Official Title: A Phase 3, Randomized, Double-Blind, Placebo-Controlled, Multicenter Study to Evaluate the Efficacy and Safety of BMN 111 in Children with Achondroplasia

NCT Number: NCT03197766

Applicant/MAH: BioMarin Pharmaceutical Inc.

Version Date: 05 November 2019

16.1.9 Documentation of Statistical Methods

105 Digital Drive

Novato, CA 94949

Statistical Analysis Plan

Study 111-301

A Phase 3, Randomized, Double-Blind, Placebo-Controlled, Multicenter Study to Evaluate the Efficacy and Safety of BMN 111 in Children with Achondroplasia

Version: Final

Date: 05 November 2019

1 SAP SYNOPSIS

TITLE OF STUDY: A Phase 3 Randomized, Double-Blind, Placebo-Controlled, Multicenter Study to Evaluate the Efficacy and Safety of BMN 111 in Children with Achondroplasia
PROTOCOL NUMBER: 111-301
STUDY SITES: Approximately 33 clinical centers worldwide
PHASE OF DEVELOPMENT: Phase 3
OBJECTIVES: The primary objective of the study is to: • Evaluate change from baseline in annualized growth velocity at 52 weeks in subjects treated with BMN 111 compared with control subjects in the placebo group The secondary objectives of the study are to: • Evaluate change from baseline in height Z-score in subjects treated with BMN 111 compared with control subjects in the placebo group at 52 weeks • Evaluate change from baseline in upper:lower body segment ratio in subjects treated with BMN 111 compared with control subjects in the placebo group at 52 weeks • Evaluate change from baseline in body proportion ratios of the extremities • Evaluate effect of BMN 111 on bone morphology/quality by X-ray and dual X-ray absorptiometry (DXA) • Evaluate potential changes in health-related quality of life (HRQoL) as measured by the Quality of Life in Short Stature Youth (QoLISSY) and the Pediatric Quality of Life Inventory (PedsQL) questionnaires • Evaluate potential changes in functional independence as measured by the Functional Independence Measure (Wee-FIM) clinician-reported outcome • Evaluate safety and tolerability of BMN 111 in children with ACH • Evaluate the pharmacokinetics of BMN 111 • Evaluate immunogenicity of BMN 111 and assess impact on safety, PK, and efficacy measures • Evaluate change from baseline in bone metabolism biomarkers The exploratory objectives of the study are to: • Evaluate sleep study scores by polysomnography in a subset of subjects • Evaluate biomarkers of BMN 111 activity • Evaluate genomic biomarkers

STUDY DESIGN AND PLAN: This is a phase 3 randomized, double-blind, placebo-controlled, multicenter study to evaluate the effect of BMN 111 on growth velocity in children with ACH. Subjects who are 5 to < 18 years old, with documented ACH confirmed by genetic testing who have been enrolled in Study 111-901 for at least a 6-month period immediately before study entry, and meet all study eligibility criteria, will participate. Approximately 33 clinical centers worldwide will participate in the study.

At least 110 eligible subjects will be stratified based on sex and Tanner stage (Stage I, and > Stage I, with no more than 20% of subjects > Tanner Stage I) at Screening, and randomly assigned in a 1:1 ratio to one of two treatment groups (placebo or BMN 111 at 15µg/kg).

After subjects have completed the treatment period of the study, those in both treatment and control groups may be eligible to receive BMN 111 in the extension study 111-302 after the Week 52/Study Completion Visit.

ANALYSIS POPULATIONS:

The Full Analysis Set (FAS) includes all randomized subjects with a signed informed consent. The FAS will be used for all efficacy analyses. When applicable, additional sensitivity analyses will be carried out for the Per-Protocol (PP) population, defined as a subset of the FAS who are compliant with the protocol and do not have major protocol violations that affect the interpretability of efficacy data.

The Safety Population is a subset of the FAS, and defined as all randomized consented subjects who receive at least one dose of double-blinded BMN 111 or placebo in this study.

The PK Population is defined as all subjects in the FAS randomized to the BMN 111 treatment group, who receive at least one dose of BMN 111 in this study and have at least one evaluable PK concentration.

The Immunogenicity Population is defined as all subjects in the Safety Population who have at least one evaluable immunogenicity sample.

STUDY ENDPOINTS AND ANALYSES:

Primary efficacy endpoint:

The primary efficacy endpoint is the change from baseline in annualized growth velocity (AGV, cm/yr) at week 52.

The primary analysis of the change from baseline in AGV at week 52 will be performed using an ANCOVA model. Missing values will be imputed, either using multiple imputation (MI) techniques or (if there is insufficient data to perform MI) alternative approaches to address missing data. Details of missing data imputation can be found in Section 14.3.2.

The following sensitivity analysis and supplementary analyses will be performed to assess the robustness of the results of the primary analysis.

- Primary analysis using the washout model (sensitivity).
- Primary analysis on completers (supplementary).
- Primary analysis excluding assessments that occur post limb lengthening or receipt of growth hormone/gonadotropin-releasing hormone (supplementary).

Key secondary efficacy endpoints:

These key secondary efficacy endpoints will be tested using hierarchical testing which controls the type I error rate (otherwise referred to as confirmatory testing):

- Change from baseline in height Z-score at week 52
- Change from baseline in upper to lower body segment ratio at week 52

Other secondary efficacy endpoints:

Non-confirmatory statistical testing will be performed for the secondary endpoints identified below (*). Other endpoints will be summarized using descriptive statistics only.

- Change from baseline in upper arm length to lower arm ratio*
- Change from baseline in upper leg length to knee to heel ratio*
- Change from baseline in upper leg length to tibial length ratio*
- Change from baseline in arm span to standing height ratio*
- Change from baseline in growth measures* (see Section 14.5.2 for list of measures)
- Change from baseline in body mass index
- Change from baseline in body mass index Z-score
- Change from baseline in weight Z-scores
- Change from baseline in pediatric quality of life inventory (PedsQL)
- Change from baseline in quality of life short statured youth (QoLISSY)
- Change from baseline in functional independence measure (Wee-FIM).

Safety endpoints and analyses:

Adverse events (AEs) will be coded using Medical Dictionary for Regulatory Activities (MedDRA) Version 22 and presented by System Organ Class (SOC) and Preferred Term (PT). Summaries of AE's will include all AE's, serious adverse events (SAEs) and events of interest (EOI). Exposure-adjusted summaries will also be provided.

Clinical laboratory tests will be summarized descriptively. Shift tables tabulating Common Terminology Criteria for Adverse Events (CTCAE) severity grade at baseline versus worst post-baseline grade will be provided.

Vital signs will be summarized descriptively. Shift tables tabulating shifts from pre-dose (diastolic BP <40mmHg, diastolic BP ≥40mmHg) to lowest post-injection value (diastolic BP <40mmHg, diastolic BP ≥40mmHg) will be provided. Additional shift tables will describe similar pre/post-injection shifts in vital signs.

X-rays and DXA will be performed on the extremities and spine to evaluate changes in bone morphology and quality. Growth and bone age will be assessed by the means of X-rays of the wrist/hand.

Anti-BMN 111 immunogenicity assessments will be summarized descriptively

Behavioral assessments with the Childhood Behavioral Checklist (CBCL) will be summarized descriptively.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

2 APPROVAL (SIGNATURE AND DATE)

Biostatistician	DocuSigned by: PI Signer Name: PI Signing Reason: I approve this document Signing Time: 14-Nov-2019 | 8:08 AM PST DocuSigned by: E24446DD0CD94CB4BB544985B5001F73 PI	DD MON YYYY
Statistical Programmer	Signer Name: PI Signing Reason: I approve this document Signing Time: 19-Nov-2019 | 3:58 AM PST DocuSigned by: D18269522B56473585DBB6E22270AC44 PI	DD MON YYYY
Clinical Science Lead	Signer Name: PI Signing Reason: I approve this document Signing Time: 14-Nov-2019 | 2:38 AM PST DocuSigned by: E07E8C6298E949C8B26BDADE4053FAEC PI	DD MON YYYY
Pharmacovigilance	Signer Name: PI Signing Reason: I approve this document Signing Time: 14-Nov-2019 | 9:18 AM PST DocuSigned by: C1D6E38155734549A2494D455FF5DD0C PI	DD MON YYYY
Head of Biostatistics	Signer Name: PI Signing Reason: I approve this document Signing Time: 13-Nov-2019 | 12:35 PM PST DocuSigned by: 027DA7AC705E447B9682C8910CF2CB1D PI	DD MON YYYY
Head of Clinical	Signer Name: PI Signing Reason: I have reviewed this document Signing Time: 18-Nov-2019 | 8:10 AM PST CF12A6516C7144E5A6ABD272E4D686EB	DD MON YYYY

3 LIST OF ABBREVIATIONS

Abbreviation	Definition
µg/kg	microgram/kilogram
ACH	achondroplasia
ADA	anti-drug antibody
ADaM	Analysis Data Model
AE	adverse event
AGV	annualized growth velocity
ANCOVA	analysis of covariance
ANP	atrial natriuretic peptide
ATC	Anatomical Therapeutic Chemical
BMC	bone mineral content
BMD	bone mineral density
BMI	body mass index
BMN	BioMarin
BNP	brain natriuretic peptide
BPV	BioMarin Pharmacovigilance
BSAP	bone-specific alkaline phosphatase
CBCL	Child Behavior Checklist
CRF	case report form
cGMP	cyclic guanosine monophosphate
CNP	c-type natriuretic peptide
CNP22	c-type natriuretic peptide (22 amino acids in length)
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events (v4.0)
DMC	data monitoring committee
DXA	dual x-ray absorptiometry
ECG	electrocardiogram
ECHO	echocardiogram
eCRF	electronic case report form
FAS	full analysis set
FDA	Food and Drug Administration
GCP	Good Clinical Practice
ICH	International Conference on Harmonisation
ISR	injection site reaction
IXRS	interactive web or voice response system
LOCF	last observation carried forward
LS	least squares
MAR	missing at random
MCAR	missing completely at random
MCP	multiple comparisons procedure
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MMRM	mixed model for repeated measures
NAb	neutralizing antibody
NCI	National Cancer Institute
NMAR	not missing at random
PD	pharmacodynamic

PedsQL	Pediatric Quality of Life Inventory
PK	pharmacokinetic
PP	per-protocol
PT	preferred term
QoLISSY	Quality of Life in Short Statured Youth
QTc-B	Bazett's corrected QT interval
QTc-F	Fridericia's corrected QT interval
SAE	serious adverse event
SAP	statistical analysis plan
SC	subcutaneous
SD	standard deviation
SDTM	Study Data Tabulation Model
SE	standard error
SMQ	Standardized MedDRA query
SOC	system organ class
TAb	total antibody
TEAE	treatment-emergent adverse event
TESAE	treatment-emergent serious adverse event
TLGs	tables, listings, and graphs
WeeFIM	Functional independence measure for children
WHO	World Health Organization

4 INTRODUCTION

Study 111-301 (original protocol effective date: 01-September-2016, Japanese-specific protocol effective date: 18-April-2018) is a Phase 3 randomized, double-blind, placebo-controlled, multicenter study to evaluate the effect of BMN 111 on growth velocity in children with achondroplasia (ACH).

This Final SAP is an amendment to the Final Draft SAP (dated: 15-May-2019). The SAP has been amended to include modifications to address the FDA written response letter dated 25-July-2019. The key updates are summarized in Section 22.

This SAP is based on Amendment 4 of the protocol (date: 01-February-2019), and describes the analyses and evaluations that will be provided for study 111-301, including the data from study 111-901 data for subjects who subsequently enrolled in study 111-301. Where there are major differences between the protocol-defined analyses, and the SAP-defined analyses, these will be identified in the SAP. The SAP-defined analyses prevail.

4.1 Objectives of Study

The primary objective of the study is to:

- Evaluate change from baseline in annualized growth velocity at 52 weeks in subjects treated with BMN 111 compared with control subjects in the placebo group

The secondary objectives of the study are to:

- Evaluate change from baseline in height Z-score in subjects treated with BMN 111 compared with control subjects in the placebo group at 52 weeks.
- Evaluate change from baseline in upper:lower segment body ratio in subjects treated with BMN 111 compared with control subjects in the placebo group at 52 weeks
- Evaluate change from baseline in body proportion ratios of the extremities
- Evaluate effect of BMN 111 on bone morphology/quality by X-ray and DXA
- Evaluate potential changes in health-related quality of life (HRQoL) as measured by QoLISSY and PedsQL questionnaires.
- Evaluate potential changes in functional independence as measured by the WeeFIM clinical-reported outcome.
- Evaluate safety and tolerability of BMN 111 in children with ACH
- Evaluate the pharmacokinetics of BMN 111
- Evaluate immunogenicity of BMN 111 and assess impact on safety, PK and efficacy measures

- Evaluate change from baseline in bone metabolism biomarkers

The exploratory objectives of the study are to:

- Evaluate sleep study scores by polysomnography in a subset of subjects
- Evaluate biomarkers of BMN 111 activity
- Evaluate genomic biomarkers

4.2 Study Design

Study 111-301 is a Phase 3 randomized, placebo-controlled, double-blind multicenter study to evaluate the effect of BMN 111 on growth in children with ACH. Subjects who are 5 to < 18 years old with documented ACH confirmed by genetic testing will have been enrolled in Study 111-901 (observational run-in study in pediatric patients with ACH) for at least a 6-month period immediately before study entry, and meet all study eligibility criteria.

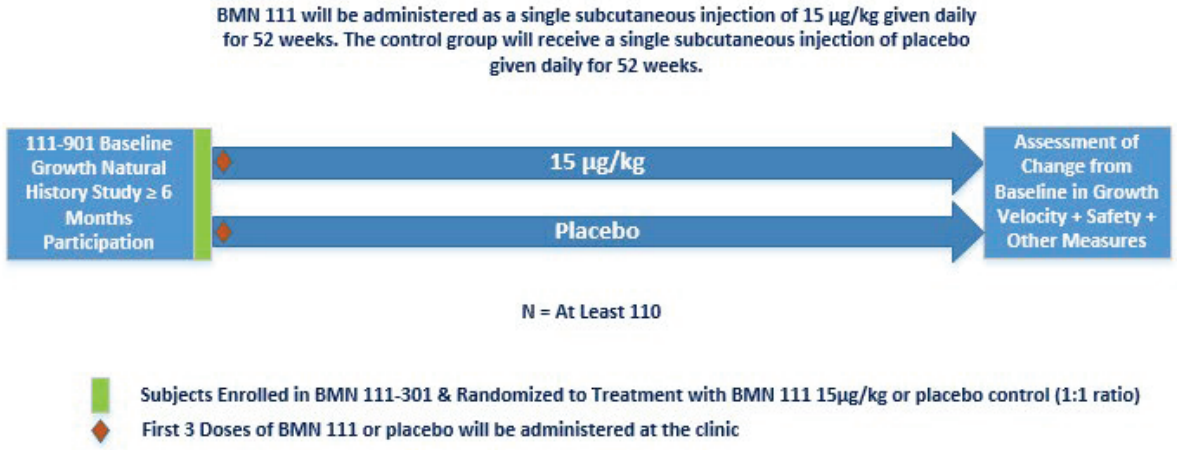
Subjects who discontinue from the study drug will be encouraged to remain in the study and continue to undergo as many of the protocol-specified procedures and assessments as possible for the remainder of the study, as long as such continued participation does not detrimentally affect the health, safety, or welfare of the subject.

This 60-week study (up to 4 weeks of screening, 52 weeks of treatment with an additional 4 weeks of safety follow up) allows for assessment of the effect of daily BMN 111 administration on change from baseline in AGV, height, and body proportions in subjects treated with BMN 111 compared with control subjects in the placebo group. In addition, this study will help to further characterize safety and tolerability of BMN 111 in children with ACH.

At least 110 eligible subjects will be stratified based on sex and Tanner stage (Stage I, and > Stage I, with no more than 20% of subjects > Tanner Stage I) at Screening, and randomly assigned in a 1:1 ratio to one of two treatment groups: placebo or BMN 111 at 15 µg/kg. Approximately 33 clinical centers worldwide will participate in the study.

The 111-301 study design is presented below.

Figure 4.2.1: Study Design



Subjects who complete the 52 week treatment period in 111-301 study on randomized treatment, will be eligible to receive open label BMN 111 15 µg/kg daily in 111-302, to assess safety and efficacy of BMN 111 over the longer term. Subjects will have the option to enroll in 111-302 after the Week 52/Study Completion Visit. Week 52 assessments will serve as Baseline assessments for entry to study 111-302 if the visit occurs on the same day or within 2 weeks. Subjects who enroll into 111-302 more than 2 weeks after the 111-301 Week 52 visit will have a separate Baseline visit; the same assessments as those performed at Week 52 will be repeated at that time.

4.3 Study Population

Subjects who are 5 to < 18 years old with documented ACH confirmed by genetic testing, and have been enrolled in Study 111-901 for at least a 6-month period immediately before study entry, and meet all study eligibility criteria will participate in the study.

4.4 Study Dosage and Administration

BMN 111 will be administered as a single subcutaneous dose of 15 µg/kg given daily for 52 weeks. The control group will receive subcutaneous injections with placebo daily for 52 weeks.

4.5 Sample Size Determination

With 55 subjects planned in each of the two randomized groups (one BMN 111 dose group and one placebo group), the power to detect a difference of 1.75 cm/year between the BMN 111 group and the placebo group in change from baseline in AGV at 12 months is

approximately 90%, assuming that the pooled standard deviation of the change from baseline in AGV is 2.80, using a two-sided two-sample t-test at a 0.05 significance level. The power calculation is based on data from Study 111-202 (a phase 2, open-label, sequential cohort dose-escalation study) and Study 111-901.

4.6 Blinding and Randomization Methods

4.6.1 Blinding Method

This is a placebo-controlled, double-blind study. Blinding is strictly maintained during the study and a study specific masking plan describes how potentially unblinding data (i.e. PK and Biomarker data) are handled during the lifetime of the study. The study will remain blinded until after all data has been entered, and the final database has been locked.

4.6.2 Randomization Method

The centralized randomization with stratification is managed by an external vendor, using IXRS technology. Eligible subjects will be randomized whilst at the site, and receive their first dose of study drug on the same day.

In the original protocol, subjects were stratified based on sex and age (<= 11, > 11 years). At the time of protocol amendment 2, the strata were changed and subjects randomized after this time were stratified by sex and Tanner Stage. The original strata will be reported for all subjects; however all statistical analyses will reported as sex and Tanner Stage.

At least 110 eligible subjects will be stratified based on sex and Tanner Stage (Stage I, and > Stage I) (with no greater than 20% of subjects > Tanner Stage I), and randomly assigned in a 1:1 ratio to one of two treatment groups: placebo or BMN 111 at 15 µg/kg.

A separate stratified randomization scheme is in place for PI subjects. This randomization scheme is set up to assure an overall approximate balance of PI subjects randomized to both treatment arms. The subjects are stratified based on sex and Tanner Stage (Stage I, and > Stage I), and in addition the blocks across the strata are sequenced to assure an overall balance across the treatment arms.

To achieve sex balance, approximately 50% of each sex will be enrolled, with neither to exceed 55%. For placebo and BMN 111, the route of administration is subcutaneous injection and the frequency is daily. Approximately 33 clinical centers worldwide were expected to participate in the study, with a cap of 11 subjects at each site.

4.6.3 Interim Analysis

No formal interim futility/efficacy analyses are planned for this study.

An independent DMC is acting as an advisory body to BioMarin and provides input on the safety data collected in the study. The safety outputs are generated by an external party with an external independent statistician. DMC data review occurs approximately every CBI during the study (or ad hoc, if necessary).

5 GENERAL ANALYSIS CONSIDERATIONS

The analyses described in this SAP are based on Amendment 4 of the protocol (dated 01-February-2019).

Individual subject data that is reflected in data summaries will be provided in PI Unless otherwise specified, all PI will be by treatment group and ordered by subject number, where subject number includes the site number.

Summary tables and figures are provided by treatment group with no overall population column. There are certain exceptions (e.g. subject disposition, demographics, baseline characteristics and safety data) which are indicated in the relevant sections.

Baseline annualized growth velocity (AGV) is defined in Section 14.3.1. Baseline growth measures (other than AGV) and baseline weight are those assessed on Day 1 of dosing, regardless of timing relative to dosing. (Note: If no assessments are available, then the baseline assessment will be the last non-missing assessment prior to Day 1). All other baseline assessments are the last non-missing on-study 301 assessments prior to administration of study drug. Change from baseline is calculated as: [post-baseline value – baseline value]; Percent change from baseline is calculated as: ([post-baseline value – baseline value]/baseline value) x 100.

The Screening visit will not be presented in descriptive summaries over time. When the analysis windows have been applied (see Section 5.4), data collected at the screening visit may be included in the Baseline assessment. All assessments will be listed, but only those assessments assigned to visits according to the windows applied (see Section 5.4) will be summarized in tables or graphically. Exceptions to this are summaries of the worst post-baseline laboratory data in shift tables where all assessments are considered, and also the standing height assessments displayed in the spaghetti plots.

Unless otherwise specified, the number of subjects, mean, 25th percentile, median, 75th percentile and estimates of precision (e.g. SD, confidence interval) will be displayed to one more decimal place than the collected data. Minimum and maximum values will be presented to the same decimal places as the collected data. For summaries by visit, the number of subjects with non-missing observations (n) according to the analysis windows (see Section 5.4 for further details) at the derived visit will be presented.

When summarizing categorical data, for mutually exclusive categories, the sum of the individual categories, including a separate “Missing” category, should equal 100%. Unless otherwise specified, if categories have no data, then they will not be displayed.

All statistical analyses will be performed using SAS® version 9.4 or a later release (SAS Institute, North Carolina, USA) and results be presented in the form of tables, PI and graphs (TLGs). All statistical analyses, including dataset creation and output generation, will be performed by Biostatistics and Statistical Programming personnel of BioMarin's Biometrics department.

5.1 Analysis Populations

A total of 5 analysis populations will be used for all summaries and analyses. Subjects who have satisfied the population criteria will be classified in the designated population and analyzed accordingly.

The analysis populations are described below. For each population, the number and percent of subjects included, excluded, and the reason for exclusion (where applicable) will be summarized by treatment group and overall. PI will be provided to identify any subjects not included in the population, and the reason for exclusion.

The number of subjects in the Full Analysis Set (FAS) will also be summarized by randomization strata (i.e., Male Tanner Stage I, Female Tanner Stage I, Male Tanner Stage >I, Female Tanner Stage >I).

5.1.1 Full Analysis Set

The Full Analysis Set (FAS) is defined according to the intention-to-treat principle and includes all randomized subjects with a signed informed consent. The FAS will be used to present the baseline characteristics, and efficacy data by randomized treatment group. All listings, excluding screen failures, will be produced on the FAS.

5.1.2 Per-Protocol

The Per-Protocol (PP) population will be determined consistent with the International Conference on Harmonisation (ICH) E9 Guideline (http://ww.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002928.pdf) and is defined as a subset of the FAS population who are compliant with the protocol.

The PP population will comprise of all subjects in the FAS who satisfy the following criteria:

1) Subjects who met all inclusion/exclusion criteria.
2) Subjects who had no major protocol deviations that could affect the results of the primary or key secondary endpoints (AGV, height Z-score, and upper to lower body segment ratio).**

3) Subjects who were assigned into the correct strata.
4) Subjects who completed the treatment originally allocated (i.e., always received the assigned treatment (BMN 111 or placebo), had a Week 52 visit within the defined visit window (see Section 5.4), and were at least 90% compliant with study drug).

Subjects who always received the assigned treatment (4) can only be determined once the study has been unblinded.

**The major protocol deviations that would exclude a subject from the PP population are:

- Evidence of taking growth hormone (coded as "somatropin and somatropin agonists") or gonadotropin-releasing hormone (coded as "gonadotropin releasing hormone analogues")
- Evidence of having a limb lengthening procedure (i.e., a MedDRA preferred term of 'limb operation').

The PP population will be used for the analysis of the primary and two key secondary endpoints.

5.1.3 Safety

The Safety Population is defined as all subjects in the FAS who receive at least one dose of double-blind BMN 111 or placebo in this study.

The Safety Population will be used to present the safety summaries by actual treatment received. Subjects randomized to one treatment group, who only receive the study drug for the other treatment group, will be analyzed in the arm of the study drug they received.

5.1.4 PK

The PK Population is defined as all subjects in the FAS randomized to the BMN 111 treatment group, who receive at least one dose of BMN 111 in this study and have at least one evaluable PK concentration.

5.1.5 Immunogenicity

The Immunogenicity Population is defined as all subjects in the Safety Population who have at least one evaluable immunogenicity sample. Subjects randomized to one treatment group, who only receive the study drug for the other treatment group, will be analyzed in the arm of the study drug they received.

5.2 Pooling of Data from Sites with Small Enrollment

Not applicable, as there are no analyses by site.

5.3 Study Day Derivation

Study day in 111-301 will be determined by subtracting the initial study drug start date from a visit date plus 1 if the visit date occurs on or after the initial study drug start date. Otherwise, the study day will be the visit date minus the initial study drug start date. Therefore, Study Day 1 will occur on the initial study drug start date. i.e.

- Study Day 1 = Day of first dose of study drug.
- If Visit Date < Study Drug Start Date, then Study Day = (Visit Date - Study Drug Start Date)
- If Visit Date > Study Drug Start Date, then Study Day = (Visit Date - Study Drug Start Date) +1

5.4 Visit Windows for Analysis

An assessment for a subject will be classified according to the study day of the assessment where it falls within a window.

Data in Study 111-901 for subjects who enrolled in Study 111-301 and are included in the corresponding analysis population will be used in efficacy summary tables and figures. The analysis visits and windows for assessments in Study 111-901 are defined in Table 5.4.1. Unless otherwise specified, only the derived visits listed in the table will be included in summary tables and figures.

Table 5.4.1: Visit Windows for Observed Assessments in BMN 111-901

Derived Visit	Target Day[a]	Analysis Window[a]
≥ 6 Months Prior	-182	> -364 to ≤ -182
≥ 12 Months Prior	-364	> -546 to ≤ -364

[a] Target days and analysis windows are relative to the 111-301 Day 1 date defined in Section 5.3 above.

For each subject and growth measure in Study 111-901, the record closest to the target day within the analysis window will be retained for inclusion in summary tables.

Spaghetti plots of standing height and standing height Z-scores will include all measures recorded in Study 111-901 up to and including the "≥ 12 Months Prior" assessment without applying windows. This will allow for all subjects with height assessments ≥ 12 months prior to Day 1 to have at least one assessment included.

The analysis windows for assessments in Study 111-301 are defined in Table 5.4.2. The analysis windows are designated for each scheduled visit and centered on a target day. If

there are two or more assessments within a designated window, the assessment that is closest to the target day will be used for analyses. If the two closest assessments to the target day are equidistant from the target day, the assessments taken on a scheduled visit per CRF page will be used. Unless otherwise specified, all analyses are based on derived visits.

Table 5.4.2: Visit Windows for Observed Assessments in 111-301

Derived Visit	Target Day[a]	Analysis Window[a]
Baseline		The last assessment prior to the first dose of study drug unless otherwise specified.
Day 1	Day 1	Day 1
Day 2	Day 2	Day 2
Day 3	Day 3	Day 3
Day 10	Day 10	Day 4 – 16 (-/+6)
Week 6	Day 43	Day 22 - 64 (-/+21)
Week 13	Day 92	Day 71 – 113 (-/+21)
Week 26	Day 183	Day 162 - 204 (-/+21)
Week 39	Day 274	Day 253 – 295 (-/+21)
Week 52	Day 365	Day 344 – 386 (-/+21)

[a] Target days and analysis windows are relative to the 111-301 Day 1 date defined in Section 5.3.

5.5 Handling of Dropouts and Missing Data

PI . Please refer to Section 14.3 for the proposed methodology for dealing with missing data in the efficacy analyses of the primary and two key secondary endpoints. PI .

5.5.1 Partial Dates

Where start and stop dates for medications or adverse events are incomplete, imputation rules will be applied.

In the event that the start date of a medication/AE is partial, the following imputation rules will be applied:

- If only day is missing, then the start date will be imputed as the first day of the month. If month and year are the same as the month and year of first dose of study drug then the start date will be imputed as the first dose date.

- If only year is non-missing, then the start date will be imputed as the first day of the year. If year is the same as the year of first dose of study drug then the start date will be imputed as the first dose date.

In the event that the stop date of a medication/AE is partial, the following imputation rules will be applied:

- If only day is missing, then the end date will be imputed as the last day of the month.
- If only year is non-missing, then the end date will be imputed as the last day of the year. If the imputed date is beyond the study completed or discontinued date, the imputed date will be replaced with the study completed or discontinued date.

6 SUBJECT DISPOSITION

At least 110 subjects (55 per treatment group) are planned to be enrolled by approximately 33 sites.

Disposition will be presented based on the FAS. The number of subjects in the FAS, the number of subjects who received study drug, the number of subjects who completed treatment on study drug, and the number of subjects who completed the study will be reported by treatment group.

The total number of subjects who were in the FAS, received study drug, and completed the study (i.e., completed the Week 52 visit) or withdrew prematurely will be summarized by treatment group and depicted by a flowchart in the clinical study report (CSR).

Disposition data will be listed by treatment group, and will indicate the investigator site. In addition, for those subjects who were screen failures, the reason for screen failure will be listed.

7 DISCONTINUATION AND COMPLETION

For subjects who prematurely discontinued from the study, the primary reason for discontinuation will be summarized. A similar summary will be provided for subjects who discontinued study drug.

The number of subjects who completed study drug, and the number of subjects who completed the study will be summarized by treatment group.

All summaries will be based on the FAS, and presented by treatment group and overall.

8 PROTOCOL EXEMPTIONS AND DEVIATIONS

The number and percentage of subjects with any protocol deviation, any major protocol deviation, and any minor protocol deviation will be presented, according to the categories described in the study specific protocol deviation plan. Reasons for major/minor deviations will be summarized.

All summaries will be based on the FAS, and presented by treatment group and overall.

All protocol deviations will be listed by treatment group, and ordered by subject. A separate listing of all major deviations which have been classified as a "Dosing Irregularity" will be provided.

9 DEMOGRAPHICS AND BASELINE CHARACTERISTICS

Demographic and baseline characteristics will be summarized for the FAS, and presented by treatment group and overall.

Demographics will be summarized as:

- Age at Day 1 (years)
- Age at Day 1 (≥ 5 to < 8, ≥ 8 to < 11, ≥ 11 to < 15, ≥ 15 to < 18 years)
- Sex (Male, Female)
- Race (American Indian or Alaska Native, Asian [Japanese/Other], Black or African American, Native Hawaiian or Pacific Islander, White, Multiple, Not Provided Due to Patient Privacy Rules)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)

Baseline characteristics will be summarized as:

- Tanner Stage (I, > I; Derived as either genitalia (male) or breast (female) Tanner stage)
- Weight (kg)
- Weight Z-Score
- Body Mass Index (kg/m^2)
- Body Mass Index Z-Score

Baseline anthropometric growth measures will be summarized as:

- Annualized Growth Velocity (AGV) (cm/yr)
- Height Z-Score
- Standing Height (cm)
- Sitting Height (cm)
- Head Circumference (cm)
- Arm Span (cm)
- Arm Span to Standing Height Ratio
- Upper Arm Length (cm)
- Lower Arm (Forearm) Length (cm)
- Upper Arm Length to Lower Arm (Forearm) Length Ratio

- Lower Body Length (cm) = Standing Height – Sitting Height
- Upper Leg Length (Thigh) (cm)
- Knee to Heel Length (cm)
- Tibial Length (cm)
- Upper to Lower Body Segment Ratio
- Upper Leg Length (Thigh) to Knee to Heel Length Ratio
- Upper Leg Length (Thigh) to Tibial Length Ratio

In addition to the demographics at baseline, subject's Tanner Stage status will be summarized [n(%)] by visit and presented by treatment group and overall:

- Tanner Stage (I, > I; Derived as either genitalia (male) or breast (female) Tanner stage)
- Pubic hair Tanner Stage (I, II, III, IV, V, Not Done)
- Breast Tanner Stage (Female only) (I, II, III, IV, V, Not Done)
- Genitalia Tanner Stage (Male only) (I, II, III, IV, V, Not Done)

10 MEDICAL HISTORY

Medical history will be solicited from each subject, including all major illnesses, diagnoses, and surgeries that the subject has ever had; any prior or existing medical conditions that might interfere with study participation or safety; and evaluation for knee, thigh, hip or groin pain, or change in gait/activity. All medical history terms will be coded in accordance with MedDRA Version 22 so that each term is assigned a system organ class (SOC) and preferred term (PT).

A predefined list of the achondroplasia-related PTs is provided in a separate project-level document. Prior to the database lock and unblinding of this study, any PTs which have not previously been reported as achondroplasia-related will be identified via medical review, and will be added to this list. The complete list of final terms will be retained in the study file, and documented in the CSR.

All recorded medical history will be summarized on the FAS by system organ class (SOC) and by preferred term (PT), ordered by descending order of frequency of SOC in the overall population. In addition, achondroplasia-related medical history will be summarized separately by SOC and PT in a similar manner.

The number of subjects with abnormal physical exam results at screening will be summarized and listed.

PI [redacted].

11 PRIOR AND CONCOMITANT MEDICATIONS

Prior and concomitant medications will be summarized for the FAS. For analysis purposes, the following definitions will be used to determine prior and concomitant medications for the entire study period:

- Prior medications: Any medications taken and ended prior to the initial study drug administration date will be considered prior medications.
- Concomitant medications: Any medications taken on or after the initial study drug administration date up to 30 days after discontinuation of study treatment will be considered concomitant medications and included in summary tables. This also includes medications initially taken prior to the initial study drug administration date but continued or ended on or after the initial study drug administration date.

All medications will be included in PI

Concomitant medications will be summarized for the overall study period, and a separate summary will be provided which includes only those concomitant medications that were initiated on study.

All medications will be coded using the latest version available within BioMarin of the World Health Organization Drug (WHO Drug) dictionary (Version March 2019). Prior and concomitant medication use will be separately summarized by Anatomical Therapeutic Chemical (ATC) medication class (Level 2) and preferred name (i.e., generic medication name). If a medication doesn't have ATC level 2, they are grouped as "ATC Level 2 classification unavailable". A subject reporting the same medication more than once will be counted once when calculating the number and percentage of subjects who received that medication. PI .

12 COMPLIANCE

Treatment compliance will be calculated based on the study drug preparation and study drug administration records captured in the eCRF, and summarized by treatment group for the safety population.

For each daily injection, the amount of study drug taken will be calculated as the number of units in the syringe pre-injection, minus the number of units remaining in the syringe post-injection. If the number of units in the syringe pre-injection is not captured in the eCRF, then it will be assumed to be the same as the planned dose.

The total amount of study drug taken will be calculated as the amount of study drug taken (units) in the corresponding treatment duration.

The total amount of study drug planned is determined based on the duration of treatment, and the planned units per day.

Percentage compliance for each subject will be derived from the total amount of study drug intake divided by the planned study drug intake, and multiplied by 100:

$$\text{Treatment Compliance (\%)} = \frac{Total\ Amount\ of\ Study\ Drug\ Taken\ (units)}{Total\ Amount\ of\ Study\ Drug\ Planned\ (unitS)} \text{ x } 100$$

Treatment compliance will be calculated and summarized by treatment group.

- Compliance with protocol-specified treatment regimen (%)
- Compliance with protocol-specified treatment regimen (≥50%, ≥60%, ≥70%, ≥80%, ≥90%, ≥100%)

13 EXTENT OF EXPOSURE TO STUDY DRUG

Dosing information is recorded by subjects (in the paper diary) and also by the site on visit days. In the case of duplicate records with the same date, if the investigator confirms the date, then investigator reported dosing data will be used. All summaries will be based on the safety population, and presented by treatment group.

The day of the first dose of study drug is referred to as Day 1. A missed dose is defined as a calendar day without a dose between the date of Day 1 and date of last dose.

Descriptive statistics will be provided for the following variables:

- Duration of treatment (days)

  = Date of last dose – Date of first dose + 1

- Total number of doses administered

  = Number of doses administered between date of Day 1 and date of last dose

- Total number of doses missed

  = Number of doses missed between date of Day 1 and date of last dose

- Total amount of weight adjusted dose administered (ug/kg)
- Average weight-adjusted daily dose administered (ug/kg/day)

The frequency and percent of subjects will be provided for the following variables, where missed doses are counted between the date of first dose (Day 1) and date of last dose:

- Number of subjects who missed at least one dose
- Number of subjects who missed more than one dose
- Number of subjects who missed more than 5 consecutive doses
- Number of subjects who missed more than 10 consecutive doses

The frequency and percent of all missed doses will be provided for the following variable:

- Reason for missed dose (Site Error, Parent/Caregiver Error, Home Health Error, Adverse Event, Other, Missing).

PI .

14 EFFICACY EVALUATIONS

14.1 General Approaches for Analyzing Growth Parameters

All growth parameters assessed by anthropometric measurements are measured at Day 1, Week 13, Week 26, Week 39 and Week 52.

Growth parameters are measured 3 times for each assessment. It is the mean of these 3 assessments that are considered for the summaries and analyses. In the event that all 3 are not available, the mean of the 2, or the individual assessment is taken.

All efficacy endpoints will be assessed using the FAS. In addition, the primary and key secondary endpoints (AGV, height Z-score, and upper to lower body segment ratio) will also be assessed using the PP population.

Estimand formulations are provided for the primary and key secondary endpoints.

Visit windows are applied for all assessments (see Section 5.4) and are used to summarize the growth measures by visit.

Confirmatory statistical testing controlling for the Type I error rate is only performed on the primary analyses for the primary and two key secondary endpoints. Beyond these analyses, there is no control for multiplicity, and p-values from additional analyses on these endpoints or testing of other endpoints are non-confirmatory.

Unless otherwise specified, all growth parameter assessments are considered for inclusion in analyses, including those collected after treatment discontinuation. Therefore, with respect to the intercurrent event of "rescue therapy received", the assessments will be included for subjects who have received growth hormone or gonadotropin-releasing hormone or had limb lengthening; however the impact of these intercurrent events will be explored in supplementary analyses censoring the data prior to the interventions.

With the exception of AGV (see Section 14.3.1), summary tables for all growth measures including height Z-score, upper to lower body ratio, body proportion ratios, growth measures (see Section 14.5.2), body mass index (BMI), BMI Z-scores and weight Z-scores will include assessments from both the 111-901 study and 111-301 studies at the following time points:

- - 12 Months (111-901)
- - 6 Months (111-901)
- Baseline (111-301)
- Week 26 (111-301)

- Week 52 (111-301).

Each table will include summaries of the absolute measures at each of these time points, and change from baseline at Week 26 and Week 52 (applying visit windows). The tables will summarize data by treatment group. These summaries are on continuous data and will include the number of subjects with assessments, mean, SD, median, 25th and 75th percentile, minimum and maximum.

Results of the statistical analyses will be provided in separate tables, including the least-squares (LS) mean change from baseline at Week 52 for each treatment group, the treatment difference in LS means (calculated as BMN 111 – Placebo), the 95% confidence interval (CI) for the treatment difference, and corresponding 2-sided p-value.

ANCOVA models will be used to determine the treatment difference at 52 weeks. Unless otherwise specified, all models include the following baseline covariates:

- Treatment group
- Stratum: (Male Tanner Stage I, Female Tanner Stage I, Male Tanner Stage >I, Female Tanner Stage >I)
- Age at baseline
- AGV at baseline
- Height Z-score at baseline

Note: If a subject is incorrectly assigned to a stratum at the time of randomization, they will be analyzed according to their actual sex and Tanner stage.

Subgroup analyses will only be performed for the primary analyses of the primary and key secondary endpoints (see Section 14.7).

PI will include the mean values used in analyses and summaries, and will be ordered by treatment group, subject and visit date. An additional PI of standing height will include all assessments (i.e., individual measures and means of the measures).

14.2 Statistical Testing

The overall type I family-wise error rate for testing the primary and key secondary efficacy endpoints will be controlled at the 0.05 significance level using the following 3-step serial gatekeeping multiple comparisons procedure (MCP). Following this MCP, advancement to the next step will only occur if the null hypotheses within a step and the previous step(s) are all rejected at the significance level of 0.05 in favor of BMN 111. If any null hypothesis within a step is not rejected or is rejected but not in favor of BMN 111, the hypothesis tests

corresponding to all subsequent steps will not be considered confirmatory. All hypothesis tests will be two-sided.

1. The first step will be the test comparing the BMN 111 group to the placebo group for the primary efficacy endpoint, the change from baseline in AGV at Week 52 time point. If the null hypothesis is not rejected (i.e., p-value > 0.05) or is rejected but not in favor of BMN 111, all subsequent statistical tests will not be considered confirmatory. The study is considered positive if the primary test is significant in favor of BMN 111.
2. The second step will be the test comparing the BMN 111 group to the placebo group for the secondary efficacy endpoint, the change from baseline in height Z-score at Week 52 time point. If the null hypothesis is not rejected (i.e., p-value > 0.05) or is rejected but not in favor of BMN 111, the subsequent statistical test will not be considered confirmatory.
3. The third step will be the tests comparing the BMN 111 group to the placebo group for the secondary efficacy endpoint, the change from baseline in upper:lower body segment ratio at Week 52 time point at a significance level of 0.05.

14.3 Primary Efficacy Endpoint

The primary efficacy objective is to evaluate the change from baseline in annualized growth velocity (AGV) at 52 weeks in subjects treated with BMN 111 compared with control subjects in the placebo group.

14.3.1 Annualized Growth Velocity

The primary efficacy endpoint is the change from baseline in AGV at the Week 52 (12-month) time point. For a given interval [Date1, Date2], the AGV is defined as follows:

$$\text{AGV (cm/yr)} = \frac{Standing\ Height\ at\ Date\ 2 - Standing\ Height\ at\ Date\ 1}{Interval\ Length\ (Days)} \times 365.25$$

where the interval length in days is calculated as Date2 - Date1.

AGV (cm/yr) will be calculated for the following visits:

- Baseline: [Date of last height measurement in study 901 at least 6 months prior to Date of Day 1 visit in study 111-301, Date of Day 1]
- Week 13: [Date of Day 1, Date of Week 13]
- Week 26: [Date of Day 1, Date of Week 26]
- Week 39: [Date of Day 1, Date of Week 39]
- Week 52 (12-month): [Date of Day 1, Date of Week 52]

The baseline AGV is established in the observational run-in Study 111-901, based on the standing height measurements at least 6 months prior to enrollment to Study 111-301.

The absolute values for AGV at Baseline, Week 26 and Week 52, and change from baseline at Week 26 and Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group.

A box and whisker plot will be provided for AGV (cm/yr) over time (Baseline, Week 26, and Week 52) for each treatment group.

In addition, for the visits at Baseline, Week 26, and Week 52, scatter plots will be provided by sex and treatment group for (1) AGV by age at visit and (2) AGV by Tanner stage (I, II, III, IV, V) (using breast Tanner Stage for females, and genitalia Tanner Stage for males) at visit.

AGV listings will include AGV for all height assessments at all scheduled visits.

14.3.2 Primary Analysis Method

Some aspects of the primary and other analyses for AGV have been revised relative to the protocol, based on interactions with health authorities and blinded data review.

The primary analysis method described below is based on the protocol, but adds criteria necessary for the multiple imputation to be applicable. In case it cannot be applied, an alternative approach for imputing missing assessments is described. Revisions have also been made to the sensitivity analyses defined in the protocol (see Section 14.3.3).

The following primary hypothesis will be tested (two-tailed):

> *H0: Difference in mean AGV change from baseline at week 52 between BMN 111 group and the placebo group = 0*
>
> *Ha: Difference in mean AGV change from baseline at week 52 between BMN 111 group and the placebo group ≠ 0*

The estimand formulation is as follows:

- **Population:** FAS
- **Variable:** Change from baseline in AGV at 52 weeks
- **Intercurrent event:** Regardless of whether or not switching to rescue medication had occurred or subjects had discontinued from the study.
- **Population-level summary:** Difference in estimated means between the treatment groups

The estimators for each of the analyses are defined in the relevant sections.

Subjects who discontinue from study drug will be encouraged to remain in the study and their non-missing height measurements will continue to be used in calculating the AGV. In the event that missing data does occur despite all efforts, missing standing height will be imputed based on multiple imputation (MI) with pattern-mixture models (Little 1993; Molenberghs and Kenward 2007, pp. 30, 34-37) implemented in PROC MI of SAS where the missing, unobserved observations of the standing height are assumed to follow a missing not at random (MNAR) mechanism. Each imputation will require at least 5 subjects in the same treatment group who also discontinue treatment prematurely at a similar time point (i.e. no more than a 90 day difference between the last dose dates) and have standing height assessments post treatment discontinuation. This is referred to as off-treatment data, and the data for these subjects with standing height assessments post treatment discontinuation will be used as the reference data in the MNAR statement within PROC MI.

PROC MI will be used to impute missing standing height assessments at Week 52. A regression method will be applied with the fully conditional specification (FCS) which can handle both monotone and non-monotone missing standing height data patterns. The baseline covariates for age, strata, height Z-score and AGV and standing height will be included in the VAR statement along with the post-baseline standing height assessments (including off-treatment assessments) at Week 13, 26, 39 and 52. Missing Week 52 standing height assessments will then be imputed from the subjects with reference data (as defined above) using the missing not at random (MNAR) option.

A fixed seed of 5122019 will be specified to ensure reproducibility of the results and the number of imputations set to 10.

In each of the imputed datasets, missing AGV at Week 52 and the corresponding missing change from baseline in AGV will be calculated using the imputed standing height. Note, missing height Z-scores at Week 52 and the corresponding change from baseline will also be calculated. For each subject whose missing Week 52 value is imputed, the 10 imputed values will be provided in a data listing that will be included in the appendix of the CSR.

Each of the 10 imputed datasets will be analyzed using an ANCOVA model that includes the following fixed-effect terms:

- Treatment group
- Stratum (Male Tanner Stage I, Female Tanner Stage I, Male Tanner Stage >I, Female Tanner Stage >I)

- Baseline age
- Baseline AGV
- Baseline height Z-score

SAS PROC MIXED will be used to perform the analysis.

PROC MIXED;

CLASS Stratum Treatment;

MODEL Change from baseline in AGV at Week 52 = Baseline_Age Baseline_AGV Baseline_Height_Z-Score Treatment Stratum;

LSMEANS Treatment / DIFF = control ("Placebo") CL;

RUN;

For each imputed dataset, PROC MIXED will generate the least-squares (LS) mean change from baseline in AGV at Week 52 for each treatment group, along with the difference in LS means (calculated as BMN 111 – Placebo). PROC MIANALYZE will then be used to combine all of the estimates across the 10 imputed datasets in order to generate an overall estimate for each treatment group, and for the difference between treatment groups, with a 95% confidence interval for the difference and 2-sided p-value. The study will be considered positive if the 2-sided p-value in favor of BMN 111 is <0.05.

If there are less than 5 subjects with reference data as defined above then missing standing height at Week 52 (i.e. Day 365) will be imputed by applying the baseline AGV (cm/yr) to the last available height assessment as follows:

Standing Height at Day 365 =

Last Assessed Height + $\frac{AGV\ at\ Baseline}{365.25}$ x (365 - Study Day of Last Height Assessment)

Based on this imputed standing height, the height z-score and AGV at Week 52 will be calculated. The ANCOVA model will then be applied to AGV at Week 52.

In the event that there are no subjects who discontinue treatment early with a missing assessment at Week 52 the ANCOVA model will be applied without any imputation.

This ANCOVA model analysis without imputation will be repeated on the PP population.

14.3.3 Sensitivity and Supplementary Analyses Defined for AGV

The following sensitivity and supplementary analyses are proposed for assessing the robustness of the results of the primary analysis. All visits referred to in the analyses, are based on the analysis windows described in Section 5.4. Summary tables for the models and descriptive summary statistics tables will be provided for each sensitivity analysis.

Sensitivity Analysis: Washout model where missing data for an endpoint for both treatment groups will be imputed using data from the placebo arm.

This approach addresses the same estimand as the primary analysis but with an alternative estimator. Employing this model, any potential BMN 111 treatment effect in those patients in the BMN 111 active group who are off-treatment will be washed out. Specifically, for the washout imputation, the missing data in an endpoint for both treatment groups will be imputed using observed endpoint data from the placebo arm.

PROC MI will be used to impute the missing standing height assessments at Week 52, as described for the primary analysis. However, for this analysis, according to the washout specifications, only placebo data (including off treatment assessments) will be used in the missing not at random (MNAR) option to implement the control-based pattern imputation. Therefore, for the subjects in the BMN 111 active treatment group, no observed post baseline assessments will be considered in the imputation process.

Supplementary Analysis 1: ANCOVA model on completers.

This approach addresses the same estimand as the primary analysis, except that the intercurrent event is defined as subject discontinuation from the study. In this analysis, the estimator will be the difference between the two randomized groups in the mean change from baseline in AGV at Week 52, observed from subjects who have non-missing AGV observation at 52 weeks. The LS means of the change from baseline in AGV at 52 weeks will be estimated for both the randomized groups by applying the ANCOVA model without multiple imputation described in the primary analysis. In this analysis, only subjects with observed AGV change from baseline at 52 weeks will be included. No imputation will be performed for missing 52-week standing height measurements.

Supplementary Analysis 2: ANCOVA model excluding assessments that occur post either limb lengthening or growth hormone/gonadotropin-releasing hormone.

This approach addresses the same estimand as the primary analysis except that the intercurrent event, defined as rescue treatment for achondroplasia, is handled differently. In the event that a subject has been receiving rescue medication from the start of the study, they

will be excluded from this supplementary analysis. The analysis will be conducted on imputed data, as described for the primary analysis; however, any assessments that are measured following either limb lengthening or growth hormone/ gonadotropin-releasing hormone will be excluded from this analysis. The ANCOVA model described in the primary analysis will be applied, and the estimator will be the difference between the two randomized groups in the mean change from baseline in AGV at Week 52.

14.4 Key Secondary Efficacy Endpoint(s)

The two secondary efficacy endpoints with confirmatory testing (referred to in this document as "key secondary endpoints") are the change from baseline at Week 52 in height Z-score and the change from baseline at Week 52 in upper to lower body segment ratio.

14.4.1 Height Z-Score

Standing height is measured at Screening, Day 1, Week 13, Week 26, Week 39, and Week 52. Each measurement of standing height will be converted to an age-and sex-appropriate standard deviation score (SDS), also referred to as a Z-score, by comparison with reference data available for average stature children from the Centers for Disease Control and Prevention (CDC).

The following key secondary hypothesis will be tested (two-tailed):

> *H0: Difference in mean height Z-score change from baseline at week 52 between BMN 111 group and the placebo group = 0*
>
> *Ha: Difference in mean height Z-score change from baseline at week 52 between BMN 111 group and the placebo group ≠ 0*

Height Z-score will be analyzed using the same methods and estimand formulation as the primary analysis. If multiple imputation is applied for testing the primary endpoint then the null hypothesis of no difference between the BMN 111 group and the placebo group in the mean change from baseline in the height Z-score at the Week 52 time point will also be tested based on 10 imputed data sets. Missing height Z-score at Week 52 will be calculated from the imputed standing height at this visit, as described for the primary endpoint. If multiple imputation is not applied for the primary endpoint, then the Week 52 height Z-score will be imputed as described above in Section 14.3.2.

SAS PROC MIXED will be used to test the null hypothesis using an ANCOVA model described below.

PROC MIXED;

CLASS Stratum Treatment;

MODEL Change from baseline in Height Z-score at Week 52 = Baseline_Age Baseline_AGV Baseline_Height_Z-score Treatment Stratum;

LSMEANS Treatment / DIFF= control ("Placebo") CL;

RUN;

The least-squares (LS) mean change from baseline in height Z-score at Week 52 will be presented for each treatment group, along with the difference in LS means (calculated as BMN 111 – Placebo), the 95% confidence interval for the difference, and 2-sided p-value. In the event that the primary endpoint test is positive, the testing on the key secondary is considered to be confirmatory. The endpoint is considered positive if the 2-sided p-value in favor of BMN 111 is < 0.05.

This analysis will be repeated on the PP population, using the same approach as described for the primary endpoint. In addition, the sensitivity and supplementary analyses of height Z-score [described in section 14.3.3] will be performed using the FAS population.

Summary tables are provided as described in Section 14.1. In addition, a line plot of the mean (+/-SD) height Z-score will be presented for visits ≥ 12 Months Prior, ≥ 6 Months Prior, Baseline, Week 26 and Week 52 by treatment group.

Height Z-scores derived for all visits will be included in data listings.

14.4.2 Upper to Lower Body Segment Ratio

Standing height and sitting height are measured at Screening, Day 1, Week 13, Week 26, Week 39, and Week 52.

The upper to lower body segment ratio will be calculated at each visit as follows:

$$\text{Upper to Lower Body Segment Ratio} = \frac{Sitting\ height(cm)}{Standing\ height(cm) - Sitting\ height(cm)}$$

The upper to lower body segment ratio at Baseline, Week 26 and Week 52, and it's change from baseline at Week 26 and Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group. Figures will be provided as described in Section 14.1.

The following key secondary hypothesis will be tested (two-tailed):

H0: Difference in mean upper:lower body segment ratio change from baseline at week 52 between BMN 111 group and the placebo group = 0

Ha: Difference in mean upper:lower body segment ratio change from baseline at week 52 between BMN 111 group and the placebo group ≠ 0

The upper to lower body segment ratio will be analyzed using the same methods and estimand formulation as the primary analysis. The primary analysis of the upper to lower body segment ratio will be assessed on the FAS. If MI methods are used to impute Week 52 results for the primary endpoint (AGV), then the same MI approach will be used to impute missing Week 52 upper to lower body segment ratios. Missing sitting height at Week 52 will be imputed in the same manner as missing standing height, using PROC MI. Missing upper to lower body segment ratios Week 52 will be calculated from the imputed values for sitting and standing height. The difference between the BMN 111 group and the placebo group in the mean change from baseline in the upper:lower body segment ratio at Week 52 will be tested based on 10 imputed data sets.

If multiple imputation is not applied for the primary endpoint, then missing sitting height and standing height at Week 52 will be imputed by applying the baseline AGV for sitting height and standing height to the last available sitting/standing height assessment (as described in Section 14.3.2). Based on this imputed sitting height and standing height at Week 52, the upper:lower body segment ratio will be calculated, and the model described above will then be applied at Week 52.

SAS PROC MIXED will be used to test the null hypothesis using an ANCOVA model described below.

PROC MIXED;

CLASS Stratum Treatment;

MODEL Change from baseline in upper:lower body segment ratio at Week 52 = Baseline_Upper:Lower Body_Segment_Ratio Baseline_Age Baseline_AGV Baseline_Height_Z-score Treatment Stratum;

LSMEANS Treatment /DIFF = control (“Placebo”) CL;

RUN;

The least-squares (LS) mean change from baseline in upper to lower body segment ratio at Week 52 will be presented for each treatment group, along with the difference in LS means (calculated as BMN 111 – Placebo), the 95% confidence interval for the difference, and

2-sided p-value. In the event that the height Z-score endpoint test is positive, the testing on this key secondary endpoint is considered to be confirmatory. The endpoint is considered positive if the 2-sided p-value in favor of BMN 111 is < 0.05.

This analysis will be repeated on the Per Protocol population. In addition, the sensitivity and supplementary analyses of upper to lower body segment ratio [described in Section 14.3.3] will be performed using the FAS population.

A line plot of the mean (+/-SD) upper to lower body segment ratio will presented at ≥ 12 Months Prior, ≥ 6 Months Prior, Baseline, Week 26 and Week 52 time points by treatment group.

All upper to lower body segment ratios will be listed alongside the standing and sitting height.

14.5 Other Secondary Efficacy Endpoint(s)

Any p-values resulting from analyses of these secondary efficacy endpoints are non-confirmatory.

14.5.1 Body Proportion Ratios

Body measurements are taken at Screening, Day 1, Week 13, Week 26, Week 39, and Week 52. Each of the body proportion ratios below will be calculated to 2 decimal places.

- Upper Arm Length to Lower Arm (Forearm) Length Ratio
- Upper Leg Length (Thigh) to Knee to Heel Length Ratio
- Upper Leg Length (Thigh) to Tibial Length Ratio
- Arm Span to Standing Height Ratio

The absolute values at ≥ 12 Months Prior, ≥ 6 Months Prior, Baseline, Week 26 and Week 52, and change from baseline at Week 26 and Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group.

A line plot of the mean (+/-SD) each body proportion ratio will be presented at ≥ 12 Months Prior, ≥ 6 Months Prior, Baseline, Week 26 and Week 52 time points by treatment group.

Box plots of each ratio at Week 52 will be provided by treatment group.

All assessments will be listed by treatment group, subject and visit.

Analysis:

No imputation will be applied for missing Week 52 assessments. Analyses will be based on observed data only, and SAS PROC MIXED will be used to test the null hypothesis of no treatment difference using an ANCOVA model described below, for each of the above ***body proportion ratios***.

PROC MIXED;

CLASS Stratum Treatment;

MODEL Change from baseline in ***body proportion ratio*** at Week 52 = Baseline_***body proportion ratio*** Baseline_Age Baseline_AGV Baseline_Height_Z-score Treatment Stratum;

LSMEANS Treatment /DIFF = control ("Placebo") CL;

RUN;

For each body proportion ratio, the least-squares (LS) mean change from baseline at Week 52 will be presented for each treatment group, along with the difference in LS means (calculated as BMN 111 – Placebo), the 95% confidence interval for the difference, and 2-sided p-value.

14.5.2 Growth Measures

Each of the following anthropometric growth measures are recorded at Screening, Day 1, Week 13, Week 26, Week 39, and Week 52:

- Standing Height (cm)
- Sitting Height (cm)
- Upper Leg Length (Thigh) (cm)
- Knee to Heel Length (cm)
- Tibial Length (cm)
- Head Circumference (cm)
- Upper Arm Length (cm)
- Lower Arm (Forearm) Length (cm)
- Arm Span (cm)

In addition, a measure of lower body length will be calculated as follows:

- Lower Body Length (cm) = Standing height – Sitting height

Baseline growth measures are those assessed on Day 1, regardless of timing relative to dosing. All summaries and analyses will be based on the FAS.

For each of the above growth measures (including lower body length), the absolute values at ≥ 12 Months Prior, ≥ 6 Months Prior, baseline, Week 26 and Week 52, and change from baseline at Week 26 and Week 52, will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group for each baseline age category (≥ 5 to < 8, ≥ 8 to < 11, ≥ 11 to < 15, ≥ 15 to < 18 years).

Analysis:

Subjects with missing standing height at Week 52 will have standing height imputed, as described in Section 14.3.2. The same approach will be applied for missing sitting height assessments at Week 52.

With the exception of sitting height and standing height, no imputation will be applied for missing Week 52 assessments. Analyses will be based on observed data only.

For each growth measure, the null hypothesis of no difference between the BMN 111 group and the placebo group in the mean change from baseline at the Week 52 visit will be tested.

SAS PROC MIXED will be used to test the null hypothesis of no treatment difference using an ANCOVA model described below, for each of the above ***growth measures***.

PROC MIXED;

CLASS Stratum Treatment;

MODEL Change from baseline in ***growth measure*** at Week 52 = Baseline***_growth_measure*** Baseline_AGV Baseline_Height_Z-score Treatment Stratum;

LSMEANS Treatment /DIFF CL;

RUN;

For each growth measure, the least-squares (LS) mean change from baseline at Week 52 will be presented for each treatment group, along with the difference in LS means (calculated as BMN 111 – Placebo), the 95% confidence interval for the difference, and 2-sided p-value.

Spaghetti plots of standing height will be provided by treatment group and sex. These 8 plots will include age-sex specific reference ranges for average stature children (from the Centers for Disease Control and Prevention (CDC)) and age-sex specific reference ranges for short

stature children (Hoover-Fong et al). All assessments will be included, including all height assessments collected in the 901 study up to the ≥ 12 months prior analysis visit.

All assessments will be listed by treatment group, subject and visit.

14.5.3 Body Mass Index

Weight (kg) and standing height are collected at Screening, Day 1, Week 13, Week 26, Week 39, and Week 52.

Body Mass Index (BMI) will be calculated at each visit as:

$$\text{Body Mass Index (kg/m}^2) = \frac{Weight\ (kg)}{Height\ (m)2}$$

Each measurement of BMI will be converted to an age-and sex-appropriate standard deviation score (SDS), also referred to as a Z-score, by comparison with reference data available for average stature children from the Centers for Disease Control and Prevention (CDC).

The absolute values at Baseline, Week 26 and Week 52, and change from baseline at Week 26 and Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group for both BMI and BMI Z-Score.

All assessments will be listed by treatment group, subject and visit.

14.5.4 Weight Z-Scores

Weight (kg) is measured at Screening, Day 1, Week 13, Week 26, Week 39, and Week 52. Each measurement of weight will be converted to an age-and sex-appropriate standard deviation score (SDS), also referred to as Z-score, by comparison with reference data available for average stature children from the Centers for Disease Control and Prevention (CDC).

The weight of the subject on Day 1 will be used as the baseline measure, regardless of whether this is recorded pre or post-dose.

The absolute values at Baseline, Week 26 and Week 52, and change from baseline at Week 26 and Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group.

All assessments will be listed by treatment group, subject and visit.

14.5.5 Health-Related Quality of Life and Functional Independence

The Health-Related Quality of Life (HRQoL) questionnaires and functional independence questionnaires are captured at Screening, Week 26 and Week 52. All summaries will be generated on the FAS.

Each table will include summaries of the absolute measures at each of these time points, and change from baseline at Week 26 and Week 52. The tables will be summarized by treatment group.

The statistical summaries will include the number of subjects with assessable data, mean, SD, median, 25th and 75th percentile, minimum and maximum. Subjects will not be included in a summary table if there is no baseline assessment available.

Separate summary tables will be provided for the total summary scores and the individual domains (scales). Individual questions (items) are not included in the summary tables.

Specific domains that are not considered clinically meaningful for the ACH indication are indicated.

Age defined versions within the same questionnaire will be summarized together. If a subject progresses during the course of the study from one age defined questionnaire to the next age defined questionnaire the results will be summarized together and the baseline assessment from the child's first questionnaire is referred to when summarizing change from baseline.

Self-reports and parent-reports will be summarized separately.

All subject data listings will include total scores, and domain (scale) scores.

14.5.5.1 Pediatric Quality of Life Inventory (PedsQL)

The PedsQL scales were designed to measure the core dimensions of health, as delineated by the WHO, as well as role (school) functioning. The following versions of the PedsQL are utilized in this protocol:

Parent/Caregiver Reported

- Parent Proxy-Report Children Ages 5-7
- Parent Proxy-Report Children Ages 8-12
- Parent Proxy-Report Adolescents Age 13-18

Self-Reported

- Child Self-Report Ages 8-12

- Adolescent Self-Report Ages 13-18

The PedsQL consists of 4 scales of functioning (Physical, Emotional, Social, and School), which make up 3 summary scores.

Each of the items within a scale is scored as 0 (never a problem), 1 (almost never a problem), 2 (sometimes a problem), 3 (often a problem), to 4 (almost always a problem). Items are then reverse scored and linearly transformed to a 0-100 scale as follows: 0=100, 1=75, 2=50, 3=25, and 4=0.

Scales/Domains:

- Physical Functioning (8 items, min score=0, max score=800)
- Emotional Functioning (5 items, min score=0, max score=500)
- Social Functioning (5 items, min score=0, max score=500)
- School Functioning (5 items, min score=0, max score=500)

If at least 50% of items are completed within a scale, then the **mean score** is calculated for each scale using the transformed values:

$$\text{Scale/Domain Score} = \frac{Sum\ of\ item\ scores}{Number\ of\ items\ answered}$$

If more than 50% of the items are missing within a scale, then scale scores should not be calculated and be set to missing. (Varni, 1999)

For each of the scales, the mean scores (based on transformed values), and change from baseline at each scheduled visit, will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) by treatment group.

Summary scores:

In addition to the mean scale scores, the following summary scores will be calculated. All calculations use the transformed data.

- Psychosocial Health Summary Score (Sum of item scores/Number of items answered on Emotional, Social and School scales)
- Physical Health Summary Score (Sum of item scores/Number of items answered on Physical scale)
- Total Score (Mean Score, i.e., Sum of all items/Number of items answered on all scales)

For each of the above summary scores, the absolute values, and change from baseline at each scheduled visit, will be summarized (n, mean, SD, median, 25th and 75th percentile,

minimum, maximum) and presented by treatment group. These scores are based on the transformed values.

Domain (scale) scores and summary scores will be listed by treatment group, subject and visit date.

14.5.5.2 Quality of Life in Short Statured Youth (QoLISSY)

The QoLISSY for parents, and the version for children and adolescents, consists of 22 likert-scaled items assigned to the core QoL dimensions: Physical, Social and Emotional, and 14 additional items in the dimensions Coping and Beliefs. In addition, the parent-reported version captures data on Future and Effects on parents.

The two versions of the questionnaire being used in this study are:

Parent/Caregiver Reported

- QoLISSY parent version for short stature (4-18)

Self-Reported

- QoLISSY child version (patient reported >= 8 years)

In both versions of the questionnaire, scoring is assigned to each individual item/question, according to whether the question is positively or negatively worded. Positively worded items are scored 1, 2, 3, 4, 5 (1: not at all/never, 2: very/very often, 3: moderately/quite often, 4: slightly/seldom, 5: extremely/always). Negatively worded items are reverse scored 5, 4, 3, 2, 1 (5: not at all/never to 1: extremely/always) so that a higher subscale score reflects higher quality of life.

Each of the questions are assigned to domains.

Domains

The following domains are common to both the parent and child versions of the QoLISSY:

- Physical (6 items)
- Social (8 items)
- Emotional (8 items)
- Coping (10 items)
- Belief (4 items)

The parent-report version contains additional items from two parent-only domains.

- Future (5 items)

- Effects on parent (10 items, excluding growth hormone question.)

For each QoLISSY domain, the individual raw scores are summed to determine the raw domain total score. In order to calculate a domain score, 80% of the items must be completed (Bullinger 2015).

The QoLISSY raw domain total scores are then transformed to standard scores on a scale of 0-100 to allow for easier comparison from one measure to another. A score closer to 0 is indicative of poor quality of life and higher scores closer to 100 are indicative of a higher quality of life.

For each domain, the raw mean is transformed to a standard score as follows:

$$Domain\ transformed\ score = \frac{(Raw\ Mean - 1)}{4} x\ 100$$

For each domain, the absolute values of the transformed scores, and change from baseline at each scheduled visit, will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group.

Summary Scores

The QoLISSY Total Score is calculated by summing the means in the physical, social, and emotional subscales and dividing by 3. The absolute value of the Total Score, and change from baseline at each scheduled visit, will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group.

Domain (scale) scores and summary scores will be listed by treatment group, subject and visit date.

14.5.5.3 Functional Independence Measure (WeeFIM)

The WeeFIM (Functional Independence Measure for children) instrument is a functional independence assessment tool that measures functional performance across three domains (self-care, mobility and cognition) from the caregiver's perspective.

Performance of a child on each of the individual items within the WeeFIM is assigned to one of seven levels on an ordinal scale that represent the function from complete and modified independence (levels 7 and 6) without a helping person to modified and complete dependence (levels 5 to 1) with a helping person.

If individual items (questions) within a domain are missing, the item result is imputed to 1, per guidance in The WeeFIM Clinical Guide v 6.49.

If all items for a domain score are missing the domain score will be considered missing, and the Total score will not be calculated.

For each of the following domains and the total score, the absolute values at each scheduled visit, and change from baseline at each scheduled visit, will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) by treatment group:

- Self-care Score (min score=8, max score=56)
- Mobility Score (min score=5, max score =35)
- Cognition Score (min score=5, max score=35)
- Total WeeFIM rating (min=18, max=126)

Domain and total scores will be listed by treatment group and subject number.

Note: Although each domain will be summarized so that all aspects of this validated scale are reported, the cognition aspect of the WeeFIM (cognition subtotal score) is considered not to be of concern for subjects with achondroplasia.

14.6 Exploratory Endpoints

14.6.1 Sleep Study Scores

An optional sleep sub-study will be performed in a limited number of qualified sleep centers for a subset of subjects. A sleep-testing device will be used to assess the presence and severity of sleep-disordered breathing by measurement of blood oxygen saturation, pulse rate, and airflow during overnight monitoring. Sleep study data is collected at Screening, and Week 52.

Due to a small number of subjects enrolled in the optional sleep study (n=8) only listings of episodes of sleep apnea will be provided. These listings will include:

- Number of episodes of apnea per hour (Apnea Index).
- Number of episodes of hypopnea per hour (Hypopnea Index).
- Number of episodes /Duration of obstructive apneas
- Number of episodes /Duration of central apneas
- Number of episodes /Duration of mixed apneas
- Number of episodes /Duration of obstructive hypopneas
- Number of desaturations

All parameters collected in the sleep study will be listed.

14.7 Examination of Efficacy by Subgroups

Subgroup analyses will be conducted on the primary and two key secondary efficacy endpoints, and will be based on the imputed data described for the primary analysis. Only subgroup categories with sufficient data will be analyzed. The ANCOVA model described in Section 14.3.2 will be applied separately for each category within a subgroup.

A Forest plot will provide an overall summary, and for each subgroup category, the difference between treatment group LS means (estimated as BMN 111-placebo), and the 95% confidence interval for the difference, at Week 52.

Subgroup categories of interest include:

- Sex (Male, Female)
- Age at Baseline (≥ 5 to < 8, ≥ 8 to < 11, ≥ 11 to < 15, ≥ 15 to < 18 years)
- Tanner stage at Baseline (I, >I)
- Stratum (Male Tanner Stage I, Female Tanner Stage I, Male Tanner Stage >I, Female Tanner Stage >I).
- Baseline Height Z-Score (≤ -6, > -6 to ≤ -5, > -5 to ≤ -4, > -4)
- Baseline AGV (≤ 3.5, > 3.5 to ≤ 4.5, > 4.5 cm/yr)

15 SAFETY EVALUATIONS

The safety population will be used for all summaries where assessments/events are reported by treatment group.

Summary tables will include all safety events up to 30 days following treatment discontinuation. PI

Safety will be assessed by examining the incidence, severity (determined using the Common Terminology Criteria for Adverse Events (CTCAE) version 4), and relationship to study drug of all treatment-emergent adverse events (TEAEs) reported during the study period. In addition, changes from baseline in clinical laboratory results and vital signs will be assessed.

15.1 Adverse Events

Adverse events (AEs) will be coded in accordance with Medical Dictionary for Regulatory Activities (MedDRA) Version 22. MedDRA coding will be performed prior to database lock. Coding of the severity of adverse events is performed by the investigators using NCI CTCAE version 4, where events are coded from CTCAE Grade 1 to Grade 5.

Only treatment-emergent adverse events (TEAEs) defined as any adverse event that newly appeared, increased in frequency, or worsened in severity following initiation of study drug administration are reported by the investigators and consequently are included in the summary tables.

In the event that the start date of an AE is incomplete or missing, conservative imputation rules are applied so that where there is uncertainty, the event is considered treatment emergent. Similarly the AE end dates are imputed in a conservative manner to a maximum length (Section 5.5.1).

15.1.1 Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) will be summarized by treatment group. In summaries by system organ class (SOC) and/or preferred term (PT), subjects with more than one AE of the same SOC or PT will be counted once within that SOC/PT. The tables are ordered by the frequency of the highest reporting level (SOC or PT) in the table within the active treatment group.

The following summary tables will be provided:

- Overview of the incidence of adverse events

- Incidence and exposure-adjusted event rates of TEAEs by system organ class (SOC), preferred term (PT) and CTCAE grade
- Incidence of TEAEs by system organ class (SOC), preferred term (PT) and highest CTCAE grade
- Incidence and exposure-adjusted event rates of TEAEs by PT
- Incidence and exposure-adjusted event rates of TEAEs by SOC
- Incidence and exposure-adjusted event rates of drug-related TEAEs by SOC, PT and CTCAE grade
- Incidence of serious adverse events (SAEs) by SOC and PT
- Incidence of TEAEs leading to study or study drug discontinuation by PT
- Incidence of TEAEs leading to study drug interruption by PT
- Incidence of TEAEs leading to study drug dose reduction by PT
- Incidence and exposure-adjusted event rates of achondroplasia-related TEAEs by SOC and PT*

* Achondroplasia-related TEAE's will be identified using the PT's identified for achondroplasia-related medical history (see Section 10). In addition, all AEs will be reviewed prior to database lock and the list updated as required.

15.1.2 Events of Interest

Events of interest (EOI) will be summarized by PT and treatment group.

The following are identified as events of interest:

Injection site reactions (ISR)

- TEAEs with a MedDRA High Level Term (HLT) of "Injection site reaction".

Hypotension

- TEAEs with PT: Blood pressure ambulatory decreased, Blood pressure decreased, Blood pressure diastolic decreased, Blood pressure orthostatic decreased, Blood pressure systolic decreased, Blood pressure systolic inspiratory decreased, Diastolic hypotension, Hypotension, Orthostatic hypotension.

Heart rate change

- TEAEs with a PT: Atrial tachycardia, Postural orthostatic tachycardia syndrome, Rebound tachycardia, Sinus tachycardia, Supraventricular tachycardia, Tachycardia, Tachycardia paroxysmal, Ventricular tachycardia, Bradycardia.

Hypersensitivity SMQ

- TEAEs with a PT included in the MedDRA hypersensitivity SMQ

Algorithmic Anaphylaxis SMQ (with sponsor-defined time restrictions)

- TEAEs with a PT included in the MedDRA anaphylactic reaction SMQ, with an additional time restriction: a narrow scope PT at any time, or two broad scope PTs from different classes where both PTs are within 24 hours of the same dose.

Fractures

- TEAEs with a MedDRA HLT: Fractures NEC, Limb fractures, Pelvic fractures, Skull and face fractures, Spinal column fractures, thoracic cage fractures non-spinal

Slipped Capital Femoral Epiphysis (SCFE)

- TEAEs with a PT: Epiphyseal disorder, Epiphyseal injury

Avascular necrosis and osteonecrosis

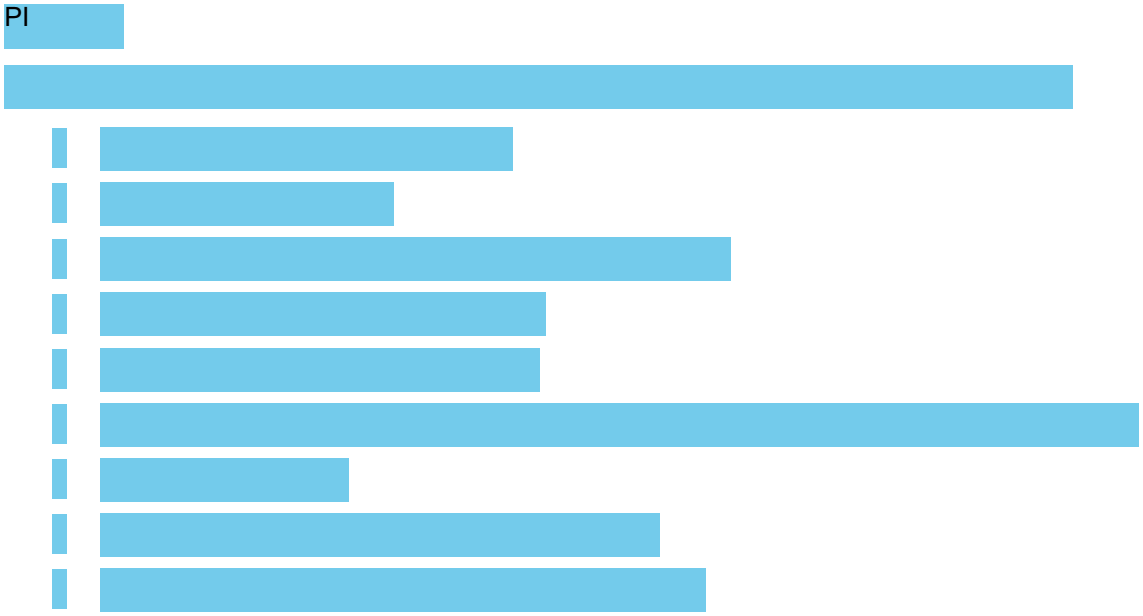
- TEAEs with a PT: Osteonecrosis, Osteonecrosis of jaw, Osteonecrosis of external auditory canal, Necrosis ischaemic

Events of interest will be summarized as follows:

- Incidence and exposure-adjusted event rates of injection site reaction events by PT
- Incidence and exposure-adjusted event rates of injection site reaction events by PT and duration of ISR (<=15min, >15min and≤60min, >60min and ≤120min, >120min and ≤24hours, >24hours)
- Incidence and exposure-adjusted event rates of hypotension by PT
- Incidence and exposure-adjusted event rates of heart rate change events by PT
- Incidence and exposure-adjusted event rates of hypersensitivity (SMQ)

- Profile of injection site reaction events
- Profile of hypotension
- Profile of heart rate change events

Profile summaries include: highest CTCAE grade, number of events per subject, time from first dose to first event onset, duration of events (days), action taken with study drug, outcome of events.



15.1.3 Injection Site Reaction Symptoms

Injection site reactions (ISRs) associated with single symptoms are recorded on the AE page (as a single symptom). ISR's with multiple symptoms are recorded as an AE of "Injection Site Reaction" on the AE page and the associated individual symptoms are recorded on the ISR symptom page.

In order to describe all ISR symptoms recorded, summaries will be based on the data collected on the ISR Symptoms page and also those single symptoms recorded as an adverse event.

Incidence and exposure-adjusted event rates of ISR Symptoms will be summarized by PT and treatment group.

ISR Symptoms will be listed.

15.2 Clinical Laboratory Tests

Clinical laboratory tests (hematology, chemistry and urinalysis), are performed at Screening, and pre-dose at Day 1, Day 10, Week 6, Week 13, Week 26, Week 39, Week 52 and the Safety Follow-up visit. This study uses a centralized laboratory. All laboratory results are presented using SI units. When results are reported beyond analytical ranges, the reported numeric value will be used and included in summaries.

All summary tables will include laboratory assessments from baseline up to 30 days post treatment discontinuation. PI ’.

Laboratory tests will be graded as low/normal/high based on laboratory normal ranges. In addition, for laboratory tests with CTC grading available, all non-missing numeric results will be used to determine CTC grade programmatically, based on CTCAE v4.0.

For the following parameters, the distinction is made between significantly low/high results:

- Glucose (Hypoglycemia/Hyperglycemia)
- Hemaglobin (Anemia/Hemaglobin increased)
- Lymphocytes (Lymphocyte count decreased/Lymphocyte count increased)
- Potassium (Hypokalemia/Hyperkalemia)
- Sodium (Hyponatremia/Hypernatremia)
- Calcium (Hypocalcemia/Hypercalcemia)

The absolute values for pre-dose laboratory results at each scheduled visit, change from baseline in pre-dose laboratory results, and percent change from baseline in pre-dose laboratory results at each scheduled visit will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group.

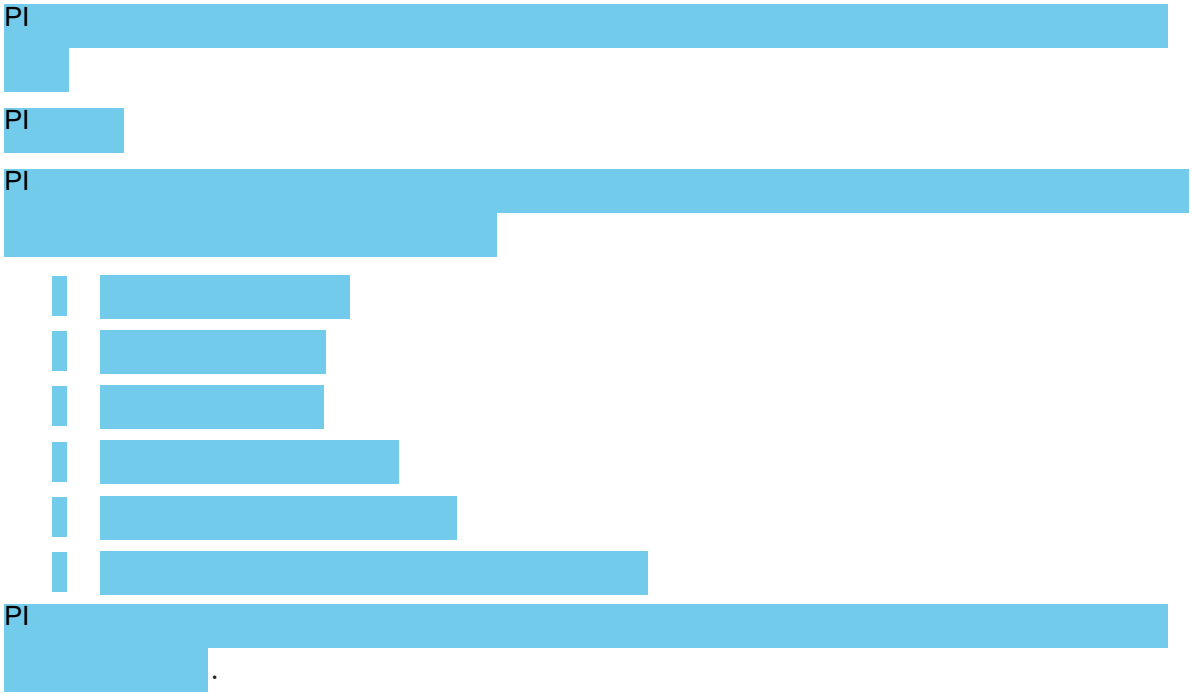
Shift tables from baseline to worst post-baseline value (including scheduled and unscheduled visits) based on the CTC grading (Normal – Grade 5) will be generated for each Lab parameter where CTCAE grading is available, excluding parameters where CTCAE Grade does not rely on quantitative results alone (e.g. potassium and uric acid). Percentages are based on the number of subjects with each Baseline CTCAE grade (Normal, Grade 1, Grade 2, Grade 3, Grade 4, Grade 5, Missing, Overall).

For laboratory tests where grades are not defined by CTCAE, or where CTCAE grading does not rely on quantitative results alone (e.g. potassium and uric acid), shift tables will be

generated using the low/normal/high classification to compare to baseline to the worst post-baseline value.

Line plots of the mean result (+/- SD) for the pre-dose absolute values by scheduled visit for each treatment group will be provided for the following parameters:

- Hematology: erythocytes, hematocrit, hemoglobin, leukocytes, neutrophils, platelets.
- Chemistry: alanine aminotransferase, albumin, alkaline phosphatase, aspartate aminotransferase, bicarbonate, bilirubin, blood urea nitrogen, calcium, chloride, creatinine, direct bilirubin, glucose, lactate dehydrogenase, phosphate, potassium, protein, sodium, thyrotropin, vitamin D.



.

15.3 Vital Signs

Vital Signs (heart rate, systolic/diastolic blood pressure, respiratory rate and body temperature,) are assessed at Screening, Day 1, Day 2, Day 3, Day 10, Week 6, Week 13, Week 26, Week 39, Week 52 and the Safety Follow-up visit.

All summary tables will include vital signs recorded from baseline up to 30 days post treatment discontinuation.

The vital signs that are collected pre-dose and post-dose are:

- **Pre-dose**: body temperature, heart rate, systolic/diastolic blood pressure and respiratory rate.

- **Post-dose**: heart rate and systolic/diastolic blood pressure only.
    - Day 1: 0-1 hr. (Q15min); 1-2hr (Q30min)
    - Day 2-3: 0-2 hr. (Q30min)
    - All other dosing visits: 0-1 hr. (Q30min)

The absolute values for pre-dose vital signs at each scheduled visit, and change from baseline in pre-dose vital signs at each scheduled visit will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group. Line plots will be generated which present the mean pre-dose absolute values (+/- SD) by scheduled visit for each treatment group.

In addition, for heart rate and systolic/diastolic blood pressure, pre-dose assessments, and change from pre-dose to multiple post-dose time points at each scheduled visit will be similarly summarized and presented by treatment group. Line plots will be generated which present the mean pre-dose and post-dose absolute values (+/- SD) by treatment group for each visit separately.

The percentage of subjects experiencing at least one instance of a decrease of 20% in diastolic blood pressure from pre-dose to post dose will be summarized [n(%)] by treatment group at Baseline, Day 2, Day 3, Day 10, Week 6, Week 13, Week 26, Week 39, and Week 52.

Shift tables (Diastolic BP <40mmHg, Diastolic BP ≥40mmHg) from pre-dose to lowest post-dose value category (Diastolic BP <40mmHg, Diastolic BP ≥40mmHg) will be generated by visit and at any time.

Shift tables (Diastolic BP <45mmHg, Diastolic BP ≥45mmHg) from pre-dose to lowest post-dose value category (Diastolic BP <45mmHg, Diastolic BP ≥45mmHg) will be generated by visit and at any time.

Shift tables (Systolic BP < (70mmHg plus 2 x Age), Systolic BP ≥ (70mmHg plus 2 x Age)) from pre-dose to lowest post-dose value category ((Systolic BP < (70mmHg plus 2 x Age), Systolic BP ≥ (70mmHg plus 2 x Age)) will be generated by visit and overall (overall will be the lowest overall post-dose value category at any time on study). Subject's age used in this calculation is the integer value.

PI

15.4 Electrocardiogram

Electrocardiogram (ECG) data are recorded at Screening, Day 1, Day 10, Week 13, Week 26, Week 39, Week 52, and at the Safety Follow-up visit.

ECG assessments are performed in triplicate on Day 1 (pre-dose and post-dose) and on study day visits (post-dose only). On days when PK samples are being drawn, post-dose ECGs should be performed within a 5-minute window prior to the 30-minute PK assessment (i.e. around the time of CMax).

Summaries will be provided over all assessments, and repeated with the requirement that all post-dose ECG assessments should occur around the expected time of CMax. The mean results of the ECG assessments meeting these criteria will be used for summaries and analyses.

All measures and means of the measures will be included in the listings.

The following ECG results will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group at Baseline (pre-dose and post-dose), and post-dose at Day 10, Week 13, Week 26, Week 39, Week 52:

- ECG Mean Heart Rate (beats/min)
- QT interval (msec)
- QTcB interval (msec)
- QTcF interval (msec)
- PR interval (msec)
- RR interval (msec)
- QRS Duration (msec)

The following endpoints will be summarized [n (%)] at Baseline, Day 10, Week 13, Week 26, and Week 52, and across all time points (using the highest mean QTc value), and presented by treatment group.

- Percentage of subjects with QTcB >450 to ≤480ms, >480 to ≤500ms, >500 msec
- Percentage of subjects with QTcF >450 to ≤480ms, >480 to ≤500ms, >500 msec

The following endpoints will be summarized [n (%)] at Baseline, Day 10, Week 13, Week 26, and Week 52, and across all time points, and presented by treatment group.

- Percentage of subjects experiencing at least one instance of a (decrease in mean heart rate from pre-dose baseline >25% and a mean heart rate <50 beats/min) post-dose.

- Percentage of subjects experiencing at least one instance of an (increase in mean heart rate from pre-dose baseline >25% and a mean heart rate >100 beats/min) post-dose.

The following endpoints will be summarized [n (%)] across all time points and presented by treatment group.

- Percentage of subjects with a QTcF increase of 60ms from pre-dose baseline at any time
- Percentage of subjects with QTcF changes from pre-dose baseline of >30 and ≤ 60 msec, or >60 msec at any time
- Percentage of subjects with both an increase in PR interval from pre-dose baseline >25% and a PR>200 msec at any time.
- Percentage of subjects with both an increase in QRS from pre-dose baseline >25% and QRS >120 msec at any time

Listings will be provided, and ordered by treatment group, subject and visit:

- All ECG results.
- All ECG results for subjects with a QTc increase of 60msec from pre-dose Baseline.

15.5 Child Behavior Checklist

The Child Behavior Checklist (CBCL) questionnaire is captured at Screening, Week 26 and Week 52. All summaries will be generated on the Safety Population and include all assessments recorded up to 30 days following treatment discontinuation. Listings will include all reported data.

The CBCL 1.5-5 years old and the CBCL 6-18 years old comprise questions addressing symptoms related to mood, behavior issues, and sleep disturbance. It is completed by the parent or primary caregiver.

The CBCL 1.5-5 years old is used in this study for children aged 5 to < 6 years old. It consists of 100 questions, scored on a three-point Likert scale (0=Not True (as far as you know), 1= Somewhat or Sometimes True, 2=Very True or Often True). The time frame for item responses is the past 2 months.

The CBCL 6-18 years old is used in children aged 6 up to 18 years old. It consists of 113 questions, scored on a three-point Likert scale (0=Not True (as far as you know), 1= Somewhat or Sometimes True, 2=Very True or Often True). The time frame for item responses is the past six months.

Note: No data is collected from the language scale, as it is supplemental and not required to calculate any of the behavior measures.

The checklist yields scores in the following areas:

CBCL 1.5-5	CBCL 6-18
• Emotionally reactive • Anxious/Depressed • Withdrawn • Somatic Complaints • Attention Problems • Aggressive Behavior	• Anxious/Depressed • Withdrawn/Depressed • Somatic Complaints • Social problems • Thought Problems • Attention Problems • Rule Breaking Behavior

The table in Appendix 23.1 shows how the domains are constructed with regards to the individual questions. The scores for each individual question within a domain are summed to give a domain score.

Each domain score above and total score will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) by treatment group at Baseline, Week 26 and Week 52, as will change from baseline at Week 26 and Week 52 (where baseline measures exist).

Where possible, the CBCL 1.5-5 and CBCL 6-18 will be summarized together. If a subject progresses during the course of the study from one version to the next the results will be summarized together and the baseline assessment from the child's first questionnaire (1.5-5) will be referred to when summarizing change from baseline.

Domain scores and total score will be listed by treatment group and subject number.

15.6 Echocardiogram

Echocardiogram results are only collected at the Screening visit, and will be listed by treatment group and subject.

15.7 On-Study Procedures, Interventions and Surgeries

Any procedures, interventions of surgeries that occur on study (post first dose of study drug) will be captured, along with start and stop date. These will be coded using the latest available MedDRA version and details provided in a listing.

15.8 Hip Monitoring and Rotation

All hip monitoring clinical assessments, including hip rotation, will be listed by treatment group, subject and assessment date.

15.9 Pregnancy Testing

A pregnancy test will be performed at the following visits for female subjects who have begun menses or are ≥10 years of age; Screening, Day 1, Week 6, Week 13, Week 26, Week 39, Week 52, Early Termination and at the Safety Follow-up visit.

PI .

16 IMAGING SECONDARY ENDPOINT(S)

All summaries of imaging data will be based on the Safety Population. All visits referred to in the analyses, are based on the analysis windows described in Section 5.4. Therefore, where summaries are based on data collected at the Screening visit, they will be referred to as Baseline assessments once the analysis window is applied.

16.1 Bone Age

An X-ray of the left hand and wrist will be performed at Screening, Week 26 and Week 52, and bone age determined using the Greulich and Pyle Atlas.

Bone age at Baseline, Week 26 and Week 52, and change from baseline at Week 26 and Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) by treatment group and presented by sex and overall.

Each measurement of bone age will be converted to an age-and sex-appropriate standard deviation score (SDS), also referred to as a Z-score, by comparison with reference data available for average stature children (Greulich and Pyle, 1959). Bone age Z-scores at Baseline, Week 26 and Week 52, and change from baseline at Week 26 and Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group.

In addition, for the visits at Baseline and Week 52, scatter plots will be provided which display bone age z-score on the y-axis, and age in years at that visit on the x-axis, identifying the two treatment groups. A separate plot will be generated for each sex and overall.

16.2 Lower Limb X-Rays

For each of the following parameters, excluding femur length, the absolute values for both the left and right leg at Baseline, Week 26 and Week 52, and change from baseline at Week 26 and Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group for each baseline age category (≥ 5 to < 8, ≥ 8 to < 11, ≥ 11 to < 15, ≥ 15 to < 18 years). As the assessment of femur length was introduced mid-study, subjects will not have a measurement of femur length at baseline, and data will be listed only.

- Left/Right Femur length (cm)
- Left/Right Tibia length (cm)
- Left/Right Fibula length (cm)

- Left/Right Tibia bowing angle (degrees)
- Left/Right Distance between ankle joint and distal growth plate of fibula (cm)

In addition, change from baseline in tibial length (y-axis) at Week 52 will be presented on a boxplot, with a separate box for each leg and treatment group with baseline age category (≥ 5 to < 8, ≥ 8 to < 11, ≥ 11 to < 15, ≥ 15 to < 18 years) on the x-axis. A similar plot will also be generated for change from baseline in fibula length at Week 52.

All assessments will be listed by treatment group, subject and visit.

16.3 Lumbar Spine X-Rays

Lumbar spine x-rays (AP and lateral views) are performed at Screening and Week 52. Assessments will be listed by treatment group, subject and visit.

16.3.1 Vertebral Height

Vertebral heights (anterior, medial, and posterior) will be measured for each individual vertebra (L1, L2, L3, L4, and L5).

The following ratios will be calculated for each individual vertebra (L1, L2, L3, L4, and L5) and reported to 1 decimal place:

- Anterior height (cm) to medial height (cm)
- Anterior height (cm) to posterior height (cm)
- Medial height (cm) to posterior height (cm)

For each of the above ratios, the absolute values at Baseline and Week 52, and change from baseline at Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group.

16.3.2 Transverse Diameter (Interpedicle Distance)

The transverse diameter (interpedicle distance) will be measured for each individual vertebra (L1, L2, L3, L4, and L5) at Baseline and Week 52.

For each vertebra, the absolute values at Baseline and Week 52, and the change from baseline at Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group for each baseline age category (≥ 5 to < 8, ≥ 8 to < 11, ≥ 11 to < 15, ≥ 15 to < 18 years).

In addition, change from baseline in transverse diameter (y-axis) for each individual vertebra (x-axis) will be presented on a boxplot at Week 52 for each treatment group. Separate plots

will be generated for each baseline age category (≥ 5 to < 8, ≥ 8 to < 11, ≥ 11 to < 15, ≥ 15 to < 18 years).

16.3.3 Sagittal Width

The spinal canal width (sagittal width) will be measured for each individual vertebra (L1, L2, L3, L4, and L5) at Baseline and at Week 52.

For each individual vertebra, the absolute values at Baseline and Week 52, and the change from baseline at Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group for each baseline age category (≥ 5 to < 8, ≥ 8 to < 11, ≥ 11 to < 15, ≥ 15 to < 18 years).

In addition, change from baseline in sagittal width (y-axis) for each individual vertebra (x-axis) will be presented on a boxplot at Week 52 for each treatment group. Separate plots will be generated for each baseline age category (≥ 5 to < 8, ≥ 8 to < 11, ≥ 11 to < 15, ≥ 15 to < 18 years).

16.3.4 Lumbar Spine Angles

The following angles will be measures on the lumbar spine x-rays:

- Sacral tilt (degrees)
- Lordosis (inward curve of the spine) (degrees)
- Kyphosis (convex curvature of the spine) (degrees)

Each angle will be measured on the spine x-rays and compared to Baseline at Week 52 to determine any worsening of sacral tilt, lordosis (inward curve of the spine) or kyphosis (convex curvature of the spine).

The absolute values of the sacral tilt, lordosis and kyphosis angles at Baseline and Week 52, and change from baseline at Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group for each baseline age category (≥ 5 to < 8, ≥ 8 to < 11, ≥ 11 to < 15, ≥ 15 to < 18 years).

The following summaries will similarly be presented for each baseline age category (≥ 5 to < 8, ≥ 8 to < 11, ≥ 11 to < 15, ≥ 15 to < 18 years):

A worsening in sacral tilt angle will be summarized [n(%)] by treatment group as follows:

- Percentage of subjects with an increase in sacral tilt angle of ≥5 to < 10 degrees at week 52
- Percentage of subjects with an increase in sacral tilt angle of ≥ 10 degrees at week 52

A worsening in lumbar lordosis will be summarized [n(%)] by treatment group as follows:

- Percentage of subjects with an increase in lordosis angle of ≥5 to < 10 degrees at week 52
- Percentage of subjects with an increase in lordosis angle of ≥ 10 degrees at week 52

A worsening in kyphosis angle will be summarized [n(%)] by treatment group as follows:

- Percentage of subjects with an increase in kyphosis angle of ≥5 to < 10 degrees at week 52
- Percentage of subjects with an increase in kyphosis angle of ≥ 10 degrees at week 52

16.4 Dual Energy X-ray Absorptiometry

A dual energy x-ray absorptiometry (DXA) scan is performed at Screening, Week 26 and Week 52, in order to collect relevant BMC/BMD data (including Z-scores) for whole body less head, and the lumbar spine. The scans of the distal forearm were removed from the original protocol at a later date, and hence data is limited.

All DXA data will be summarized separately for each scanner manufacturer (GE - Lunar Prodigy or Hologic – Discovery Horizon). Subjects who have results from more than one scanner type will be excluded from summaries, and their data will be listed only.

16.4.1 Whole Body Less Head and Lumbar Spine

The absolute values at Baseline, Week 26 and Week 52, and change from baseline at Week 26 and Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group for the following measures of bone mineral content (BMC) and bone mineral density (BMD):

- Whole Body Less Head BMC (g)
- Whole Body Less Head BMD (g/cm^2)
- Whole Body BMC (g)
- Whole Body BMD (g/cm^2)
- Lumbar Spine BMC (g)
- Lumbar Spine BMD (g/cm^2)

BMD Z-scores will be provided in the DXA reports, and the absolute values at Baseline, Week 26 and Week 52, and change from baseline at Week 26 and Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group and scanner for each of the following:

- Whole Body Less Head BMD Z-Score
- Whole Body BMD Z-Score
- Lumbar Spine BMD Z-Score

The absolute values of total body percent fat, android percent fat, and gyenoid percent fat at Baseline, Week 26 and Week 52 will be listed.

Box plots of whole body less head BMD Z-scores will be provided by treatment group, and scanner.

16.4.2 Regions of Interest in Distal Forearm

Due to sparsity of data, available data will be listed by treatment group, subject and visit. No summaries will be presented.

17 BONE METABOLISM BIOMARKERS

Bone metabolism biomarkers will be collected to assess changes in bone metabolism. The safety population will be used for all summaries.

Samples for bone metabolism blood biomarkers (collagen X and bone-specific alkaline phosphatase [BSAP]) are collected pre-dose at Day 1, Week 13, Week 26, Week 39, Week 52 and at the Safety Follow-up visit.

For each biomarker, the absolute values, and change from baseline at Week 13, Week 26, Week 39, and Week 52 will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group.

For each biomarker, the change from baseline will also be summarized by visit in a box plot and presented by treatment group.

Bone metabolism biomarkers will be listed by treatment group and subject.

18 BMN 111 ACTIVITY URINE BIOMARKERS AND URINE CHEMISTRY

Samples for BMN 111 activity urine biomarker, cyclic guanosine monophosphate [cGMP], and urine chemistry are collected pre-dose and 1, 2 and 4hr post-dose at Day 1, Day 2, Day 3, Week 6, Week 26, and Week 52. Each collection will be tested for biomarker concentration and urine creatinine concentration for normalization. The safety population will be used for all summaries.

The absolute values for pre-dose urine cGMP normalized by creatinine and urine chemistry (creatinine, potassium, sodium) results at each scheduled visit, and change from pre-dose to multiple post-dose time points at each scheduled visit will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group.

The absolute values for pre-dose urine cGMP normalized by creatinine and urine chemistry (creatinine, potassium, sodium) results at each scheduled visit, and the maximum change from pre-dose to post-dose will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) and presented by treatment group at each scheduled visit and overall.

The maximum change from pre-dose cGMP normalized by creatinine will also be summarized in a box plot by visit and presented by treatment group.

BMN 111 activity urine biomarkers will be listed by treatment group and subject.

19 IMMUNOGENICITY ASSESSMENT

Serum samples for anti-drug antibody (ADA) tests will be drawn pre-dose on Day 1, Week 13, Week 26, Week 39, Week 52 and the Safety Follow-up or Early Termination visit. Blood (serum) samples for testing drug-specific IgE will be drawn on Day 1 and in the event of a Grade 3 or significant hypersensitivity adverse event, or at the investigator's discretion.

ADA tests will include anti-BMN 111 total antibody (TAb); TAb cross-reactive with endogenous CNP, BNP or ANP; and anti-BMN 111 neutralizing antibody (NAb).

NAb testing and TAb testing for cross-reactivity with endogenous CNP, ANP or BNP will be performed only on baseline and TAb positive samples. NAb testing will not be performed if the TAb is negative. A listing for each ADA test will be provided.

Summary tables will include all safety events up to 30 days following treatment discontinuation. PI .

The immunogenicity population will be used for all summaries. The data conversion rules for immunogenicity analysis are listed below:

Operational Data Conversion Table for Immunogenicity Analysis

Assay	Result Type	Result = "Concentration (Titer Units)"	Pos/Neg for Incidence Table	Numerical Value for Display	Numerical Value for Calculation
Anti-BMN111 TAb	Numeric Titer	Negative Screen	Negative	1	0
		Negative Immunodepletion	Negative	1	0
		Negative Titer (≤10)	**Negative**	**1**	**0**
		* Value * (e.g. 20, 30, …)	**Positive**	**e.g. 20, 30, …**	**e.g. 20, 30, …**
		Imputed values - none			
		MRD of assay: 10			
Anti-BMN111 NAb	Numeric Titer	Negative Screen	Negative	1	0
		Negative Immunodepletion	Negative	1	0
		Negative Titer (≤5)	**Negative**	**1**	**0**
		* Value * (e.g. 10, 20, …)	**Positive**	**e.g. 10, 20, …**	**e.g. 10, 20, …**
		Imputed values – if TAb negative then impute result as 'Negative' for same study visit			
		MRD of Assay: 5			
ANP Reactivity	Binary	Negative Screen	Negative		
		Negative Immunodepletion	Negative		
		Positive Immunodepletion	**Positive**		
		Imputed values – if TAb negative then impute result as 'Negative' for same study visit			
BNP Reactivity	Binary	Negative Screen	Negative		
		Negative Immunodepletion	Negative		
		Positive Immunodepletion	**Positive**		
		Imputed values – if TAb negative then impute result as 'Negative' for same study visit			
CNP-22 Reactivity	Binary	Negative Screen	Negative		
		Negative Immunodepletion	Negative		
		Positive Immunodepletion	**Positive**		
		Imputed values – if TAb negative then impute result as 'Negative' for same study visit			

The incidence [n(%)] of TAb titer positive results will be summarized for each scheduled visit, and overall (described as 'Ever Positive') as follows:

$$\frac{Number\ of\ subjects\ with\ a\ positive\ test\ result}{Number\ of\ subjects\ with\ non-missing\ TAb\ result} * 100$$

The incidence [n(%)] of NAb positive test results, and positive cross-reactivity results for ANP, BNP and CNP, will also be calculated based on the number of subjects with a non-missing TAb result, and summarized at each scheduled visit and overall.

The absolute values (numerical values) of TAb titers and NAb titers will be summarized (n, mean, SD, median, 25th and 75th percentile, minimum, maximum) at Day 1, Week 13, Week 26, Week 39 and Week 52, and presented by treatment group.

The mean (SD) TAb titer, number of hypersensitivity adverse events (HAEs) excluding injection site reactions, and number of subjects experiencing HAE's will be summarized by the highest HAE CTCAE Grade, and presented by TAb status (negative or positive) and treatment group.

A line graph of the mean TAb titer (+/- SD) will be presented over time for each visit by treatment group. A similar figure will be generated for NAb titers.

The change from baseline in AGV at Week 52 will be plotted against the mean TAb titer for each subject, where the mean is calculated across all post-baseline TAb titers for each subject. The Pearson correlation coefficient will be included in the figure to describe the association.

The change from baseline in AGV at Week 52 will be plotted against the mean NAb titer for each subject, where the mean is calculated across all post-baseline NAb titers for each subject. The Pearson correlation coefficient will be included in the figure to describe the association.

The relationship between immunogenicity (ADA status: negative or positive) and measures of safety and efficacy will further be presented graphically by treatment group:

- Box plot of change from baseline in AGV at week 52 by TAb status
- Box plot of change from baseline in AGV at week 52 by NAb status
- Box plot of number of hypersensitivity adverse events by TAb status
- Box plot of number of hypersensitivity adverse events (excluding ISRs) by TAb status
- Box plot of maximum severity hypersensitivity adverse event by TAb status
- Box plot of number of injection site reaction adverse events by TAb status
- Box plot of maximum duration of injection site reaction adverse event by TAb status

In addition, the following listings will be provided:

- Total antibody titers and neutralizing antibody titers
- Hypersensitivity adverse events with antibody results
- Drug-specific IgE, total IgE, C4 and serum tryptase for hypersensitivity reaction visits
- TAb cross reactivity to endogenous natriuretic peptides

- TAb cross reactivity by subject and cardiac disorder or HLGT ‘fluid and electrolyte imbalance’ related adverse event.
- Mean antibody titers and AGV

All immunogenicity listings will be ordered by treatment group, 4-digit subject number (i.e., the subject ID, excluding the site number) and visit.

20 PHARMACOKINETICS AND PHARMACODYNAMICS

BMN 111 concentrations and pharmacokinetic (PK) parameters will be summarized descriptively by visit for all subjects in the PK population. If supported by the data, the exposure-response relationship between BMN 111 exposure and immunogenicity, efficacy, biomarker, and safety pharmacodynamic (PD) endpoints of interest will be explored. Details of the PK/PD analyses are separately documented in the Clinical Pharmacology Analysis Plan.

21 REFERENCES

Bullinger M, Sommer R, Pleil A, et al. Evaluation of the American-English Quality of Life in Short Stature Youth (QoLISSY) questionnaire in the United States, Health Qual Life Outcomes, 2015: 13:43.

Greulich, W and Pyle, S. Radiographic Atlas of Skeletal Development of the Hand and Wrist, 1959: Stanford University Press.

Hoover-Fong, JE, Schulze, KJ, McGready, J, Barnes, H et. al. Age-appropriate body mass index in children with achondroplasia: interpretation in relation to indexes of height. Am J ClinNutr 88[2], 364-371. 2008.

Little, R. J. A., "Pattern-Mixture Models for Multivariate Incomplete Data," Journal of the American Statistical Association, 1993; 88, 125–134.

Molenberghs, G. and Kenward, M. G. (2007), Missing Data in Clinical Studies, New York: John Wiley & Sons.

Varni JW, et al. The PedsQL™: Measurement Model for the Pediatric Quality of Life Inventory. Medical Care, 1999; 37(2):126-139

22 SUMMARY OF CHANGES TO STUDY SAP

Version		Affected Section (Number of Section in Previous SAP)	Summary of Revisions
Number	Date		
Final SAP	05-Nov-19	1.0 (NA)	Signature page added.
Final SAP	05-Nov-19	4.6.2 (2.6.2)	The Japanese randomization has been described in more detail.
Final SAP	05-Nov-19	5.1.2 (3.1.2)	Gonadotropin-releasing hormones have been added as an exclusion from the Pre Protocol analysis population.
Final SAP	05-Nov-19	5.4 (3.4)	111-901 analysis visit windows have been updated.
Final SAP	05-Nov-19	5.4 (3.4)	A spaghetti plot of standing height Z-scores has been added.
Final SAP	05-Nov-19	8.0 (6.0)	A listing of major deviations classified as “dosing irregularity” has been added.
Final SAP	05-Nov-19	9.0 (7.0)	Baseline characteristics growth measures are re-ordered. Tanner Stages, other than (I,>I) are not included in baseline summary.
Final SAP	05-Nov-19	10.0 (8.0 and Appendix)	Details of Achondroplasia-related terms are removed from the SAP appendix, and located in a study file.
Final SAP	05-Nov-19	10.0 (8.0)	Physical exam results listing added.
Final SAP	05-Nov-19	11.0 (9.0)	The concomitant medications are to be summarized by ATC level 2, not 4.
Final SAP	05-Nov-19	13.0 (11.0)	Definition for missed dose has been added. Dose interruptions have been replaced with a summary of “Reasons for missed dose”.
Final SAP	05-Nov-19	14.0 (12.0)	Estimands have been clarified for the primary and key secondary endpoints. Details have been added to describe imputation methods for missing Week 52 assessments for standing height, AGV, height Z-Score, sitting height and upper-to-lower body segment ratio.
Final SAP	05-Nov-19	14.2 (12.2)	Terminology for the testing strategy has been changed from “statistically significant “ to “confirmatory”.
Final SAP	05-Nov-19	14.3.3 (12.3.3)	There is now one sensitivity analysis and 2 supplementary analyses: A washout model has been introduced as the sensitivity analysis The completers analysis is now a supplementary analysis (not sensitivity). The analysis excluding data post limb lengthening/growth hormone is now a supplementary analysis (not sensitivity). The random coefficient model sensitivity analysis has been removed.
Final SAP	05-Nov-19	14.7 (12.7)	More detail has been added for subgroup analyses, and subgroups of height Z-score thresholds and AGV thresholds have been added
Final SAP	05-Nov-19	15.1.1 (13.1.1)	A table of the incidence of TEAE’s by SOC, PT and highest CTCAE grade has been added.
Final SAP	05-Nov-19	15.1.1 (13.1.1)	The Achondroplasia-related AE table is now exposure-adjusted
Final SAP	05-Nov-19	15.1.2 (13.1.2)	Events of Special Interest are now Events of Interest.
Final SAP	05-Nov-19	15.1.2 (13.1.2)	Events suggestive of hypotension are removed. Profile of events suggestive of hypotension, and profile of hypersensitivity (SMQ) tables have been removed.
Final SAP	05-Nov-19	15.1.2 (13.1.2)	Incidence and exposure-adjusted ISR tables by duration of ISR have been added.
Final SAP	05-Nov-19	16.1 (14.1)	Bone Age presented by sex and overall. Plots of Z-scores.
Final SAP	05-Nov-19	16.2 (14.2)	Lower limb x-rays: Left and Right leg assessments included. Percent change from baseline removed.

Final SAP	05-Nov-19	16.2 (14.2)	Lumbar spine x-rays: Percent change from baseline removed.
Final SAP	05-Nov-19	16.4 (14.4)	DXA data will be presented separately for each scanner (GE or Hologic). Summaries of fat removed.
Final SAP	05-Nov-19	17.0 (15.0)	Biomarker plots added.
Final SAP	05-Nov-19	18.0 (16.0)	Urine chemistry parameters added.

23 APPENDICES

23.1 Child Behavior Checklist Domains

Domain/Subscale	CBCL 1.5-5	CBCL 6-18
EMOTIONALLY REACTIVE	21. Disturbed by change 46. Twitches 51. Panics 79. Shifts between sad-excite 82. Sudden mood change 83. Sulks a lot 92. Upset by new 97. Whining 99. Worries	
ANXIOUS/DEPRESSED	10. Too dependent 33. Feelings easily hurt 37. Upset when separated 43. Looks unhappy 47. Nervous 68. Self-conscious 87. Fearful 90. Unhappy, sad, depressed	14. Cries a lot 29. Fears 30. Fears school 31. Fears doing bad 32. Must be perfect 33. Feels unloved 35. Feels worthless 45. Nervous, tense 50. Fearful, anxious 52. Feels too guilty 71. Self-conscious 91. Talks or thinks of suicide 112. Worries
WITHDRAWN/DEPRESSED		5. Enjoys a little 42. Rather be alone 65. Won't talk 69. Secretive 75. Shy, timid 102. Lacks energy 103. Sad 111. Withdrawn
WITHDRAWN	2. Acts too young 4. Avoids eye contact 23. Doesn't answer 62. Refuses active games 67. Unresponsive to affection 70. Little affection 71. Little interest 98. Withdrawn	
SOMATIC COMPLAINTS	1. Aches, pains 7. Can't stand things out of place 12. Constipated 19. Diarrhea 24. Doesn't eat well 39. Headaches 45. Nausea 52. Painful bowel movements	47. Nightmares 49. Constipated 51. Feels dizzy 54. Overtired 56a. Aches, pains 56b. Headaches 56c. Nausea 56d. Eye problems

	78. Stomachaches 86. Too concerned with neatness 93. Vomits	56e. Skin problems 56f. Stomachaches 56g. Vomiting
SOCIAL PROBLEMS		11. Too dependent 12. Lonely 25. Doesn't get along 27. Jealous 34. Others out to get him/her 36. Accident prone 38. Gets teased 48. not liked 62. Clumsy 64. Prefers younger kids 79. Speech problems
THOUGHT PROBLEMS		9. Can't get mind off thoughts 18. Harms self 40. Hears things 46. Twitching 58. Picks skin 59. Sex parts in public 60. Sex parts too much 66. Repeats acts 70. Sees things 76. Sleeps less 83. Stores things 84. Strange behavior 85. Strange ideas 92. Sleep talks/walks 100. Trouble sleeping
ATTENTION PROBLEMS	5. Can't concentrate 6. Can't sit still 56. Clumsy 59. Quickly shifts 95. Wanders away	1. Acts young 2. Fails to finish 8. Can't concentrate 10. Can't sit still 13. Confused 17. Daydreams 41. Impulsive 61. Poor schoolwork 78. Inattentive 80. Stares blankly
RULE BREAKING BEHAVIOR		2. Drinks alcohol 26. Lacks guilt 28. Breaks rules 39. Bad friends 43. Lies, cheats 63. Prefers older kids 67. Runs away 72. Sets fires 73. Sex problems 81. Steals at home 82. Steals outside home 90. Swearing

		96. Thinks of sex too much 98. Tardy 99. Uses tobacco 101. Truant 105. Uses drugs 106. Vandalism
AGGRESSIVE BEHAVIOR	8. Can't stand waiting 15. Defiant 16. Demands must be met 18. Destroys others things 20. Disobedient 27. Lacks guilt 29. Easily frustrated 35. Gets in fights 40. Hits others 42. Hurts unintentionally 44. Angry moods 53. Attacks people 58. Punishment doesn't change behavior 66. Screams 69. Selfish 81. Stubborn 85. Temper 88. Uncooperative 96. Wants attention	3. Argues a lot 16. Mean to others 19. Demands attention 20. Destroys own things 21. Destroys others' things 22. Disobedient at home 23. Disobedient at school 37. Gets in fights 57. Attacks people 68. Screams a lot 86. Stubborn, sullen 87. Mood changes 88. Sulks 89. Suspicious 94. Teases a lot 95. Temper 97. Threatens others 104. Loud