Consent) (Years)

Number of whole years (rounded down) derived from age in months on the subjects demographic eCRF page.

Age (years) = age (months) / 12

Age at Visit (Years)

Age at visit in years (rounded down) will be calculated as:

Age (years) at visit = (date of visit – date of informed consent + 1) / 365.25 + (age **at screening (informed consent)** in months) / 12

Change from Baseline

For a given variable, study baseline of interest, and timepoint of interest, subtract the endpoint value at the timepoint of interest minus the corresponding baseline value.

Percent change from baseline

The change from baseline divided by study baseline value of interest and multiplied by 100:

(Change from Baseline/ Baseline) * 100

6.7 Study Endpoints

6.7.1 DXA Assessments

Only General Electric Lunar or Hologic bone densitometers will be allowed for this study. The same DXA machine must be used for all DXA assessments for a particular subject. The left side should be used for proximal femur, unless prohibited (eg, hip implant). If the right side must be used or is inadvertently used at screening, then it must be used consistently throughout the study.

Lumbar spine scans should include L1 through L4. At least 2 lumbar vertebrae from L1 to L4 must be evaluable by DXA. Individual vertebral levels may be excluded due to artifact. A vertebral level excluded from one visit will be excluded from all visits with the total lumbar spine BMD calculated based on the evaluable vertebral levels by the central imaging vendor.

Baseline DXAs will be performed as duplicate scans; DXA assessments at all other visits will be performed as single scans. After analysis of the scans by the central imaging vendor, the study site may be asked to re-acquire a scan, because of poor positioning or other technical reasons.

6.7.2 Fracture-related Trauma Severity Definitions

Low Trauma Severity

Assessed by the investigator and collected on the Additional Fracture Details eCRF for each clinical fracture event with the following trauma severity:

- Fall from standing height or less than 20 inches
- Minimal or moderate trauma other than a fall from standing height or less than 20 inches
- Unknown/don't know (including missing)

High Trauma Severity

Assessed by the investigator and collected on the Additional Fracture Details eCRF for each clinical fracture event with the following trauma severity:

- Severe trauma

Pathologic

Assessed by the investigator and collected on the Additional Fracture Details eCRF for each clinical fracture event with the following trauma severity:

- Pathologic fractures

6.7.3 Vertebral Fractures

Lateral radiographs of the thoracic and lumbar spine (T4 through L4) will be obtained at times to determine clinical or morphometric vertebral fractures.

The lateral spine radiographs will be assessed for prevalent, new, worsening and improving vertebral fractures using the Genant Semiquantitative (SQ) scoring method:

- Grade 0: normal (approximately $<$ 20% reduction in anterior, middle, and/or posterior height)
- Grade 1: mild fracture (approximately 20% - 25% reduction in anterior, middle, and/or posterior height)
- Grade 2: moderate fracture (approximately 25% - 40% reduction in anterior, middle, and/or posterior height)
- Grade 3: severe fracture ($>$ 40% reduction in anterior, middle, and/or posterior height)

All vertebral fractures data will be verified by Bioclinica for analysis purpose.

Prevalent Vertebral Fracture

A subject has prevalent vertebral fracture if any thoracic or lumbar vertebrae (T4 to L4) evaluated have SQ grade ≥ 1 at baseline. A subject does not have any prevalent vertebral fracture when the SQ grades of all evaluated vertebrae are 0 on the baseline spinal radiograph. Otherwise, the subject will have an unknown status for prevalent vertebral fracture.

New Vertebral Fractures

Any thoracic or lumbar vertebrae (T4 to L4) evaluated at a post-baseline visit is deemed to have a vertebral fracture when the post-baseline SQ is ≥ 1 compared to a baseline SQ score of 0.

A subject is identified as having a new vertebral fracture at a timepoint of interest if at least one thoracic or lumbar vertebra from T4 to L4 at the timepoint of interest is deemed to have a new vertebral fracture (SQ ≥ 1) compared to the baseline SQ score of 0, excluding any symptomatic new vertebral fracture associated with high trauma severity or a pathologic fracture.

Because the vertebral fractures in the pediatric population can worsen or improve over time, the SQ score of the fractured vertebra may increase or decrease over time, respectively. The vertebral fracture status at a timepoint of interest will always be based on the last radiograph assessment on or before the timepoint of interest relative to the baseline radiograph assessment. For example, if a vertebra is assessed as having 0, 2, 0, and 2 SQ scores at baseline, 12, 24, and 36 months, respectively, then the vertebra will be identified as having a new vertebral fracture at the 12- and 36-month timepoints. However, the vertebra will not be assessed as having a new vertebral fracture at the 24-month timepoint.

Multiple New Vertebral Fractures

A subject will be identified as having multiple new vertebral fractures when more than one thoracic or lumbar vertebra from T4 to L4 is identified as having a new vertebral fracture at the timepoint of interest, and the total number of new vertebral fractures will be the sum of new vertebral fractures at the timepoint of interest.

Worsening Vertebral Fractures

Worsening vertebral fractures occur in vertebrae with prevalent fractures. It is identified when a vertebra SQ score increases by more than 1 grade relative to a baseline SQ score ≥ 1, in any thoracic or lumbar vertebra evaluated, excluding any symptomatic

worsening vertebral fracture associated with high trauma severity or a pathologic fracture.

A subject is identified as having a worsening vertebral fracture at a timepoint of interest if at least one thoracic or lumbar vertebra from T4 to L4 is assessed as having a worsening vertebral fracture compared to the baseline SQ score of ≥ 1. For example, if a vertebra is assessed as having SQ scores of 2, 3, 3, and 2 at baseline, 12, 24 and 36 months, respectively, the vertebra will not be identified as having a worsening vertebral fracture at the 36-month timepoint because the SQ score closest to the 36 months is not greater than the baseline SQ score. However, the vertebra will be assessed as having a worsening vertebral fracture at the 12 and 24-month timepoints.

Multiple Worsening Vertebral Fractures

A subject will be identified as having multiple worsening vertebral fractures when more than one thoracic or lumbar vertebra from T4 to L4 is identified as having a worsening vertebral fracture at the timepoint of interest, and the total number of worsening vertebral fractures will be the sum of worsening vertebral fractures at the timepoint of interest.

New or Worsening Vertebral Fracture

A new and worsening vertebral fracture is identified when there is ≥ 1 grade increase from the baseline grade in any vertebra from T4 to L4, excluding any symptomatic vertebral fracture associated with high trauma severity or a pathologic fracture.

A subject is identified as having a new or worsening vertebral fracture if at least one thoracic or lumbar vertebra from T4 to L4 is assessed as having a new or worsening vertebral fracture at the timpoint of interest.

Multiple New or Worsening Vertebral Fractures

A subject will be identified as having multiple new or worsening vertebral fractures when more than one thoracic or lumbar vertebra is identified as having a new or worsening vertebral fracture at the timepoint of interest, and the total number of new or worsening vertebral fractures will be the sum of new or worsening vertebral fractures at the timepoint of interest.

Improving Vertebral Fractures

For each postbaseline timepoint of interest (12, 24 and 36 months), an improvement in a prevalent vertebral fracture is identified when the SQ score decreases by at least 1 grade relative to baseline SQ score, ie, the last SQ score assessment closest to the time point of interest is smaller than its baseline SQ score.

For each timepoint of interest (12, 24 and 36 months), a subject is identified as having an improving vertebral fracture if at least one thoracic or lumbar vertebra from T4 to L4 is assessed as an improving vertebral fracture at the timepoint of interest. For example, if a vertebra is assessed as having 2, 2, 1, and 1 SQ scores at baseline, 12, 24 and 36 months, respectively, then the vertebra will be identified as having an improving vertebral fracture at the 24- and 36-month timepoints.

6.7.4 Nonvertebral Fracture

Nonvertebral fracture **(other than long bone fractures)**, reported on the Events eCRF and Additional Fractures Details eCRF, is defined as any nonvertebral (excluding **fingers, toes, skull or face**) fracture confirmed by the central imaging vendor based on a copy of radiographs or other diagnostic images such as computerized tomography (CT) or magnetic resonance imaging (MRI), and/or documented in a copy of the radiology report, surgical report, or discharge summary, excluding fractures associated with high trauma severity or pathologic fractures. See Appendix C for the list of fracture locations.

Long-bone Fractures

Long bone fractures are a subset of nonvertebral fractures including the femur, tibia, fibula and the humerus, radius, and ulna bones.

Long bone fractures reported on the Events eCRF, are defined as any long bone fracture confirmed by imaging excluding fractures associated with high trauma severity or pathologic fractures.

Any Historical Fracture

Any fracture recorded on the Subject Fracture History eCRF regardless of trauma severity.

Historical Nonvertebral Fracture

Any nonvertebral fracture recorded on the Subject Fracture History eCRF and not associated with known high trauma severity or pathologic fractures.

6.7.5 Growth Velocity Endpoints

Height-for-age Z-score is defined as the difference between the subject's height and the median height for the population with the same age and gender, divided by the population standard deviation:

$$Height - for - age\ Z - score = \frac{(measured\ value\ -\ median\ of\ reference\ population)}{standard\ deviation\ of\ the\ reference\ population}$$

Growth velocity is given by the change from baseline in height-for-age Z-scores at 12, 24 and 36 months. During normal growth, the change in growth velocity score should equal 0. Growth acceleration is indicated by a positive change, and growth deceleration is indicated by a negative change.

The definitions of growth velocity based on weight and BMI are similarly calculated; weight-for-age and BMI-for-age Z-scores will be calculated for each subject.

To programmatically calculate the Z-scores, the National Center for Health Statistics percentiles growth charts, based on the 2000 Center for Disease Control and Prevention (CDC) (http://www.cdc.gov/growthcharts/cdc_charts.htm), and the CDC Anthropometric Software Package 3.0 Z scores will be used. See details in Appendix D (Code Fragment).

6.7.6 PRO Questionnaires

CHQ-PF-50

The CHQ-PF50 is a 50-item questionnaire to be completed by the parents or guardian of children between 5 and 18 years of age. The 50 questions measure 14 domains which are summarized as the physical and psychological summary scores. Each domain is scored from 0 to 100, with higher score indicating better physical and psychosocial health.

For more information on the contribution of the 50 questions to calculate the CHQ-PF50, see Appendix E.

CHAQ Disability Index Score

The disability domain (questions 1-54) of the CHAQ is used to measure the subject's assessment of physical functioning or the parent's assessment of the child's physical functioning. The disability index comprises 8 categories (dressing and grooming, arising, eating, walking, hygiene, reaching, gripping, and activities). Scoring ranges from 1 to 5; 1 is "without any difficulty," 2 is "with some difficulty," 3 is "with much difficulty," and 4 is "unable to do." An answer of "not applicable" is scored as a 5, but is not counted. If a child requires assistance from another person or uses an aid or other device for any of the 8 categories, the minimum score for that category will be recorded as a 3. The CHAQ questions will then be converted to a 0 to 3 scale score.

Missing values will not be imputed. If any of the domain questions are missing, but the aids/device indicator is non-missing, the domain can still be computed. However, if all domain questions are missing (or are scored as "not applicable") and the aids/device

indicators are missing, then the domain score is considered missing. The CHAQ score is not computed when the subject/parent provides answers in fewer than 6 domains (normally coded as "UNK" in the database).

For more information on the contribution of the 54 questions to calculate the CHAQ, see Appendix E.

WBFPRS

The WBFPRS is validated for use in children 3 years of age or older. At each timepoint, subjects are asked to choose the face that best describes their own pain. See details in Appendix E.

6.7.7 Adverse Events

Treatment-emergent Adverse Events for the 12-month Double-blind Treatment Period

Adverse events with an onset date during the 12-month double-blind treatment period

Treatment-emergent Adverse Events for the 24-month Treatment Period (Double-blind and Open-label)

Adverse events with an onset date during the 24-month treatment period

Treatment-emergent Adverse Events for the 36-month Overall Study Period

Adverse events with an onset date during the 36-month overall study period

Adverse Events for the 12-month Observation Period

Adverse events with an onset date during 12-month observation period

7. Analysis Subsets

7.1 Full Analysis Set

The full analysis set (FAS) is defined according to the intent-to-treat analysis to include all subjects randomized into the study.

7.2 Primary DXA Analysis Set

The primary DXA analysis set includes all subjects in the FAS with baseline and ≥ 1 post-baseline lumbar spine provided by the central imaging vendor during the first 12 months and will only be used to analyze the change from baseline in lumbar spine BMD Z-score as assessed by DXA at 12 months, as the primary analysis for the primary efficacy endpoint. Subjects will be analyzed according to their randomized treatment group.

7.3 Per Protocol DXA Analysis Set

The per protocol DXA analysis set will only be used to analyze the change from baseline in lumbar spine BMD Z-score as assessed by DXA at 12 months, as an additional analysis for the primary efficacy endpoint.

The per protocol DXA analysis set for the primary efficacy endpoint includes subjects who are in the primary DXA analysis set, received the planned doses at day 1 and month 6, and did not significantly deviate from the protocol through month 12. Subjects will be analyzed according to their randomized treatment group.

7.4 DXA Analysis Set

The DXA analysis set for each endpoint of interest (lumbar spine, total hip or femoral neck) includes all subjects in the FAS with baseline and $\geq$ 1 post-baseline valid DXA assessments as provided by the central imaging vendor for the relevant endpoint at or before the timepoint under consideration (6, 12, 18, 24 or 36 months). Subjects will be analyzed according to their randomized treatment group.

7.5 Vertebral Fracture Analysis Set

For each time point of interest (12, 24 and 36 months), the vertebral analysis set includes all subjects in the FAS who have a non-missing baseline and $\geq$ 1 non-missing postbaseline X-ray vertebral evaluation as provided by the central imaging vendor, on or before the time point under consideration. This analysis set will be used to analyze incidence of vertebral fracture.

7.6 Nonvertebral Fracture Analysis Set

The nonvertebral fracture analysis set includes all subjects in the FAS, and will be used to analyze the subject incidence of nonvertebral fracture endpoints. All nonvertebral fractures **(other than long bone fractures)** included in the analysis will be identified or confirmed by the central imaging vendor. **Long bone fractures reported on the Events eCRF and confirmed by imaging will be included in the analysis.**

7.7 PRO Analysis Set

For each PRO (CHQ-PF-50, CHAQ disability index score and WBFPRS), the PRO analysis set includes all subjects in the FAS with baseline and at least 1 postbaseline valid PRO response for the respective PRO at the time point under consideration. Note that this subset could potentially be different from endpoint to endpoint, and from time point to time point, due to missing data.

7.8 Growth Velocity Analysis Set

For each growth velocity endpoint (weight-for-age, height-for-age, and BMI-for-age Z-scores) and time point (12, 24, and 36 months), the analysis set includes all subjects in the FAS who have the relevant data (age in total months, weight, height, and BMI, respectively) at baseline, and on or before the time point of interest.

7.9 Safety Analysis Set

The safety analysis set includes all subjects in the FAS who received ≥ 1 dose of IP. This analysis set will also be used to summarize safety data in the 12-month double-blind treatment period, 24-month treatment period and the entire 36-month overall study period. These subjects will be analyzed according to their actual treatment received during the 12-month double-blind treatment period, where subjects who received ≥ 1 dose of denosumab will be analyzed in the denosumab treatment group regardless of the randomized treatment.

7.9.1 Safety Analysis Set for 12-month Observation Period

The safety analysis set for the 12-month observation period includes all subjects in the FAS who received ≥ 1 dose of IP, completed the 24-month treatment period, and remained in the observation period. These subjects will be analyzed according to their actual treatment received during the 12-month double-blind treatment period, where subjects who received ≥ 1 dose of denosumab will be analyzed in the denosumab treatment group regardless of the randomized treatment.

7.10 PK Analysis Set

The PK analysis set includes all subjects in the FAS who have ≥ 1 serum denosumab reported result.

[redacted]

8. Interim Analysis and Early Stopping Guidelines

8.1.1 Interim Analyses

No interim analysis is planned for the study.

8.1.2 Data Monitoring Committee (DMC)

An external, independent DMC will be used to oversee progress of the study and make recommendations relating to early closure/extension or alteration of the study based on ongoing monitoring of the study data. The DMC will be comprised of members external to Amgen. The DMC members will have access to treatment assignments in order to monitor safety and efficacy results to protect subjects. To minimize the potential introduction of bias, the DMC members will not have any direct contact with the study team, site personnel, or subjects. An independent statistical service provider will generate unblinded reports for review by the DMC. If at any time there are safety concerns, the DMC will communicate the concerns to a representative from Amgen senior management. A charter specifying the DMC functions will be written with the agreement of the DMC members. Last DMC was in March 2022 as all subjects have completed 12-month double blind treatment period of the study.

8.1.3 Primary Analysis

The primary analysis will be conducted at the time of the final analysis.

8.1.4 Final Analysis

The final analysis for the study, including the analysis of the primary efficacy endpoint, will be performed when all enrolled subjects have had the opportunity to complete the 36-month follow-up.

9. Data Screening and Acceptance

9.1 General Principles

The objective of the data screening is to assess the quantity, quality, and statistical characteristics of the data relative to the requirements of the planned analyses.

9.2 Data Handling and Electronic Transfer of Data

The Amgen Global Study Operations-Data Management (GSO-DM) department will provide all data to be used in the planned analyses, except for the imaging and laboratory data. The central imaging vendor will provide DXA scans, and X-ray (lateral thoracic, lumbar spine, knees, molars, cephalogram, and panoramic radiograms or radiographic assessment) data to Amgen cumulatively.

The central laboratory vendor will provide all lab-related data, except for the samples analyzed at local laboratories. All laboratory values will be transferred cumulatively from the central laboratory.

Data screening will be performed periodically during the conduct of the study. The objective of the data screening is to assess the quantity, quality, and statistical characteristics of the data relative to the requirements of the planned analyses. Any questionable values or situations will be reported to GSO-DM for review or confirmation.

As part of the data acceptance procedure, all datasets, planned tables, listings, and graphs will be generated and reviewed to identify any additional data issues. Any critical issues identified must be resolved with GSO-DM before final acceptance of the data.

Cumulative data from the central vendors and Amgen will be transferred via a secure electronic transfer directly to the Independent Biostatistical Group supporting the DMC on a monthly basis initially. The frequency may change later during the study.

This study will use the RAVE database.

9.3 Handling of Missing and Incomplete Data

Subjects may have missing specific data points for a variety of causes. In general, data may be missing due to a subject's early withdrawal from study, a missed visit, or non-evaluability of a specific clinical measurement at its planned clinical visit.

9.3.1 Missing Data

Missing data will not be imputed for this study, **except for the sensitivity analysis for the primary endpoint in which missing data will be imputed based on placebo data using a multiple imputation approach incorporating baseline covariates.**

9.3.2 Incomplete Dates

NA

9.3.2.1 Other Incomplete Dates

The following data will be imputed using the following algorithm:

- Adverse Events
- Concomitant Medications

Imputation Rules for Partial or Missing Start Dates

Start Date		Stop Date						
		Complete: yyyymmdd		Partial: yyyymm		Partial: yyyy		Missing
		< 1st dose	≥ 1st dose	< 1st dose yyyymm	≥ 1st dose yyyymm	< 1st dose yyyy	≥ 1st dose yyyy	
Partial: yyyymm	= 1st dose yyyymm	2	1	n/a	1	n/a	1	1
	≠ 1st dose yyyymm		2	2	2	2	2	2
Partial: yyyy	= 1st dose yyyy	3	1	3	1	n/a	1	1
	≠ 1st dose yyyy		3		3	3	3	3
Missing		4	1	4	1	4	1	1

1=Impute the date of first dose; 2=Impute the first of the month; 3=Impute January 1 of the year; 4=Impute January 1 of the stop year
Note: For subjects who were never treated (first dose date is missing), partial start dates will be set to the first day of the partial month or first day of year if month is also missing.

Imputation Rules for Partial or Missing Stop Dates

Initial imputation

- If the month and year are present, impute the last day of that month.
- If only the year is present, impute December 31 of that year.
- If the stop date is entirely missing, assume the event or medication is ongoing.

If the imputed stop date is before the start date, set stop date to missing.

If the imputed stop date is after the death date, impute as death date.

9.4 Detection of Bias

The study has been designed to minimize potential bias by allocating subjects randomly to treatment groups, assessing endpoints, and handling withdrawals without knowledge of the treatment. Other factors that may bias the results of the study include:

- important protocol deviations (IPDs) likely to impact the analysis and interpretation of the efficacy endpoints
- inadvertent breaking of the blind before formal unblinding
- IP dosing non-compliance
- the timing of and reasons for early withdrawal from treatment and from study

The incidence of these factors will be assessed. IPDs likely to impact the analysis and interpretation of the efficacy endpoints will be listed and/or tabulated. Any breaking of

the blind for individual subjects prior to formal unblinding of the study will be documented. The impact of such unblinding on the results observed will be assessed.

9.5 Outliers

Scatter plots will be examined to identify outliers in continuous variables. Frequencies of the categorical data will be examined to identify questionable values. Outliers resulting from data entry error will be corrected in the database. Outliers that are not due to data entry error will be included in the analysis. The validity of any questionable values or outliers will be confirmed. No valid measurement will be excluded from descriptive or inferential analyses. However, sensitivity analyses may be conducted to evaluate the influence of extreme values in the data. These analyses will be documented in the clinical study report.

9.6 Validation of Statistical Analyses

Programs will be developed and maintained, and output will be verified in accordance with current risk-based quality control procedures.

Tables, figures, and listings will be produced with validated standard macro programs where standard macros can produce the specified outputs.

The production environment for statistical analyses consists of Amgen-supported versions of statistical analysis software; for example, the SAS System version 9.4 or later.

10. Statistical Methods of Analysis

10.1 General Principles

Descriptive statistics will be provided for demographics and subject characteristics, efficacy, PROs, and safety data. Descriptive statistics of continuous measurements will include mean, standard deviation, minimum, 25th percentile, median (50th percentile), 75th percentile, maximum and number of nonmissing observations, while categorical data will be summarized using counts and percentages. Subjects will be analyzed according to their randomized treatment group for efficacy endpoints and actual treatment received for safety endpoints. Where appropriate safety data for the 12-month double-blind period, 24-month treatment period, 36-month overall study period and 12-month observation period will be summarized separately.

Data will be subject to ongoing checks for integrity, completeness and accuracy in accordance with the Data Management Plan with the expectation that all outstanding data issues are resolved ahead of the final lock.

10.2 Subject Accountability

The disposition of all enrolled subjects will be tabulated. Disposition for number of enrolled subjects participating in the PK/[redacted] substudy, disposition for number of randomized subjects, successfully completing IP administration during double-blind period and 24-month treatment period, completing double-blind period and 24-month period and completing the study will be included. The disposition of subjects will also include the number of subjects who withdrew from the IP, their reasons for withdrawal including protocol specified criteria reasons and the number of subjects who withdrew from study and their reasons for withdrawal. IPDs including eligibility violations will be summarized and listed.

10.3 Important Protocol Deviations

IPD categories are defined by the study team before the first subject's visit and updated during the IPD reviews throughout the study prior to database lock. These definitions of IPD categories, sub-category codes, and descriptions will be used during the course of the study. Eligibility deviations are defined in the protocol. Summary table and listing of the deviations from eligibility criteria will be generated.

10.4 Demographic and Baseline Characteristics

Subject demographic and baseline disease characteristics will be summarized based on FAS. They include the following endpoints:

- Race
- Ethnicity
- Age (in years)
- Age (in months)
- Gender
- Body composition (height [cm], weight [kg], and BMI [kg/m^2])
- Growth Velocity (height-for-age, weight-for-age, and BMI-for-age z-scores)
- Armspan
- Baseline BMD Z-score
- Fracture history including any fractures and nonvertebral fractures
- Prevalent vertebral fracture at baseline (based on screening spinal radiograph)
- Number of prevalent vertebral fractures at baseline (0, 1, $\geq$ 2; based on screening spinal radiograph)
- Selected Laboratory assessments

[redacted]

10.5 Efficacy Analyses

Table 1. Endpoint Summary Table

Endpoint	Primary Summary and Analysis Method (Specify Analysis Set if FAS is Not Used)	Sensitivity Analysis
Primary Efficacy Endpoint		
Change from baseline in lumbar spine BMD Z-score as assessed by DXA at 12 months	• Primary DXA Analysis Set: Analysis of covariance (ANCOVA) model including treatment (denosumab vs placebo), baseline age (age at informed consent) and baseline BMD Z-score. Missing baseline and postbaseline BMD Z-scores will not be imputed.	• DXA Analysis Set: Repeated measures model including treatment (denosumab vs placebo), study visit (6 and 12 months), baseline age, baseline BMD Z-score, and treatment-by-visit included as an interaction with no imputation for missing data • **Full Analysis Set:** **An additional sensitivity analysis of the primary efficacy endpoint will be conducted including all randomized and treated subjects, with missing data for the primary endpoint imputed based on placebo data using a control-based pattern multiple imputation approach**

Endpoint	Primary Summary and Analysis Method (Specify Analysis Set if FAS is Not Used)	Sensitivity Analysis
		incorporating baseline covariates . This sensitivity analysis will be performed using an ANCOVA model as described for the primary analysis of the primary endpoint.
Key Secondary Endpoints		
Change from baseline in lumbar spine BMD Z-score as assessed by DXA at 6, 18, 24, and 36 months	• DXA Analysis Set: Repeated measures model as described for the primary efficacy endpoint	
Change from baseline in total hip and femoral neck BMD Z-score as assessed by DXA at 6, 12, 18, 24, and 36 months	• DXA Analysis Set: Repeated measures model as described for the primary efficacy endpoint	
Subject incidence of X-ray confirmed long-bone fractures and new and worsening vertebral fractures during 12, 24, and 36 months	• Full Analysis Set: Summary of the subject incidence of confirmed long-bone fractures and new and worsening vertebral fracture endpoints will	

Endpoint	Primary Summary and Analysis Method (Specify Analysis Set if FAS is Not Used)	Sensitivity Analysis
	be provided for each timepoint of interest	
Subject incidence of improving vertebral fractures at 12, 24, and 36 months compared to baseline	• Vertebral Fracture Analysis Set: Summary of the subject incidence of improving vertebral fracture will be provided for each timepoint of interest	
Subject incidence of new and worsening vertebral and nonvertebral fractures fractures during 12, 24, and 36 months	• Full Analysis Set: Summary of the subject incidence of new and worsening vertebral and nonvertebral fracture endpoints will be provided for each timepoint of interest	
Change from baseline in patient reported outcomes (CHQ-PF-50, CHAQ disability index score, and WBFPRS) at 12, 24, and 36 months	• PRO Analysis Set: Summary statistics without imputation for missing data	
Change from baseline in growth velocity, determined by calculating age-adjusted Z-scores for	• Growth Velocity Analysis Set:	

Endpoint	Primary Summary and Analysis Method (Specify Analysis Set if FAS is Not Used)	Sensitivity Analysis
height, weight, and BMI, at 12, 24, and 36 months	Summary statistics without imputation for missing data	

10.5.1 Analyses of Primary Efficacy Endpoint

10.5.1.1 Primary Analysis

The change from baseline in lumbar spine BMD Z-score at 12 months will be analyzed based on the primary DXA analysis set using an analysis of covariance (ANCOVA) model including treatment (denosumab vs placebo), baseline age (age at informed consent) and baseline BMD Z-score. Missing baseline and postbaseline BMD Z-scores will not be imputed.

The superiority of denosumab compared to placebo for the primary efficacy endpoint will be estimated from the 12-month least-squares (LS) mean of the treatment difference (denosumab – placebo) and the corresponding 95% confidence interval and associated p-value. See Appendix D (Code Fragment).

The above primary analysis will be repeated based on the Per Protocol DXA Analysis Set, as an additional analysis for the primary efficacy endpoint.

10.5.1.2 Sensitivity Analysis

A sensitivity analysis of the primary efficacy endpoint will be conducted in the DXA analysis set, in subjects with baseline and ≥ 1 postbaseline valid DXA assessments, using a repeated measures analysis. Treatment (denosumab vs placebo), study visit (6, and 12, months), baseline age, baseline BMD Z-score, will be included as fixed effects, and treatment-by-visit included as an interaction. Study visit will be treated as a categorical variable. The 12-month least-squares (LS) mean of the treatment difference (denosumab – placebo) and the corresponding 95% confidence interval and associated p-value will be estimated.

Unstructured variance–covariance structure will be used for estimation in the repeated

measures analysis. Other variance-covariance structure may be substituted if convergence problem arises.

An additional sensitivity analysis of the primary efficacy endpoint will be conducted including all randomized and treated subjects, with missing data for the primary endpoint imputed based on placebo data using a control-based pattern multiple imputation approach incorporating baseline covariates. This sensitivity analysis will be performed using an ANCOVA model as described for the primary analysis of the primary endpoint.

10.5.1.3 Covariate Analysis

The ANCOVA model specified for the primary analysis will be explored to include the remaining covariates of interest in Section 4.2.1; summary statistics for the results will include 12-month LS mean point estimate of the treatment difference (denosumab - placebo) and 2-sided 95% confidence interval. No imputation will be performed for missing data.

10.5.2 Analyses of Secondary Efficacy Endpoint(s)

Secondary efficacy endpoints will be assessed without imputation for missing data and without multiplicity adjustment. Due to the lack of a placebo control group after the double-blind treatment period, point estimates of the difference in a given endpoint between the two treatment groups based on their original randomization after M12 cannot be interpreted as the treatment effect of denosumab compared with placebo anymore. Instead, the correct interpretation at different time points after M12 is listed in Table 2 and applicable to the following secondary efficacy endpoints:

- Change from baseline in BMD Z-score as assessed by DXA at 6, 18, 24, and 36 months for each skeletal site (lumbar spine, total hip, and femoral neck).

Table 2. Summary of Description for Interpretation of LS Mean Difference Between Treatment Estimates After 12 Months

Time point	Interpretation of LS mean difference
Months 18, 24	M18: Effect of 18 months of denosumab treatment vs 12 months of placebo followed by 6 months of denosumab treatment. M24: Effect of 24 months of denosumab treatment vs 12 months of placebo followed by 12 months of denosumab treatment.
Months 36	M36: Effect of 24 months of denosumab followed by 12 months of no treatment versus 12 months of placebo followed by 12 months of denosumab treatment followed by 12 months of no treatment

10.5.2.1 Analyses of Other Changes From Baseline in BMD Z-score by DXA

For each skeletal site (lumbar spine, total hip, and femoral neck), the change from baseline in BMD Z-score will be analyzed using repeated measures analysis with original randomized treatment group (denosumab vs placebo), visit (6,12, 18, 24, and 36 months), baseline age and baseline BMD Z-score as fixed effects and treatment by visit included as an interacton term. Visit will be treated as categorical variable. LS mean difference between the two treatment groups (denosumab-placebo) (95% CI) and assocaited p-value will be presented for total hip and femoral neck, and for each visit excluding the 12-month visit for lumbar spine. However, due to the lack of a placebo control group after the double-blind treatment period, all such estimates after month 12 cannot be interpreted as the treatment effect anymore. Instead, the appropriate interpretations of these estimates are listed in Table 2.

All analyses will be based on the skeletal site specific DXA analysis set at 36 months. Missing baseline and post-baseline BMD Z-scores will not be imputed; only subjects with baseline and at least one skeleton-specific DXA assessment visit will be included in the analysis.

10.5.2.2 New and Worsening Vertebral Fractures and Nonvertebral Fractures

Fracture analysis will use the data as mentioned below :

1. Vertebral Fractures : central imaging vendor i.e fractures confirmed by vendor

2. Nonvertebral Fractures other than long bone fracture : fractures reported on the Events eCRF which is confirmed by the central imaging vendor

3. Long bone Fractures : fractures reported on the Events eCRF i.e fractures confirmed by imaging

Summary of the subject incidence of the confirmed long-bone and new and worsening vertebral fractures and new and worsening vertebral and nonvertebral fractures will be provided from Baseline to each timepoint of interest (12, 24, and 36 months).

10.5.2.3 Improved Vertebral Fracture

For each timepoint (12, 24, and 36 months), the subject incidence of improving vertebral fracture will be summarized based on the vertebral fracture analysis set.

10.5.2.4 Analyses of Patient Reported Outcomes

Patient reported outcomes (CHQ-PF-50, CHAQ disability index score, and WBFPRS) and their changes from baseline will be summarized at 12, 24, and 36 months based on their respective PRO analysis set. There is no imputation for missing baseline or postbaseline data. WBFPRS, CHQ-PF-50 physical and psychological summary scores, and CHAQ disability score will be calculated as detailed in the Appendix E.

10.5.2.5 Growth Velocity

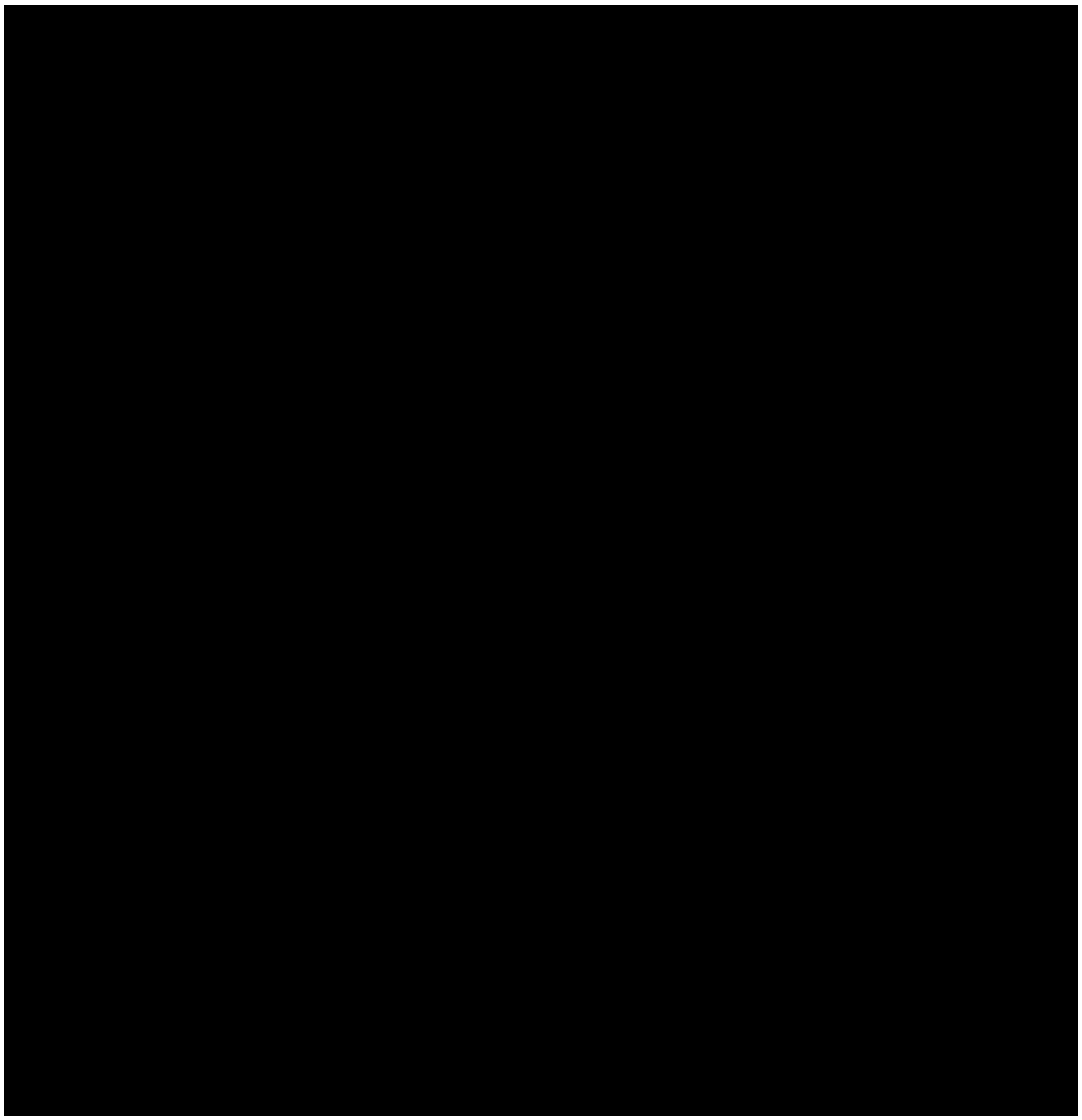
For each time point, descriptive statistics of height-for-age, weight-for-age and BMI-for-age Z-scores, and the respective changes from baseline at 12, 24, and 36 months based on the growth velocity analysis set. There is no imputation for missing data.

10.5.2.6 Denosumab Serum Concentrations

Descriptive statistics of denosumab serum concentration by visit (1, 10 and 30 days, and 3, 6, 12 and 18 months) will be provided based on the PK analysis set.

10.5.3 Analyses of Exploratory Endpoints

[redacted]



10.6 Safety Analyses

Where appropriate safety endpoints will be analyzed separately for the 12-month double-blind period, 24-month treatment period and overall study period (36-month), Adverse events of hypercalcemia will be presented for 12 month observation period (24-36 month) separately.

10.6.1 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) version 26.0 or later will be used to code all events categorized as adverse events (AEs) to a system organ class and a preferred term. Treatment-emergent AEs will be summarized separately for the 12-month double-blind treatment period, 24-month treatment period, 36-month overall

study period according to the event onset date and based on the respective safety analysis set.

Treatment-emergent AEs of Hypercalcemia events will be summarized during 12-month observation period separately under safety analysis set for 12-month observation period.

The subject incidence of AEs will be summarized for all treatment-emergent AEs, serious AEs, AEs leading to withdrawal of investigational product, and fatal AEs by system organ class and preferred term in descending order of frequency**.**

10.6.2 Adverse Events of Interest

Subject incidence of treatment-emergent AEs of interest (EOIs) will also be summarized according to their categories. This tentative list has been based on the EOIs identified for denosumab in the osteoporosis indication, and may be revised and changed later on to better reflect the GiOP pediatric population.

- hypocalcemia
- hypercalcemia
- hypersensitivity
- serious bacterial cellulitis (skin infection)
- osteonecrosis of the jaw (ONJ)

In addition, subject incidence of Hypercalcemia following denosumab during the 12-month observation period will also be summarized.

10.6.3 Laboratory Test Results

Actual values and changes from baseline in each parameter will be descriptively summarized at each visit. For serum calcium, serum albumin-adjusted calcium, phosphorus, and alkaline phosphatase, summary of the percent change from baseline will also be provided. Shifts in laboratory parameters between baseline and the most extreme postbaseline values will be assessed based on the Common Terminology Criteria for Adverse Events v4.0. See Appendix F.

All laboratory analyses will be based on the safety analysis set for the respective period of interest. Subjects with missing data for a scheduled visit will not contribute to the tabulation for that timepoint (no imputation). Visit windows will be used for these summaries as described in Appendix A.

Potential Hy's law cases will be summarized. A listing of liver function tests will be produced for the subjects who meet the criteria for potential Hy's law.

10.6.4 Vital Signs

Descriptive statistics of the actual values and changes from baseline in vital signs (heart rate, respiration rate, temperature) will be presented by visit based on the safety analysis set. Subjects with missing data for a scheduled visit will not contribute to the tabulation for that time point (no imputation).

10.6.5 Antidenosumab Antibody

Subjects receiving at least 1 dose of denosumab will be tested for antidenosumab antibodies. The percentages of subjects who tested positive (binding or neutralizing) for antidenosumab antibodies will be descriptively summarized by visit.

If a subject tests positive for antibodies against denosumab, the relationship between the presence of antibodies, adverse events, bone mineral density, and [redacted] will be evaluated.

Immunogenic response during the study will be described by tabulating the numbers and percentages of subjects who tested positive for (binding and neutralizing) antidenosumab antibodies based on the safety analysis set**.**

10.6.6 Exposure to Investigational Product

Descriptive statistics will be produced to describe the exposure to investigational product by treatment group.

10.6.7 Exposure to Concomitant Medication

During the 12-month observation period, bisphosphonate use will be summarized.

11. Literature Citations / References

Centers for Disease Control and Prevention, National Center for Health Statistics. CDC growth charts: United States. May 2000 http://www.cdc.gov/growthcharts/2000 Centers for Disease Control and Prevention Growth Charts for the United States:Methods and Development:http: //www.cdc.gov/growthcharts/2000growthchart-us.pdf

Rudge S, Hailwood S, Horne A, et al. Effects of once-weekly oral alendronate on bone in children on glucocorticoid treatment. *Rheumatology.* 2005;44:813-818.

Sbrocchi AM, Forget S, Laforte D, et al. Zoledronic acid for the treatment of osteopenia in pediatric Crohn's disease. *Pediatrics International.* 2010;52:754-761.

Simm PJ, Johannesen J, Briody J, et al. Zoledronic acid improves bone mineral density, reduces bone turnover and improves skeletal architecture over 2 years of treatment in children with secondary osteoporosis. *Bone.* 2011;49:939-943.

Ward K, Cowell CT and Little DG. Quantification of metaphyseal modeling in children treated with bisphosphonates. *Bone*. 2005; 36: 999-1002

12. Changes From Protocol-specified Analyses

In response to a FDA request, imputation will be used in an additional sensitivity analysis for the primary efficacy endpoint to use a control based multiple imputation approach based on placebo data incorporating baseline covariates, as described in Section 10.5.1.2.

As described in Section 10.5.2.2, fracture analysis will use the data as mentioned below:

1. Vertebral Fractures : central imaging vendor i.e fractures confirmed by vendor

2. Nonvertebral Fractures other than long bone fracture : fractures reported on the Events eCRF which is confirmed by the central imaging vendor

3. Long bone Fractures : fractures reported on the Events eCRF i.e fractures confirmed by imaging

13. Appendices

Appendix A. Technical Detail and Supplemental Information Regarding Statistical Procedures and Programs

A.1 Analysis Visit Windows

Per protocol, all tests and procedures scheduled on Study Day 1, 10 and 30 will be performed ± 3 days of the scheduled day and ± 7 days for visits at Months 3, 6, 12, 18, 24, 30, and 36. To allow for variations in scheduling, the following sets of analysis visit windows will be used to assign evaluations to a most appropriate nominal visit for analysis. Furthermore, there will be no gaps between consecutive analysis visit windows in order to include as many data points as possible for summarization.

Regardless of the width of the analysis visit window, if more than 1 visit falls within the defined window, the result from the visit closest to the target day will be used in the analysis. If 2 evaluations are of the same distance from the target day, the result from the later visit will be used in the analysis.

Table A.1.1 Lateral Thoracic and Lumbar Spine X-ray,CHQ-PF50, CHAQ Disability Score and WBFPS

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	1	Last evaluation on or before Study Day 1
Month 12	365	Study Day 2 to 548
Month 24	731	Study Day 549 to 913
Month 36	1096	≥ Study Day 914

Table A.1.2 DXA

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline [a]	1	Last evaluation on or before Study Day 1.
Month 6	183	Study Day 2 to 275
Month 12	366	Study Day 276 to 458
Month 18	549	Study Day 459 to 641
Month 24	732	Study Day 642 to 915
Month 36	1098	≥ Study Day 916

[a] If results from baseline DXA are not available, the results from the scan taken on or before Study Day 60 will be considered as baseline values and not the 6-month values.

Table A.1.3 X-ray AP Knees

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	1	Last evaluation on or before Study Day 1
Month 6	183	Study Day 2 to 275
Month 12	366	Study Day 276 to 458
Month 18	549	Study Day 459 to 641
Month 24	732	Study Day 642 to 824
Month 30	915	Study day 825 to 1007
Month 36	1098	≥ Study Day 1008

[redacted]

Table A.1.5 Vital Signs, [redacted] and Serum Chemistry

Nominal Visit	Target Day	Window Definition (Study Day)
Screening	**-38**	**Last evaluation on or before Study Day -1**
Baseline	**1**	**Study day 1 to 5**
Study Day 10	10	Study day 6 to 20
Study Day 30	30	Study day 21 to 40
Month 3	92	Study day 46 to 138
Month 6	183	Study Day 139 to 275
Month 12	366	Study Day 276 to 458
Month 18	549	Study Day 459 to 641
Month 24	732	Study Day 642 to 824
Month 30	915	Study day 825 to 1007
Month 36	1098	≥ Study Day 1008

Table A.1.6 Physical Examination and Haematology

Nominal Visit	Target Day	Window Definition (Study Day)
Screening	**-38**	**Last evaluation on or before Study Day -1**
Baseline	**1**	**Last evaluation on or before Study Day 1**
Month 3	91	Study day 2 to 137
Month 6	183	Study Day 138 to 275
Month 12	366	Study Day 276 to 458
Month 18	549	Study Day 459 to 641
Month 24	732	Study Day 642 to 824
Month 30	915	Study day 825 to 1007
Month 36	1098	≥ Study Day 1008

Table A.1.7 Height and Weight

Nominal Visit	Target Day	Window Definition (Study Day)
Screening	-38	Last evaluation on or before Study Day -1
Baseline	**1**	**Last evaluation on or before Study Day 1**
Month 6	183	Study Day 2 to 275
Month 12	366	Study Day 276 to 458
Month 18	549	Study Day 459 to 641
Month 24	732	Study Day 642 to 824
Month 30	915	Study Day 825 to 1007
Month 36	1098	≥ Study Day 916 1008

Table A.1.8 Pregnancy Test

Nominal Visit	Target Day	Window Definition (Study Day)
Screening[a]	-38	Last evaluation on or before Study Day-1
Baseline	**1**	**On study day 1**
Month 6	183	Study Day 2 to 275
Month 12	366	Study Day 276 to 458
Month 18	549	Study Day 459 to 641
Month 24	732	Study Day 642 to 915
Month 36	1098	≥ Study Day 916

[a] Only for Pregnancy Test

Table A.1.9 Armspan

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	**-38**	**Last evaluation on or before Study Day 1**
Month 12	366	Study Day 2 to 549
Month 24	732	Study Day 550 to 915
Month 36	1098	≥ Study Day 916

Table A.1.10 Antidenosumab Antibody Assay

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	**1**	**On study day 1**
Day 30	30	Study Day 2 to 61
Month 3	92	Study Day 62 to 138
Month 6	183	Study Day 139 to 275
Month 12	366	Study Day 276 to 549
Month 24	732	≥ Study Day 550

Table A.1.11 Tanner Stage

Nominal Visit	Target Day	Window Definition (Study Day)
Baseline	1	Last evaluation on or before Study Day 1
Month 6	183	2 to 275
Month 12	366	276 to 458
Month 18	549	459 to 641
Month 24	732	642 to 824
Month 30	915	825 to 1007
Month 36	1098	≥ Study Day 1008

Appendix B. Age Conversion in Months

Whole-year Age		Partial-year Age (Months)											
Years	Months	0 to < 1	1 to < 2	2 to < 3	3 to < 4	4 to < 5	5 to < 6	6 to < 7	7 to < 8	8 to < 9	9 to < 10	10 to < 11	11 to < 12
5	60	60	61	62	63	64	65	66	67	68	69	70	71
6	72	72	73	74	75	76	77	78	79	80	81	82	83
7	84	84	85	86	87	88	89	90	91	92	93	94	95
8	96	96	97	98	99	100	101	102	103	104	105	106	107
9	108	108	109	110	111	112	113	114	115	116	117	118	119
10	120	120	121	122	123	124	125	126	127	128	129	130	131
11	132	132	133	134	135	136	137	138	139	140	141	142	143
12	144	144	145	146	147	148	149	150	151	152	153	154	155
13	156	156	157	158	159	160	161	162	163	164	165	166	167
14	168	168	169	170	171	172	173	174	175	176	177	178	179
15	180	180	181	182	183	184	185	186	187	188	189	190	191
16	192	192	193	194	195	196	197	198	199	200	201	202	203
17	204	204	205	206	207	208	209	210	211	212	213	214	215

Appendix C. Bone Codes for Nonvertebral Fractures

Table C1 Bone Codes for Nonvertebral Fractures[a]

Code	Name	Code	Name
100	Skull	480	Acetabulum
120	Facial	502	Femur Neck
122	Mandible	503	Femur Intertrochanter
210	Cervical Vertebrae	504	Femur Subtrochanter
240	Sacrum	512	Femur Midshaft
250	Coccyx	513	Femur Distal
310	Ribs	520	Patella
320	Sternum	530	Fibula
401	Clavicle	531	Fibula Proximal
402	Scapula	532	Fibula Shaft
410	Humerus	533	Fibula Distal
411	Humerus Proximal	540	Tibia
412	Humerus Shaft	541	Tibia Proximal
413	Humerus Distal	542	Tibia Shaft
420	Radius	543	Tibia Distal
421	Radius Proximal	550	Metatarsus
422	Radius Shaft	560	Tarsus
423	Radius Distal	570	Toe Phalanges
430	Ulna	620	Ilium
431	Ulna Proximal	630	Ischium
432	Ulna Shaft	640	Pubis
433	Ulna Distal	888	Other (Specify)
450	Carpus		
460	Metacarpus		
470	Finger Phalanges		

[a] From BioClinica Charter for Independent Imaging Assessment

Table C2 Mapping Nonvertebral Fracture Locations[a]

Fracture Location Assessed by BioClinica	eCRF Locations
Acetabulum	Acetabulum Fracture
Carpus	Carpus Fracture
Cervical Vertebrae	Cervical Vertebrae Fracture
Clavicle	Clavicle Fracture
Coccyx	Coccyx Fracture
Femur Distal	Femur Distal Fracture
Femur Intertrochanter	Femur Intertrochanter Fracture
Femur Midshaft	Femur Midshaft Fracture
Femur Neck	Femur Neck Fracture
Femur Subtrochanter	Femur Subtrochanter Fracture
Fibula	Fibula Fracture
Fibula Distal	Fibula Distal Fracture
Fibula Proximal	Fibula Proximal Fracture
Fibula Shaft	Fibula Shaft Fracture
Humerus	Humerus Fracture
Humerus Distal	Humerus Distal Fracture
Humerus Proximal	Humerus Proximal Fracture
Humerus Shaft	Humerus Shaft Fracture
Ilium	Ilium Fracture
Ischium	Ischium Fracture
Mandible	Mandible Fracture
Metacarpus	Metacarpus Fracture
Metatarsus	Metatarsus Fracture
Patella	Patella Fracture
Pubis	Pubis Fracture
Radius	Radius Fracture
Radius Distal	Radius Distal Fracture
Radius Proximal	Radius Proximal Fracture
Radius Shaft	Radius Shaft Fracture
Ribs	Ribs Fracture
Sacrum	Sacrum Fracture
Scapula	Scapula Fracture
Sternum	Sternum Fracture
Tarsus	Tarsus Fracture
Tibia	Tibia Fracture
Tibia Distal	Tibia Distal Fracture
Tibia Proximal	Tibia Proximal Fracture
Tibia Shaft	Tibia Shaft Fracture

.Page 1 of 2

Footnotes defined on last page of the table

Table C2 Mapping Nonvertebral Fracture Locations[a]

Fracture Location Assessed by BioClinica	eCRF Locations
Ulna	Ulna Fracture
Ulna Distal	Ulna Distal Fracture
Ulna Proximal	Ulna Proximal Fracture
Ulna Shaft	Ulna Shaft Fracture

Page 2 of 2

[a] From BioClinica Charter for Independent Imaging Assessment **and Events eCRF. Finger, toe, skull, and facial fractures are not included for nonvertebral fractures in the study.**

Appendix D. Code Fragments

D.1 Analysis of covariance (ANCOVA) Model for the Change from Baseline in BMD Z-score.

D.2 Repeated-measure Analysis for the Change from Baseline BMD Z-score.

D.3 Growth Velocity

D.3.1 Age-adjusted Z-score

The instructions below are given for the calculation of age-adjusted Z-scores for height (also called height-for-age Z-scores). Similar calculations can be performed for weight-for-age and BMI-for-age Z-scores.

Height measurements are normalized for age (given in months) and sex by converting them to Z-score (the number of SD above or below the mean for sex and age) using the National Center for Health Statistics percentiles based on the 2000 Center for Disease Control and Prevention (CDC) growth charts for ages 0 to less than 20 years of age *(http://www.cdc.gov/growthcharts/cdc_charts.htm).* The CDC Anthropometric Software Package 3.0. Z scores are calculating as:

$$Height - for - age\ Z - score = \frac{\left[\left(\frac{height}{M}\right)^{L} - 1\right]}{S * L},$$

where the age-specific M, S and L values are provided by the CDC growth charts.

To calculate height-for-age Z-scores follow the steps below:

Step 1. From site http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm, download the SAS program (cdc-source-code.sas) and the reference data file (CDCref_d.sas7bdat). Do not alter these files, but move them to a folder (directory) that SAS can access.

Step 2. Create a libname statement in your SAS program to point at the folder location of 'CDCref_d.sas7bdat'.

Step 3. Set your existing dataset containing height, weight, sex, age and other variables [ie, subject ID, visit (baseline , 12, 24 and 36 months), cohorts], into a temporary dataset, named *mydata*. Make sure the following variables are renamed and coded as follows:

MYDATA Variables

Variable	Description
agemos	Child's age in months; must be present. The program assumes you know the number of months to the nearest day based on the dates of birth and examination. For example, if a child was born on Oct 1, 2007 and was examined on Nov 15, 2011, the child's age would be 1506 days or 49.48 months, rounded to 1 decimal place (ie 49.5 months). In everyday usage, this age would be stated as 4 years or as 49 months. However, if 49 months were used as the age of all children who were between 49.0 and <50 months in your data, the estimated z-scores would be slightly too high because, on average, these children would be taller, weigh more, and have a higher BMI than children who are exactly 49.0 months of age. This bias would be greater if only completed years of age were known, and the age of all children between 4 and <5 years was represented as 48 months. If age is known only as the completed number of months (as is data from NHANES 1988-1994 and 1999-2010), consider adding 0.5 so that the maximum error would be 15 days. If age is given as the completed number of years, multiply by 12 and consider adding 6.
sex	Coded as 1 for boys and 2 for girls.
height	Height in cm. This is either standing height (for children who are ≥ 24 months of age or recumbent length (for children < 24 months of age); both are input as height. If standing height was measured for some children less than 24 months of age, you should add 0.8 cm to these values (see page 8 of

	http://www.cdc.gov/nchs/data/series/sr_11/sr11_246.pdf). If recumbent length was measured for some children who are ≥ 24 months of age, subtract 0.8 cm.
weight	Weight (kg)
bmi	BMI (Weight (kg) /Height (m)^2). If your data doesn't contain BMI, the program calculates it. If BMI is present in your data, the program will not overwrite it.
headcir	Head circumference (cm)

Z-scores and percentiles for variables that are not in *mydata* will be coded as missing (.) in the output dataset (named *_cdcdata*). Sex (coded as 1 for boys and 2 for girls) and agemos must be in *mydata*. It's unlikely that the SAS code will overwrite other variables in your dataset, but you should avoid having variable names that begin with an underscore, such as _bmi.

Step 4. Copy and paste the following line into your SAS program to call for the SAS macro; location means the full path to where the macro program was copied:
%include '*location*\CDC-source-code.sas'; run;

Step 5. Submit the %include statement. This will create a dataset, named *_cdcdata*, which contains all of your original variables along with z-scores, percentiles, and flags for extreme values. The names and descriptions of these new variables in *_cdcdata* are in the table below. Additional information on the extreme z-scores is given in a separate section that follows the "Example SAS Code".

Z-Scores, Percentiles, and Extreme Values

Description	**Variable**				**Cutoff for Extreme Z-Scores**	
	Percentile	**Z-score**	**Modified Z-score to Identify Extreme Values**	**Flag for Extreme Values**	**Low z-score (Flag coded as -1)**	**High z-score (Flag coded as +1)**
Weight-for-age for children between 0 and 239 (inclusive) months of age	wapct	waz	_Fwaz	_bivwt	< -5	> 5
Height-for-age for children between 0 and 239 (inclusive) months of age.	hapct	haz	_Fhaz	_bivht	< -5	>3
Weight-for-height for children with heights between 45 and 121 cm (this height range	whpct	whz	_Fwhz	_bivwh	< -4	>5

approximately covers ages 0 to 6 y) [a]						
BMI-for-age for children between 24 and 239 months of age	bmipct	bmiz	_Fbmiz	_bivbmi	< -4	>5
Head circumference-for-age for children between 0 and 35 (inclusive) months of age [a]	headcpct	headcz	_Fheadcz	_bivhc	< -5	>5

[a] Not applicable to this study

Step 6. Examine the new dataset, *_cdcdata*, with PROC MEANS or some other procedure to verify that the z-scores and other variables have been created. If a variable in the table was not in your original dataset (eg, head circumference), the output dataset will indicate that all values for the percentiles and z-scores of this variable are missing. If values for other variables are unexpectedly missing, make sure that you've renamed and recoded variables as indicated in the table and that your SAS dataset is named *mydata*. The program should not modify your original data, but will add new variables to your original dataset.

Additional Information

Z-scores are calculated as =

$$Z = \frac{\left(\left(\frac{\text{value}}{M}\right)^{L} - 1\right)}{S*L},$$

in which 'value' is the child's height, BMI, weight, etc. The L, M, and S values are in CDCref_d.sas7bdat and vary according to the child's sex and age or according to the child's sex and height.

Validation of the SAS program (cdc-source-code.sas) will be performed by checking the percentiles associated to a subset of the subjects against the CDC percentiles tables available at the site.

D.3.2 Growth Velocity

Growth velocity at 36 months is then determined by the change from baseline in height-for-age Z-scores (haz). Similar calculations are done for weight-for-age (waz) and BMI-for-age (bmiz) Z-scores.

Appendix E. Patient-reported Outcome Forms/Instruments

E.1 Wong-Baker FACES Pain Rating Scale (WBFPRS)

At each assessment (baseline, 12, 24 and 36 months), subjects are asked to choose a face that best describes their own pain; the corresponding number (0, 2, 4, 6, 8, or 10) is recorded.

This rating scale is recommended for subjects age 3 years and older, therefore only subjects at the age of 3 or older at baseline will be asked to respond.

There is no imputation for missing values.

Change from baseline will be calculates for those subjects with baseline and post-baseline assessments at the timepoint of interest (12, 24 or 36 months).

E.2 Child Health Questionnaire – Parent Form 50

The scales and corresponding items for the CHQ-PF50 are listed below in the order in which they appear in the health survey. Specific instructions for their scoring are given in the Child Health Questionnaire Scoring and Interpretation Manual ©, 2013 HealthActCHQ Inc Boston MA. A summary of those instructions is provided below.

Scale/Item Name (and abbreviation)	Questions	No. of Items	Questionnaire Section
Global Health (GH)	1.1	1	Your Child's Global Health (1)
Physical Functioning (PF)	2.1.a - 2.1.f	6	Your Child's Physical Activities (2)
Roles/Social Limitations-Emotional/Behavioral (REB)	3.1.a - 3.1c	3	Your Child's Everyday Activities (3.1)
Roles/Social Limitations-Physical (RP)	3.2.a - 3.2.b	2	Your Child's Everyday Activities (3.2)
Bodily Pain/Discomfort (BP)	4.1, 4.2	2	Pain (4)
Behavior (BE)	5.1.a - 5.1.e, 5.2	6	Behavior (5)
Mental Health (MH)	6.1.a - 6.1.e	5	Well-Being (6)
Self Esteem (SE)	7.1.a - 7.1.f	6	Self Esteem (7)
General Health Perception (GH)	8.1.a - 8.1.e	5	Your Child's Health (8.**1**)
Change in health (CH)	**8.2**	**1**	**Your Child's Health (8.2)**
Parental Impact – Emotional (PE)	9.1.a - 9.1.c	3	You and Your Family (9.1)
Parental Impact – Time (PT)	9.2.a - 9.2.c	3	You and Your Family (9.2)
Family Activity (FA)	**9.3a – 9.3 f**	**6**	**You and Your Family (9.3)**
Family Cohesion (FC)	**9.4**	**1**	**You and Your Family (9.4)**

Items should be recoded and recalibrated:

- Recoding is necessary to ensure that all items are positively scored so that a higher score indicates better health. Scores vary from 5 to 1.
- Recalibrating is performed for two items (Question 1.1 and 5.2) with an "excellent to poor" response as given below.

Table Recalibration

Response choices	Precoded Value	Final Value
Excellent	1	5.0
Very good	2	4.4
Good	3	3.4
Fair	4	2.2
Poor	5	1.0

- Recalibrating is performed for one item (Question 4.1) with a “none to very severe” response as given below.

Response choices	Precoded Value	Final Value
None	1	6
Very mild	2	5
Mild	3	4
Moderate	4	3
Severe	5	2
Very severe	6	1

- Recalibrating is performed for one item (Question 4.2) with a “none of the time to every/almost every day” response as given below.

Response choices	Precoded Value	Final Value
None of the time	1	6
Once or twice	2	5
A few times	3	4
Fairly often	4	3
Very often	5	2
Every/almost every day	6	1

- Recalibrating is performed for one item (Question 6.1.e) with an “all of the time to none of the time” response as given below.

Response choices	Precoded Value	Final Value
All of the time	1	5
Most of the time	2	4
Some of the time	3	3
A little of the time	4	2
None of the time	5	1

- Recalibrating is performed for six items (Question 7.1.a - 7.1.f) with a “very satisfied to very dissatisfied” response as given below.

Response choices	Precoded Value	Final Value
Very satisfied	1	5
Somewhat satisfied	2	4
Neither satisfied nor dissatisfied	3	3
Somewhat dissatisfied	4	2
Very dissatisfied	5	1

- Recalibrating is performed for two items (Question 8.1.b, 8.1.d) with a “definitely true to definitely false” response as given below.

Response choices	Precoded Value	Final Value
Definitely true	1	5
Mostly true	2	4
Don't know	3	3
Mostly false	4	2
Definitely false	5	1

- **Recalibrating is performed for a item (Question 8.2) with a “Much better now than 1 year ago” to “Much worse now than 1 year ago” response as given below.**

Response choices	Precoded Value	Final Value
Much better now than 1 year ago	**1**	**5**
Somewhat better now than 1 year ago	**2**	**4**
About the same now as 1 year ago	**3**	**3**
Somewhat worse now than 1 year ago	**4**	**2**
Much worse now than 1 year ago	**5**	**1**

- Recalibrating is performed for three items (Question 9.1.a - 9.1.c) with a “none at all to a lot” response as given below.

Response choices	Precoded Value	Final Value
None at all	1	5
A little bit	2	4
Some	3	3
Quite a bit	4	2
A lot	5	1

For subjects that complete half or more of the items in each scale, the mean of the scale will be calculated based on the available responses by adding all non-missing values and divided by the number of non-missing items; otherwise scale is set to missing. Mean is provided with 1 decimal place. After calculating the mean for each subject, it will need to be standardized usinf the formula as: [(Mean -1)/N]*100. The transformed score should be rounded to 0 decimal place.

For each of the 10 scales in Table below, from the manual, each subject's non-missing mean is then standardized by using the Mean and SD of the reference population. For example, the mean for REB is standardized as:

REB_Z = (REB - 90.4013015) / 19.5067502

CHQ-PF50 Scale and Factor Score Coefficients

CHQ-PF50 Scale	No. of Completed Items*	N	Mean	SD	Factor Score Coefficients	
					PhS	PsS
GH	3	4	66.6958379	19.3564297	0. 29460	-0.05547
PF	3	3	90.8525408	16.3826344	0. 37138	-0. 09243
REB	2	3	90.4013015	19.5067502	-0.01178	0. 21155
RP	1	3	91.4951246	18.9079749	0. 34493	-0.06973
BP	1	5	78.6833515	20.7355708	0. 27883	-0. 05514
BE	3	4	72.3086051	17.1447913	-0.12675	0. 27911
MH	3	4	77.2595806	13.6861999	-0.08263	0. 25335
SE	3	4	79.2555314	17.8308361	-0.09480	0. 24792
PE	2	4	73.9788476	21.406013	0. 06063	0. 19823
PT	2	4	83.8816188	20.2901603	0. 09113	0. 16944

* The raw scores can be created by computing the algebraic mean of items when the respondents who completed at least the number of items listed in the table; otherwise both the raw scores and transformed scores should be set to missing.

Values are given with 2 decimal places.

The aggregate CHQ-PF50 physical summary score (PhS) consists of multiplying each CHQ-PF50 scale z-score by its respective physical factor score coefficient (PhS) and summing the ten products. Result is given with 6 decimal places and is called PhS_raw. The formula of calculating aggregate PhS (PhS_raw) is:

PhS_raw = (GH_Z * 0.29460) + (PF_Z * 0.37138) + (REB_Z * -0.01178) + (RP_Z * 0.34493) + (BP_Z * 0.27883) + (BE_Z * -0.12675) + (MH_Z * -0.08263) + (SE_Z * -0.09480) + (PE_Z * 0.06063) + (PT_Z * 0.09113)

And finally the transformed physical (PhS) is calculates as:

$$PhS = 50 + (PhS\_raw * 10)$$

PhS should be rounded to 1 decimal place.

Similarly, the aggregate CHQ-PF50 psychological summary score (PsS) consists of multiplying each CHQ-PF50 scale z-score by its respective psychological factor score coefficient (PsS) and summing the ten products. Result is given with 6 decimal places and is called PsS_raw. The formula of calculating aggregate PsS (PsS_raw) is:

$$PsS\_raw = (GH\_Z * -0.05547) + (PF\_Z * -0.09243) + (REB\_Z * 0.21155) + (RP\_Z * -0.06973) + (BP\_Z * -0.05514) + (BE\_Z * 0.27911) + (MH\_Z * 0.25335) + (SE\_Z * 0.24792) + PE\_Z * 0.19823) + (PT\_Z * 0.16944)$$

And finally the transformed physical (PsS) is calculates as:

$$PsS = 50 + (PsS\_raw * 10)$$

PsS should be rounded to 1 decimal place.

For each subject, the summary scale scores (PhS and PsS) are set to missing if the subject is missing any one of the ten CHQ-PF50. To minimize the number of summary scores missing and for a further comparison with original PhS and PsS, each of the ten scale scores is calculated if half or more of the items are completed.

E.3 Childhood Health Assessment Questionnaire (CHAQ)

The disability domain (questions 1-54) of the CHAQ was used to measure the subject's assessment of physical functioning or the parent's assessment of the child's physical functioning. The disability index comprises eight categories (dressing and grooming, arising, eating, walking, hygiene, reaching, gripping, and activities). Scoring ranged from 1 to 5; 1 was "Without ANY difficulty," 2 was "With SOME difficulty," 3 was "With MUCH difficulty," and 4 was "UNABLE to do". An answer of "not applicable" was scored as a 5, but was not counted. If a child required assistance from another person or used an aid or other device for any of the eight categories, the minimum score for that category was to be recorded as a 3. The CHAQ questions would then need to be converted to a 0 to 3 scaled score.

Afterwards, the CHAQ score using the aids (and/or devices) indicator is computed by taking the maximum score of the questions in each domain (range: 0, 3) and whether or not aids/devices are used (range: 0, 1):

$$A = \max(\text{dressing \& grooming domain questions}, 2*\text{aids indicator}) +$$

+ max(rising domain questions, 2*aids indicator) +

+ max(eating domain questions, 2*aids indicator) +

+max(walking domain questions, 2*aids indicator) +

+ max(hygiene domain questions, 2*aids indicator) +

+ max(reach domain questions, 2*aids indicator) +

+ max(grip domain questions, 2*aids indicator) +

+ max(usual activities domain questions, 2*aids indicator).

CHAQ= A/(total number of sub-domains with at least 6 non-missing)

The CHAQ score is given with 2 decimal places. No missing values are imputed. If any of the domain questions are missing, but the aids/device indicator is non-missing, the domain can still be computed. However, if a domain is missing all domain questions (or is scored as “not applicable”) and is missing aids/device indicators, then the domain score is considered missing. The CHAQ score is not computed when the subject/parent provides answers in fewer than six domains (normally coded as “UNK” in the database).

The following table shows the contribution of the 54 questions used to calculate the CHAQ:

CHAQ Domains and Scores

CHAQ Domains	Domian Components: At least 6 domains must have scores to compute the CHAQ.		CHAQ Component Score Definition:
	Domain Questions	Aids/Devices Questions	
Dressing/Grooming	1, 2, 3, 4	16, 20	First set the Domain Component to missing if the original question = 5 for Scale Range Questions. Then, convert values of each question to a scale of 0-3 (Without ANY difficulty = 0, With SOME difficulty = 1, With MUCH difficulty = 2, UNABLE to do = 3). If Aids/Devices = Yes for Yes/No questions, then set the minimum CHAQ Domain to be equal to 2. Specific terms are mapped to current component for question 19, 53 or 54. Set max of the Domain Questions and Aid/Devices Questions as the component score. Note: Questions 19, 53 and 54 use format to define which category.
Arising	5, 6	18, 21	
Eating	7, 8, 9	17, 22	
Walking	10, 11	12, 13, 14, 15, 23	
Hygiene	24, 25, 26, 27, 28	43, 44, 46, 48, 49	
Reach	29, 30, 31, 32	47, 50	
Grip	33, 34, 35, 36, 37	45, 51	
Activity	38, 39, 40, 41, 42	52	

Appendix F. Reference Values/Toxicity Grades and Lab Assessments

Central Laboratory Chemistry	Central Laboratory Coagulation	Central Hematology	Serology	Other assessments
Sodium	PT/INR[c]	RBC	HIV-1, -2 antibody	Serum denosumab[a]
Potassium		Hemoglobin	HBsAg	Anti denosumab antibody[a]
Chloride		Platelets	Hep C Ab	[redacted]
Bicarbonate		WBC		
Total protein		Differential		
Albumin		Eosinophils		
25(OH) vitamin D		Basophils		
Calcium[a]		Lymphocytes		
Adjusted calcium[a]		Monocytes		
Magnesium		Neutrophils		
Phosphorus[a]				
Glucose				
BUN				
Creatinine				
Total bilirubin				
Alk phos[a]				
AST (SGOT)				
ALT (SGPT)				
GGT[b]				
CPK[b]				

ALT = Alanine aminotransferase; AST = Aspartate aminotransferase; [redacted] BUN = Blood urea nitrogen; CPK = Creatine phosphokinase; DILI = Drug-induced liver injury; GGT = Gamma-glutamyltransferase; HBsAg = Hepatitis B surface antigen; Hep C Ab = Hepatitis C antibody; HIV = Human immunodeficiency virus; INR = International normalized ratio; PT = Prothrombin time; RBC = Red blood cell; [redacted] ULN = Upper limit of normal; WBC = White blood cell.

[a] Results of post-day 1 assessments will be blinded to any study-related personnel (including the sites) except for serum calcium, albumin-adjusted calcium, phosphorus, or ALP in the event of a panic value

[b] To be performed only in subjects with dystrophinopathies at screening and at subsequent visits for the purpose of DILI ascertainment (if applicable).

[c] To be performed at screening only in subjects with dystrophinopathies who have AST or ALT > 5 x ULN and at subsequent visits in any subjects for the purpose of DILI ascertainment.

Assessment of severity for each adverse event and serious adverse event reported during the study will be based on:

The Common Terminology Criteria for Adverse Events (CTCAE), version 4.0 which is available at the following location: https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf

Appendix G. CDC Growth Chart for stature z-scores by sex and age

https://www.cdc.gov/growthcharts/data/zscore/zstatage.xls