STATISTICAL ANALYSIS PLAN

A Phase 2, Randomized, Double-Blind, Placebo-Controlled Efficacy and Safety Study of Palovarotene in Subjects with Multiple Osteochondromas

Study PVO-2A-201 (MO-Ped Trial)

Sponsor:	Clementia Pharmaceuticals Inc. 1000 De La Gauchetière West, Suite 1200 Montreal, QC H3B 4W5 Canada
Sponsor Representative:	PPD
Protocol Version/ Date	Amendment 2: 23 April 2019
Document Date:	14 May 2020
Document Version:	1.0

STATISTICAL ANALYSIS PLAN

A Phase 2, Randomized, Double-Blind, Placebo-Controlled Efficacy and Safety Study of Palovarotene in Subjects with Multiple Osteochondromas

Study PVO-2A-201 (MO-Ped Trial)

Cytel Author/Date:

_____

PPD

Sponsor Author/Date:

_____

PPD

Sponsor Approver/Date:

_____

PPD

_____

PPD

TABLE OF CONTENTS

ABBREVIATIONS

Abbreviation	Definition
aBMD	areal bone mineral density
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ATC	Anatomic Therapeutic Class
$AUC_{0\text{-}24ss}$	area under the plasma concentration-time curve at steady-state from time zero to 24 hours
BMC	bone mineral content
BMI	body mass index
CI	confidence interval
CL/F	apparent total clearance of study drug from plasma after oral administration
$C_{max,\ ss}$	maximum plasma concentration, steady state
$C_{min,\ ss}$	trough plasma concentration, taken 24 hours after dose and prior to subsequent dose, steady state
C-SSRS	Columbia-Suicide Severity Rating Scale
CTCAE	Common Terminology Criteria for Adverse Events
DMC	data monitoring committee
DXA	dual x-ray absorptiometry
ECG	electrocardiogram
EOS	end of study visit
EOT	end of treatment visit
EoTS	end of trial study
Ext	exostosin gene
FAS	Full Analysis Set
FDA	United States Food and Drug Administration
FOP	Fibrodysplasia Ossificans Progressiva
FPS-R	Faces Pain Scale – Revised
LDFA	lateral distal femoral angle
LDTA	lateral distal tibial angle
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MI	myocardial infarction
MO	Multiple Osteochondromas
MPTA	medial proximal tibial angle
MRI	magnetic resonance imaging
OC	osteochondroma
PCSA	potentially clinically significant abnormality
PedsQL	Pediatric Quality of Life Inventory
PPC	premature physeal closure
PROMIS	Patient-Reported Outcomes Measurement Information System

Abbreviation	Definition
PKS	Pharmacokinetic Set
PT	Preferred Term
SAP	statistical analysis plan
SD	standard deviation
SOC	System Organ Class
TEAE	treatment-emergent adverse event
$T_{max, ss}$	time to maximum plasma concentration, steady state
λ_z	Elimination rate constant
$t_{½z}$	Elimination half-life
ULN	upper limit of normal
WHO-DD	World Health Organization Drug Dictionary

1 OVERVIEW AND INVESTIGATIONAL PLAN

1.1 Study design

Study PVO-2A-201 is a Phase 2, multicenter, randomized, double-blind, placebo-controlled study designed to evaluate the efficacy and safety of two dosage regimens of palovarotene in pediatric subjects with Multiple Osteochondromas (MO). Approximately 240 subjects were planned to be randomized in a 1:1:1 ratio to one of two treatments of weight-adjusted dose equivalents of 2.5 or 5.0 mg palovarotene or a placebo, administered orally, once daily for up to 24 months. Randomization was stratified by age (≤7 years, >7 years), by sex (M, F), and by exostosin (Ext) gene mutation (Ext1, Ext2).

Effective 04 December 2019, all recruitment and dosing in Study PVO-2A-201 stopped, due to a partial clinical hold instituted by the FDA. The partial clinical hold was a result of concerns about premature physeal closure (PPC) reported in studies of palovarotene in Fibrodysplasia Ossificans Progressiva (FOP). As of 04 December 2019, 193 subjects had been randomized and dosed in Study PVO-2A-201, and 28 subjects had completed their 12-month efficacy assessments. The sponsor informed sites on 24 March 2020 that Study PVO-2A-201 was going to be terminated early. Subjects are to be followed for safety assessment without dosing. Because of the institution of the partial clinical hold and subsequent decision to terminate the study, as well as the ongoing COVID-19 pandemic and effect on the ability of subjects to come to study sites for assessments, planned end of treatment assessments have not been consistently performed. The plan is for subjects to return to the sites for their planned EOS assessments.

This Statistical Analysis Plan (SAP) was developed based on the protocol Amendment 2 dated 23 April 2019, and also on the changes in study procedures due to early termination of the study. As of the approval date of this SAP, the sponsor remains blinded to treatment assignments.

An interim efficacy analysis (to declare early efficacy or futility) and a final analysis were planned, but only a modified interim analysis – to support the interim CSR to be used in the palovarotene FOP NDA – and modified final analysis will be performed, due to early termination of the study. The statistical analyses and summary tabulations described in this SAP will provide the basis for the results sections of the MO-Ped clinical study report.

The primary efficacy analysis will compare the effect of two palovarotene dosage regimens with placebo on the annualized rate of new osteochondromas (OCs) over a maximum treatment period, which could be either 12 or 24 months. Efficacy data at other timepoints, if collected, may also contribute to the modified final analysis.

Note that, although the Study PVO-2A-201 protocol references an open-label extension study (Study PVO-2A-202), the scope of this statistical analysis plan is restricted to data collected during Study PVO-2A-201; due to the early termination, no subject entered the open-label extension.

1.2 Objectives

1.2.1 Primary Objective

The primary study objective is to evaluate the efficacy of two palovarotene dosage regimens compared with placebo in preventing new OCs in subjects with MO due to Ext1 or Ext2 gene mutations.

1.2.2 Secondary Objectives

Secondary objectives are to compare the following effects of palovarotene with placebo:

- The volume of OCs as assessed by magnetic resonance imaging (MRI).
- The proportion of subjects with no new OCs as assessed by whole body MRI.
- The annualized rate of new or worsening skeletal deformities.
- The annualized rate of MO-related surgeries.

Additional secondary objectives are to evaluate overall palovarotene safety, the pharmacokinetics of palovarotene at steady-state, and the palatability of drug product when sprinkled onto specific foods.

1.2.3 Exploratory Objectives

Exploratory objectives are to compare the following effects of palovarotene with placebo:

- The change in volume of OC cartilage caps as assessed by MRI.
- The annualized rate of new or worsening functional limitations.
- Pain and pain interference due to OCs.
- Quality of life.

1.3 Determination of sample size

The sample size required for the MO-Ped Trial was initially determined by simulation. The annualized rate of new OCs was assumed to be 1.15 based on analyses of the Instituto Orthopedico Rizzoli Registry for Multiple Exostoses (Section 4.5 of protocol). Osteochondroma counts were assumed to follow a negative binomial distribution parameterized with a variance inflation factor of 2.0.

1.4 Study plan

The schedule of the safety, efficacy, and pharmacokinetic assessments can be found in Table 1 of the study protocol.

1.5 Modifications to the statistical section of the protocol

The protocol (Original Protocol, Version 2, and Amendment 2) incorrectly lists the one-sided p-value significance threshold used in the interim analysis comparing the 5.0-mg palovarotene and placebo groups and the 2.5-mg palovarotene and placebo groups as 0.007. The correct value of 0.00073 is used in this SAP. Due to study early termination, the stopping criteria described in the protocol are no longer applicable.

1.6 Statistical modifications from Protocol

The study was terminated early with substantially reduced sample size and study duration. Power for efficacy comparisons is substantially reduced, and there should be no expectation of observing statistically significant findings from efficacy analyses. In lieu of formal hypothesis testing, efficacy analyses will be presented with point estimates and associated confidence intervals. For exploratory purposes, the two treatment groups will be compared with the placebo group for primary and key secondary/exploratory endpoints, and no multiple-testing adjustments on the p-value will be implemented.

2 STATISTICAL AND ANALYTICAL PROCEDURES

Data analyses will be performed for the MO-Ped Trial study period. The results will be presented for four groups (two palovarotene dose groups, total palovarotene group, and placebo group) for both safety and efficacy analyses.

2.1 Analysis endpoints

2.1.1 Demographic and baseline characteristics

Definition of baseline

Baseline is generally defined as the last available value prior to first administration of study drug, except for endpoints relying on the following assessments: whole body MRI, radiographs of upper/lower limbs, dual x-ray absorptiometry (DXA), and hearing and visual acuity tests. As specified in the protocol Schedule of Assessments, these "Day 1" imaging and sensory assessments may be performed within 4 weeks of other Day 1 procedures, including initiation of study drug treatment, to allow flexibility for scheduling. Because these assessments characterize endpoints for efficacy and safety that change slowly over the course of months, the baseline values measured shortly after the initiation of treatment should not markedly differ from values measured before administration of palovarotene. Therefore, endpoints based on these assessments will use "Day 1" as baseline, even if after the first dose is administered.

Demographic characteristics

The following demographic characteristics will be summarized using descriptive statistics:

- Sex (male, female)
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Age (years and months)
- Age (≤7 years, >7 years)

The height z-scores will be calculated by using US Centers for Disease Control and Prevention (CDC) growth charts, using the SAS code from

https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm.

Medical or surgical history

This information will be coded using a current or recent version of the Medical Dictionary for Regulatory Activities (MedDRA). Subjects' medical history will also be summarized.

MO history

The following MO history will be summarized by the following categories:

- Age at diagnosis (years)
- Time since diagnosis (years)
- Other family members with MO (mother, father, siblings)
- MO-associated mutation
- MO-associated mutation type
- MO surgeries performed (yes/no)
- Body location of MO surgeries performed
- Number of prior MO-related surgeries

2.1.2 Prior or concomitant medications

All medications taken during the screening period and until the end of the study are to be reported in the case report form pages.

All medications will be coded using the using the most current version of the World Health Organization– Drug Dictionary (WHO-DD) by Anatomic Therapeutic Class (ATC) class 3 and the preferred drug names. If the ATC class 3 is not available, ATC class 2 will be used instead.

Prior medications are those taken during the screening period. Prior medications can be discontinued before first dose of study drug or can be ongoing during the treatment phase.

Concomitant medications are any treatments received after the first dose of study drug, no matter whether the start date was prior to or after the first dose.

If a medication started prior to the first dose and continued after the first dose, it will be categorized as both prior medication and concomitant medication.

2.1.3 Efficacy endpoints

2.1.3.1. Primary efficacy endpoint

The primary efficacy endpoint is the annualized rate of new OCs as assessed by whole body MRI (i.e., the total number of new OCs divided by the time in years between the baseline and latest post-baseline MRI). Whole body MRI is performed at baseline and at Months 12 and 24, and at early termination for subjects who are withdrawn or discontinue early. The MRI may be performed within ±4 weeks of a study visit to allow flexibility of scheduling. All images will be interpreted by a central imaging laboratory, as described in the imaging read charter.

The total number of new OCs will be the count of OCs identified in the post-baseline MRI scans that were not present in the baseline scan. Documented counts of OCs that have been surgically

removed will be carried forward and added to the counts of OCs identified in the post-baseline scans.

Both total number of new OCs and annualized new OCs will be summarized. The annualized new OCs will be calculated using the following formula:

$$li \qquad Cs = \frac{To \qquad Cs}{of\ last\ MR\ \ sc \quad - \quad of\ firs\ \ st\ \ dy\ dr \quad + 1} * 365.25$$

2.1.3.2. Secondary efficacy endpoints

The following secondary efficacy endpoints will be assessed:

- The change from baseline in the total volume of OCs as assessed by whole body MRI at Months 12 and 24.
- The proportion of subjects with no new OCs as assessed by whole body MRI at Months 12 and 24.
- The annualized rate of new or worsening deformities as assessed by radiographic imaging of both upper and lower limbs.
- The annualized rate of MO-related surgeries. Surgeries include any procedure indicated for the treatment of MO, such as an excision of a symptomatic OC or correction of a limb deformity.

Change from baseline in total volume of OCs

The change from baseline in the total volume of OCs is the difference between the total volume of OCs at Months 12 and 24 and the total volume of OCs at baseline, as assessed by whole body MRI.

Volume of OCs that have been surgically removed will be carried forward and added to the volumes based on whole body MRI for the post-surgery timepoints.

Proportion of subjects with no new OCs

The proportion of subjects with no new OCs will be assessed by whole body MRI at Months 12 and 24.

Subjects with new OCs not identified by MRI due to surgical resection during the treatment period will be categorized as having new OCs for this analysis.

Annualized rate of new or worsening deformities

The annualized rate of new or worsening deformities will be derived using the same methods as described for the primary endpoint. Radiographs to assess limb deformity will be performed at baseline and at Months 12 and 24. Radiographs may be performed within ±4 weeks of a study

visit to allow flexibility of scheduling. All radiographs will be interpreted by a central imaging laboratory, as described in the imaging read charter.

The criteria for new limb deformity are listed below. Presence of any of these criteria constitutes a deformity.

- Relative ulnar length/height <14%
- Radial articular angle >30°
- Radiocapitellar articulation other than normal
- >5 degrees deviation from physiologic range for
    - Lateral Distal Tibial Angle (LDTA) (physiologic range 86-92°)
    - Medial Proximal Tibial Angle (MPTA) (physiologic range 85-90°)
    - Lateral Distal Femoral Angle (LDFA) (physiologic range 85-90°)
    - Femoral neck shaft angle (physiologic range 124-136°)
- Leg length discrepancy >2 cm

The criteria for worsening deformity (in previously deformed joints) are listed below. Presence of any of these criteria constitutes a deformity:

- Decrease in relative ulnar length/height from previous measurement by 1%
- Increase in radial articular angle by >5°
- Radiocapitellar articulation progression
    - Displaced→subluxed
    - Displaced→dislocated
    - Subluxed→dislocated
- >5° deviation from previous measurement for
    - LDTA
    - MPTA
    - LDFA
    - Femoral neck shaft angle
- Leg length discrepancy increase >1 cm/year from previous measurement

Deformities that have been surgically corrected will be considered as their pre-surgical state and carried forward.

Annualized rate of MO-related surgeries

The annualized rate of MO-related surgeries will be derived using the same methods as described for the primary endpoint. MO-related surgeries include any procedure indicated for the treatment of MO, which may include excisions to eliminate symptomatic OCs (pain, neurovascular, or tendon impingement) or procedures to correct deformities and/or functional limitations. To provide standards for assessing these surgical events for efficacy endpoints, corrections for

deformities of a ≥10° angle deviation, 2.5-cm leg length discrepancy, or restricted joint movement will be counted as an endpoint event.

Planned surgeries within 6 months after enrollment to remove symptomatic OCs present at baseline, or to correct deformities present at baseline, will be recorded but will not contribute to the endpoint. Similarly, surgical procedures that are a continuation of a previous procedure, such as removal of a surgical plate for guided growth, will not be counted as a separate MO-related surgery endpoint event.

2.1.3.3. Exploratory efficacy endpoints

The following exploratory efficacy variables will be assessed:

- The change from baseline in the total volume of cartilage caps of OCs as assessed by whole body MRI at Months 12 and 24.
- The annualized rate of new or worsening functional limitations. Functional limitations are defined as restrictions in joint range of motion.
- The effect of pain on daily activities, as assessed with the Patient Reported Outcomes Measurement Information System (PROMIS) pain interference pediatric item bank at Months 6, 12, 18, and 24.
- Pain intensity, as assessed with the Faces Pain Scale – Revised (FPS-R) at Months 6, 12, 18, and 24.
- Quality of life, as assessed with the Pediatric Quality of Life Inventory (PedsQL, version 4.0) at Months 6, 12, 18, and 24.

Change from baseline in total volume of cartilage caps

The change from baseline in the total volume of cartilage caps is the difference between the total volume of cartilage caps at Months 12 and 24 and the total volume of cartilage caps at baseline, as assessed by whole body MRI.

If surgery is performed after baseline, the last pre-surgery measurement of volume of the affected cartilage cap will be carried forward (ie, change from baseline in the affected cap is assumed to be 0).

Annualized rate of new or worsening functional limitations

The annualized rate of new or worsening function limitations will be derived using the same methods as described for the primary endpoint (Section 2.1.3.1). A functional limitation will be defined as a restriction in joint range of motion. Range of motion will be assessed by a treatment-blinded trained assessor using a goniometer at baseline and at Months 6, 12, 18, and 24. Whenever possible, the same assessor will be used to standardize the procedures and to minimize variability.

The following joint ranges of motion will be assessed:

- Hip extension and flexion
- Knee flexion and extension
- Shoulder flexion
- Elbow flexion, extension, supination, and pronation
- Ankle dorsiflexion and plantarflexion

A functional limitation is defined as a deviation >2 standard deviations below the mean range of motion per age category and gender versus normal individuals. Functional limitations, unless corrected by surgery, must persist for the duration of observation. Normative values (Section 6.3) are from individuals from 2 to 19 years of age as published in the Normal Joint Range of Motion Study conducted by the CDC.[1] A worsening restriction is defined as a further decrease of >5º or >20% of the CDC age category-by-gender mean range of motion.

PROMIS pain interference

The PROMIS pain interference items assess the extent to which pain hinders social, psychological, physical, recreational activities and sleep over the preceding 7 days. An age-appropriate form of the PROMIS pain interference short form is administered to subjects (or parent proxy); self-report forms are used for subjects 8 to 17 years of age (PROMIS Pediatric Item Bank v2.0 – Pain Interference – Short Form 8a) and parent proxy forms are used for subjects from 2 to 7 years of age (PROMIS Parent Proxy Item Bank v2.0 – Pain Interference – Short Form 8a).

The PROMIS pain interference short form consists of eight questions. Each question has five response options scored from 1 to 5. The total raw score will be calculated as the sum of the values of the response to each question. If more than four questions are unanswered, the total raw score will not be computed and will be recorded as missing. If four or fewer questions are unanswered, the total raw score will be set to the mean of the answered questions × 8. The computed total raw score will be transformed into a T-score for analysis as described in the Appendix (Section 6.1). The T-score rescales the raw score into a standardized score with a minimum value of 8 and maximum value of 40. Reference is at the following:

http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Pain_Interference_Scoring_Manual.pdf.

Faces Pain Scale – Revised

Pain will be further evaluated using the Faces Pain Scale – Revised (FPS-R), which is a pediatric subject self-report measure of acute pain intensity for subjects ≥4 years of age. The FPS-R was adapted from the original FPS to make it possible to score pain on a numeric rating scale from 0 (no pain) to 10 (very much pain).

Pediatric Quality of Life Inventory

The Pediatric Quality of Life Inventory (PedsQL, version 4.0) is a reliable and validated pediatric quality-of-life survey for children from the ages of 2 to 18 years, designed to measure core health

dimensions outlined by the World Health Organization. It comprises 23 items, with eight items assessing physical health and 15 items assessing psychosocial health. The Parent Proxy Report form for toddlers is used for subjects 2 to 4 years of age; the Parent Proxy Report is used for subjects 5 to 7 years of age; the Child Self Report is used for subjects from 8 to 12 years of age; and the Teen Self Report is used for subjects 13 years of age and older.

Four subscales, including physical functioning (eight items), emotional functioning (five items), social functioning (five items), and school functioning (three items for toddlers ages 2-4, and five items for the others) are collected. Three summary scores will be presented: total scale score (all subscales), physical health summary score (physical functioning scale only), and psychosocial health summary (emotional, social, and school functioning scales combined).

To obtain summary scores, item scores (from 0 to 4) are reverse transformed to a 0-100 scale (0=100, 1=75, 2=50, 3=25, and 4=0). To create summary scores, the mean is computed by totaling the item scores and dividing by the number of items answered. If >50% of the items in the scale are not completed, the summary score will not be computed. Specifically, to create the psychosocial health summary score, the mean is computed as the sum of the items over the number of items answered in the emotional, social, and school functioning subscales. The physical health summary score is the same as the physical functioning subscale score. To create the total scale score, the mean is computed as the sum of all of the items over the number of items answered on all of the scales.

2.1.4 Safety endpoints

The safety analysis for this double-blind study will be based on the reported adverse events, serious adverse events, and retinoid-associated adverse events. Other safety information, such as the Columbia-Suicide Severity Rating Scale (C-SSRS), electrocardiograms (ECGs), vital signs (temperature, respiratory rate, blood pressure, and heart rate), physical examination, body weight (and weight-for-age z-score) and height (and height-for-age z-score) and body mass index (BMI) (and BMI-for-age z-score), laboratory parameters (hematology, biochemistry, and urinalysis), blood or urine pregnancy tests for females of child-bearing potential, Tanner staging, and concomitant medication reporting will be summarized by visit, consistent with the methods described for the baseline summaries. Evaluation of subjects under the age of 18 years enrolled with open epiphyses will include growth assessment with linear height and knee height measurements every 6 months, hand/wrist and knee plain radiographs evaluating growth plate and bone age every 6 months, and bilateral long bone lengths (ulna, radius, femur, tibia, and fibula) by radiograph every 12 months. Bone mineral content (BMC) and density at lumbar spine, hip, and mid radius will be assessed by DXA every 6 months. MRI scans obtained every 12 months for the assessments of OCs will also be evaluated for potential osteonecrosis of the hip, shoulders, and knees. Hearing and visual acuity will be assessed at baseline and every 12 months. All assessments are performed subject to the double-blind; any exceptions, e.g. if a serious adverse event requires unblinding the investigator, will be described in the clinical study report.

Observation period

The observation period will be divided into three periods: screening, treatment, and post-treatment. The treatment-emergent adverse event (TEAE) period includes both the treatment and post-treatment period:

- **Screening period** is defined as the date from the signed informed consent up to the date prior to first dose of study drug (Day 1) for Adverse events and concomitant medications, while it is defined as the date from the signed informed consent up to pre-dose time of the first study drug (Day 1) for other safety and efficacy measurements.
- **Treatment period** is defined as the date/time after the first dose of study drug (Day 1) to 7 days after the last dose of study drug for subjects who terminated study drug before Month 24 (end-of-treatment visit [EOT]+7 days), or to 7 days after scheduled EOT (nominal Month 24 visit+7 days) for subjects who complete the Month 24 study drug administration.
- **Post-treatment period** is defined as the date immediately following the end of the treatment period (as defined above) to completion or discontinuation of the MO-Ped Trial/end-of study (EOS) (EOT + 6 months). Because of the discontinuation of study treatment due to clinical hold, the post-treatment period will commence on 13 December 2019, which was calculated as last dose plus 7 days (Clinical hold notification to sites was on Dec 5, 2019; patients were notified on Dec 6, 2019).

2.1.4.1. Adverse event variables

Adverse event observation period

- Pretreatment adverse events are adverse events that developed, worsened, or became serious during the screening period.
- On-treatment TEAEs are adverse events that developed, worsened, or became serious during the treatment period.
- Post-treatment TEAEs are adverse events that developed, worsened, or became serious during the post-treatment period.

All adverse events (including serious adverse events) will be coded to SOC and PT from the most current version during the study of MedDRA. Adverse event severity will be assessed and reported according to criteria defined in the protocol (mild, moderate, severe). Adverse event causality will also be assessed in terms of relationship to study drug and reported according to criteria described in the protocol (not related, possibly related, probably related, definitely related).

The occurrence of adverse events (including serious adverse events) will be recorded from the start of the screening period until the end of the post-treatment period.

Post-baseline abnormal findings on laboratory assessments, physical examinations, vital signs or ECGs assessed as clinically significant by the Investigator will be recorded as adverse events.

Retinoid-associated adverse events

Adverse events known to be associated with retinoids (e.g., mucocutaneous events) will be further graded according to Common Terminology Criteria for Adverse Events (CTCAE), Version 4.03, 14 June 2010 to keep consistency with previous studies (Table 1). Infectious pharyngitis and conjunctivitis are coded to the Infection and Infestation SOC and do not require CTCAE grading.

Table 1. Retinoid-associated adverse events by CTCAE severity criteria

Adverse Event	**CTCAE Page Number (Version 4.03, 14 June 2010)**
Corneal ulcer	22
Conjunctivitis	22
Dry eye	23
Keratitis	24
Night blindness	24
Cheilitis	30
Dry mouth	33
Mucositis oral	45
Pancreatitis	48
Pharyngitis	81
Alanine aminotransferase increased	107
Aspartate aminotransferase increased	107
Blood bilirubin increased	107
Lipase increased	111
Serum amylase increased	112
Hypertriglyceridemia	116
Alopecia	179
Dry skin	179
Erythroderma	180
Photosensitivity	183
Pruritus	184
Rash maculo-papular	185
Skin and subcutaneous tissue disorders – other, specify	187

CTCAE, Common Terminology Criteria for Adverse Events

2.1.4.2. Deaths

The deaths observation periods align with the periods defined above.

- Deaths on-treatment: deaths occurring during the treatment period
- Death post-treatment: deaths occurring during the post-treatment period

2.1.4.3. Laboratory safety variables

Clinical laboratory data include hematology, biochemistry (including lipids), serum or urine pregnancy tests (when applicable), and urinalysis. Clinical laboratory values will be converted to conventional units; conventional units will be used in all listings and tables.

Hematology, biochemistry, and urinalysis data will be collected at Screening, every 3 months (Months 3, 6, 9, 12, 15, 18, 21, and 24), and during the 6 months post-treatment safety follow-up if laboratory abnormalities were observed during the treatment period. If no laboratory abnormalities were observed in the first 12 months of treatment, the frequency of hematology, lipids and biochemistry assessments may be changed to every 6 months.

The laboratory parameters will be classified as listed in Table 2.

Table 2. Clinical laboratory parameters

Biochemistry:	
Sodium	Globulin
Potassium	Alkaline phosphatase (ALP)
Chloride	Aspartate transaminase (AST)
Bicarbonate	Alanine transaminase (ALT)
Blood urea nitrogen	Gamma glutamyl transferase (GGT)
Creatinine	Uric acid
Calcium	Total thyroxine (T4)
Inorganic phosphorous	Free T4
Glucose	Thyroid-stimulating hormone
Total bilirubin	Amylase
Total proteins	Lipase
Albumin	Parathyroid hormone (at baseline)
Lipid Profile[1]:	
Triglycerides	High-density lipoprotein (HDL)
Total cholesterol	Low-density lipoprotein (LDL)
	Very low-density lipoprotein (VLDL)
Hematology:	
Hemoglobin	Platelets
Hematocrit	White blood cell count (including differentials)
Red blood cell count	Neutrophils
Packed cell volume	Lymphocytes
Mean corpuscular volume	Monocytes
Mean corpuscular hemoglobin	Eosinophils
Mean corpuscular hemoglobin concentration	Basophils
Urinalysis[2]:	
pH	Blood (free hemoglobin)
Protein	Nitrite
Glucose	Urobilinogen
Ketones	Specific gravity
Bilirubin	Color & appearance

[1] Non-fasting lipid profile will be obtained. If the triglyceride is abnormal, a fasting triglyceride should be repeated.
[2] If results are abnormal, then a microscopic evaluation should be completed.

2.1.4.4. Electrocardiogram variables

A 12-lead ECG will be performed at screening, baseline, and at Months 12 and 24 . Heart rate, RR, PR, QRS, QT, QTcF, and QTcB intervals will be determined using centralized readings.

The following define the categories for ECG abnormalities.

- Rhythm:
    - Atrial premature complexes
    - Ventricular premature complexes
    - Sinus arrhythmia
    - Sinus tachycardia

- Conduction:
    - Nonspecific intraventricular conduction delay
    - Right bundle branch block
    - Incomplete right bundle branch block
    - Left posterior fascicular block
    - Left bundle branch block
    - Incomplete left bundle branch block
- Morphology/chamber enlargement:
    - Left ventricular hypertrophy
    - Left ventricular hypertrophy with repolarization abnormality
- Axis deviation:
    - Right axis deviation
    - Left axis deviation
- Myocardial infarction (MI):
    - Pathologic Q waves
    - Acute ST elevation MI
    - Acute non-ST elevation MI
- ST segment/T waves/U waves:
    - Early repolarization
    - Nonspecific ST and T wave abnormality
    - Nonspecific ST elevation
    - Nonspecific U wave abnormalities

2.1.4.5. Vital Signs

Vital sign variables including heart rate, systolic blood pressure, diastolic blood pressure, respiratory rate, and body temperature will be assessed at screening, baseline, every 3 months during site and remote visits during treatment, and at safety follow-up, if appropriate.

2.1.4.6. Physical Exams

A comprehensive physical examination including body weight and BMI is performed by a physician at screening, baseline, every 6 months (Months 6, 12, 18, and 24), and at the 6-month safety follow-up to monitor whether objective changes occur. Body weight and BMI will be calculated and summarized relative to age- and gender-matched peers by calculating z-scores.

Note that, as stated in the protocol, any post-baseline abnormal findings assessed as clinically significant are recorded as adverse events.

2.1.4.7. Columbia-Suicide Severity Rating Scale

In accordance with the FDA guidance for industry on suicidal ideation and behavior[2], all subjects 8 years of age and older are assessed for suicidal ideation and behavior using the C-SSRS at Screening and every 3 months (Months 3, 6, 9, 12, 15, 18, 21, and 24). The adult form is used for subjects 12 years and older and the pediatric form is used for subjects 8 to 11 years old.

2.1.4.8. Knee and hand/wrist radiographs for assessment of epiphyseal growth plate

All subjects undergo knee and hand/wrist radiographs (anterior/posterior view) at screening, every 6 months (Months 6, 12, 18, and 24) and at the 6-month safety follow-up visit. Once a subject has achieved 100% skeletal maturity (ie, complete fusion of growth plates) as determined by the knee and hand/wrist radiographs, further radiographs are no longer necessary. If 100% skeletal maturity is not achieved at both anatomical locations, then only the location that is still maturing needs to be monitored.

Two treatment-blinded and independent readers from a central imaging laboratory interpret all radiographs to ensure consistent assessment of the radiographs. The Investigator is provided with all results and reviews and assesses abnormal results for clinical significance. Any post-baseline abnormal result assessed as clinically significant is recorded as an adverse event. For each abnormality, independent reviewers assess changes from baseline and changes from the prior assessment time point.

The following parameters are assessed:

- Hand/wrist and knee radiograph images
    - Epiphyseal growth plate abnormalities assessment:
        - Frayed metaphyseal edge
        - Cupping
        - Calcinosis (punctate or streaky calcifications, not orderly closure of the physes)
        - Widening of the epiphyses
        - Sclerosis of the adjacent growing bone (metaphyseal edge)
        - Under-mineralization or osteopenia of the adjacent growing bone
        - Dense metaphyseal lines (also known as growth recovery lines)
        - Periosteal elevation
        - Other, specify (comments required)

- Epiphyseal growth plate closure assessment with premature closure of the epiphyses as defined by having partially closed, or closed epiphyses at a chronological age that is less than 2 SD below mean age of expected epiphyseal closure
- Bone age (years, months), if applicable (hand/wrist only)

  The average of bone age assessments from two independent readers will be averaged and summarized, as described in the imaging read charter.

Discrepancies in the assessment of growth plate abnormalities or status of closure of the epiphysis in the forearms and knee are subject to adjudication based on a consensus review by two independent reviewers, as described in the imaging read charter. The adjudicated assessment will be summarized.

2.1.4.9. Bone Densitometry

Bone mineral content, areal bone mineral density (aBMD), and bone area at the lumbar spine, hip, and mid-third radius will be assessed with DXA at baseline and every 6 months (Months 6, 12, 18, 24, and the 6-month safety follow-up visit if appropriate). Analyses will be conducted to evaluate BMC, aBMD, and bone area (percentage changes from baseline) in the lumbar spine, hip, and mid-third radius. Spine and hip BMC and aBMD z-scores (changes from baseline) will be calculated by the imaging vendor (per availability of reference data) and adjusted for height for lumbar spine aBMD.

Vertebral compression fractures (VCF) are potential clinical sequelae of low BMD. To avoid additional radiation exposure and procedures, existing DXA and MRI will be assessed by the central imaging reader for the presence and absence of VCF. Using the existing post baseline MRI sequences centered on the thoracic and pelvic girdles, VCFs will be assessed as present or absent from the fourth thoracic to the fifth lumbar vertebrae on an annual basis. These assessments are supplemented by spine DXA for severe VCF (>50% height loss, or > grade 3). If VCFs are detected in the postbaseline DXA or MRI scans. Baseline MRI will be assessed for background prevalence of VCF in this population.

2.1.4.10. Linear Growth Assessment

Linear growth is assessed in all subjects by a stadiometer and by measuring the knee-height in both knees at baseline and every 6 months (Months 6, 12, 18, 24 and the 6-month safety follow-up visit, if appropriate). Linear heights are performed in triplicate and averaged to generate a height-for-age z-score.

2.1.4.11. Osteonecrosis

The presence of potential osteonecrosis at the hips, shoulders, and knees is assessed by whole body MRI at baseline and at Months 12 and 24, as it detects early osteonecrosis as low signal regions with T1 sequences and high signal regions with T2/STIR sequences. Other abnormalities that may

be seen with osteonecrosis such as cysts, subchondral lucency, subchondral collapse, and irregular or narrowing of the joint space are also assessed. All images will be interpreted by a central imaging laboratory, as described in the imaging read charter.

Each body region will be assessed with documentation of one of the following assessments:

- No (osteonecrosis not present)
- Not evaluable (image quality not sufficient)
- Yes (possible, probable, definite osteonecrosis present)

If osteonecrosis is assessed as possibly/probably/definitely present, the independent reviewers will assess change across time points as follows:

- Change from baseline and change from the prior assessment time point
    - Not present at baseline or prior timepoint
    - No change from baseline or prior timepoint
    - Improved from baseline or prior timepoint
    - Worse than baseline or prior timepoint
    - Unknown: baseline was not evaluable/missing (NE/missing)

2.1.4.12. Long bone Length

Bilateral long bone lengths (radius, ulna, femur, tibia, and fibula) are assessed with forearm and lower extremity anterior/posterior radiographs at baseline and at Months 12 and 24. Lengths of paired bones (radius and ulna; tibia and fibula) are assessed and ratio of paired bone lengths will be described using summary statistics. If surgery was undertaken for guided growth involving one of the paired bones, these paired bones will be excluded from subsequent analyses.

2.1.4.13. Tanner Staging

To assess growth and development, Tanner staging is performed at baseline and at Months 12 and 24. Tanner staging is not necessary once Stage V is reached.

2.1.4.14. Hearing and Visual acuity

Hearing and visual acuity will be assessed at baseline and at Month 12 and 24. Clinically significant abnormalities will be reported as adverse events.

2.1.5 Palatability of sprinkled drug product

Palatability is assessed with a 5-point hedonic face scale following the first dose at Study Day 1 and at Month 1 in subjects ≥4 years of age who sprinkle the drug product or placebo onto a spoonful of specific foods.

2.1.6 Pharmacokinetic endpoints

The steady-state pharmacokinetics of palovarotene will be assessed at Month 1. Pharmacokinetic samples may be obtained at the site or obtained by a visiting nurse during the remote visit at the discretion of the subject and investigator. Sparse pharmacokinetic blood samples are collected before dosing and at 3, 6, 10, and 24 hours post-dose. Pharmacokinetic sampling may be repeated at a subsequent visit if the first attempt was unsuccessful.

The determination of palovarotene plasma concentrations is performed using a validated liquid chromatography–tandem mass spectrometry method. The time of sample collection as it relates to the time of dosing on the pharmacokinetic sample collection days is recorded.

2.2 Disposition of subjects

This section describes the disposition for subject status and the patient populations.

Screened subjects are defined as all subjects who sign the informed consent.

Randomized subjects include all screened subjects who are allocated to a randomized treatment group.

The total number of subjects for each of the following categories will be presented in a table:

- Screened subjects
- Screen failure subjects and the reasons for screen failure
- Randomized subjects
- Randomized and treated subjects
- Subjects who permanently discontinued study drug and reasons for discontinuations.
- Subjects who completed the MO-Ped Trial
- Subjects who the discontinued the MO-Ped Trial and reasons for discontinuations

For all categories of subjects (except for the screened category), percentages will be calculated using the number of randomized and treated subjects as the denominator. Reasons for study drug discontinuation will be presented in tables using the total number of subjects in the safety population as denominator.

All major protocol deviations potentially affecting the efficacy analyses and other major deviations will be summarized in tables giving numbers and percentages of deviations by treatment grouping.

Additionally, the analysis populations will be summarized in a table by subject counts based on the safety populations (see definitions in Section 2.3). A listing for subject disposition will be also provided.

2.3 Analysis populations

Subjects will be grouped into the following analysis sets:

- The Randomized Set will include any subject allocated to a randomized treatment group, regardless of whether the study drug was administered.

- The Full Analysis Set (FAS) will include randomized subjects who received at least one dose of study drug. Analyses using the FAS will be according to the treatment group allocated by randomization and not the actual treatment received. All efficacy analyses will be based on the FAS unless specified otherwise.

- The Safety Set (Safety) will be a subset of the Randomized Set and include randomized subjects who receive at least one dose of study drug. The Safety Set will be analyzed according to the treatment received, unless otherwise specified. All safety analyses will use the Safety population unless specified otherwise.

- The Pharmacokinetic Set (PKS) will be a subset of the Safety population and include subjects with evaluable pharmacokinetics data. All Pharmacokinetic analyses will use the PKS population unless specified otherwise.

2.4 Statistical methods

2.4.1 Demographic and baseline characteristics

Continuous parameters will be summarized using the number of subjects with available data, mean, SD, median, and minimum and maximum for each treatment group. Categorical and ordinal parameters will be summarized using number and percentage of subjects in each treatment group.

Parameters will be summarized for the FAS and Safety populations.

Parameters described in Demographic and baseline characteristics will be summarized by treatment grouping specified earlier using descriptive statistics.

Medical and surgical history will be summarized by SOC and PT sorted by decreasing frequency of SOC and by decreasing frequency of PT within SOC in the Safety populations.

2.4.2 Prior or concomitant medication

The prior and concomitant medications will be summarized based on the Safety Set as follows:

- Prior medications
- Concomitant medications

The medications will be summarized according to the WHO-DD by ATC class 3 (or class 2 if class 3 was not coded) and the preferred drug names.

The tables for prior and concomitant medications will be sorted by decreasing frequency of ATC therapeutic subgroup, followed by the preferred drug names based on the overall incidence in the Safety populations.

2.4.3 Study drug exposure

The extent of study drug exposure will be assessed and summarized based on the Safety population.

The extent of study drug exposure is defined as [*last study drug date – first study drug date + 1*], regardless of unplanned study drug interruptions.

The duration of dosing, the number of subjects with at least one dose reduction or interruption and the number of subjects discontinuing study drug will be summarized.

Treatment compliance is defined as the total cumulative doses received divided by the total planned dose at or before study termination × 100%.

The duration of dosing, treatment compliance, the number of subjects with at least one dose modification, the number of dose modifications, and the number of subjects with interrupted study drug will be summarized.

2.4.4 Analyses of efficacy endpoints

Efficacy analyses will be performed on FAS with all subjects with post-baseline visits on or prior to 04 December 2019; and on all subjects with post-baseline efficacy assessments, regardless of timing.

For all efficacy endpoints, the point estimates for each treatment group, together with 95% confidence interval will be provided. The statistical comparisons on the two treatment groups vs. placebo will be performed on the primary and key secondary endpoints. Details are provided in the related sections. For the proportions, the exact confidence interval using Clopper-Pearson method will be presented. Where appropriate, p-values will also be presented, with no adjustment for multiple-testing.

Due to early termination of the study, some efficacy endpoints at some scheduled visits, like 6 months safety follow up, are not performed. The summary tables and analyses will use the actual study visits that are available in the analysis dataset.

2.4.4.1. Analyses of primary efficacy endpoint

Primary analyses

The primary efficacy endpoint is the annualized rate of new OCs as assessed by whole body MRI. The annualized rate of new OCs will be estimated using a negative binomial regression model on FAS. The last assessment in the study will be used in the analysis. All new OCs observed at post-baseline will be carried forward to the last assessment. Covariates included in the regression model

are three treatment groups, baseline age, sex, and Ext1/2 mutation status. As siblings may be enrolled, repeated measures with unstructured covariance structure will be used to address the potential family-level correlation. An offset variable equal to the log of the number of years of follow-up will be included to convert the estimate to annualized rate. The treatment effect on two dose levels vs. placebo will be estimated for exploratory purpose since the power to detect the difference would be small due to reduced sample size. Non-confirmatory exploration of the annualized rate of new OCs for each treatment group, together with the ratios of annualized rate of new OCs with placebo and the p-values of the ratio estimates, will be obtained through the point estimate and its confidence interval using the following statistical model. No multiple testing adjustment will be implemented on the p-values. If there are no subjects from the same family, the "repeated" statement in the model will be removed.

```
proc genmod descending;
    class TRTA AGE SEX MUTATION FAMILY;
    model New_OCs = TRTA AGE SEX MUTATION /dist=negbin link=log offset=L_YR;
    lsmeans TRTA/CL DIFF PDIFF EXP;
    repeated subject= family/ type=un;
run;
```

For anatomical regions where counts of OCs cannot be ascertained (e.g., due to motion artifacts), last observation carried forward (or first observation carried backward if the missing region is at baseline) will be used to complete the count of OCs. Assessments that cannot be recovered will not be imputed and missing at random (MAR) will be assumed automatically in the SAS mixed model repeated measurement procedure (SAS PROC GENMOD procedure).

Descriptive statistics summarizing total new OCs, and annualized new OCS, by study visit will be provided.

Subgroup analyses

The following subgroup analyses will also be performed to examine the consistency of the estimated treatment effect by age, gender and exostosin gene mutation. The analysis method will be the same as the primary analysis, except that treatment group will be the only independent variable included in the model, as below using age as example:

```
proc genmod descending;
    by age;
    class TRTA;
    model New_OCs = TRTA/dist=negbin link=log offset=L_YR;
    lsmeans TRTA/CL DIFF PDIFF EXP;
    repeated subject= family/ type=un;
run;
```

The variables for subgroup analyses are as follows:

- Baseline age (≤7 years, >7 years)
- Sex
- Ext1/2 mutation status

2.4.4.2. Analyses of secondary efficacy endpoints

Change from baseline in total volume of OCs (key secondary endpoint)

Descriptive statistics summarizing the total OCs volume and change from baseline by study visit will be provided by treatment groups. The treatment effect for two treatment groups vs. placebo will be estimated for exploratory purposes, since the power to detect the difference would be small due to reduced sample size. No multiple testing adjustment will be implemented on the p-values. Non-confirmatory exploration on the change from baseline for the total volume of OCs between each treatment group and placebo will be obtained through the following statistical model.

```
proc genmod descending;
      class TRTA AGE SEX MUTATION FAMILY;
      model Change = TRTA BASE_VOLUME AGE SEX MUTATION /dist=normal;
      lsmeans TRTA/CL DIFF PDIFF;
      repeated subject= family/ type=un;
run;
```

Proportion of subjects with no new OCs

The proportion of subjects with no new OCs as assessed by whole body MRI will be reported by study visit by treatment group, together with exact confidence interval, using the Clopper-Pearson method.

Subgroup analyses will also be provided by Age, Sex, and Ext1/2 mutation status.

Annualized rate of new or worsening deformities

The annualized rate of new or worsening deformities (as assessed by radiographic imaging of both upper and lower limbs) will be analyzed as described for the primary endpoint analysis. Deformities that were corrected will be assessed as the pre-surgical state. Descriptive statistics summarizing new or worsening deformities by study visit will be provided.

Subgroup analyses will also be provided by Age, Sex, and Ext1/2 mutation status.

Annualized rate of MO-related surgeries

The annualized rate of MO-related surgeries will be analyzed as described for the primary endpoint analysis. Planned surgeries within 6 months of enrollment to remove symptomatic OCs present at baseline, or to correct deformities present at baseline, will be recorded but not contribute to the endpoint. Similarly, surgical procedures that are a continuation of a previous procedure, such as removal of a surgical plate for guided growth, will not be counted as a separate MO-related surgery endpoint event. Descriptive statistics summarizing MO-related surgeries by study visit will be provided.

Subgroup analyses will also be provided by Age, Sex, and Ext1/2 mutation status.

2.4.4.3. Analysis of exploratory efficacy endpoints

The change from baseline in the total volume of cartilage caps of OCs as assessed by whole body MRI at Months 12 and 24 will be analyzed using the same statistical methods employed for the analysis of the secondary efficacy endpoint change from baseline in the total volume of OCs.

The annualized rate of new or worsening functional limitations, defined as restrictions in joint range of motion, will be analyzed as described from the primary endpoint analysis. Functional limitations that were surgically corrected will be assessed as the pre-surgical state.

The change from baseline in PedsQL summary scores (total, physical health, and psychosocial health), FPS-R, and PROMIS Pain Interference T-score will be analyzed using summary statistics by study visit.

Subgroup analyses will also be provided by Age, Sex, and Ext1/2 mutation status.

2.4.5 Analysis of safety endpoints

General rules

All safety analyses will be performed on the Safety population defined in Section 2.3, using the following rules:

- Baseline is defined as the last available value prior to the first dose of study drug or on the same day as the first dose of study drug if the time of evaluation is not available for finding domains. If no value prior to first dose of study drug is available, value obtained within 48 hours of the first dose may be used unless specified separately in the protocol.
- The potentially clinically significant abnormality (PCSA) values are defined as abnormal values according to predefined criteria/thresholds (see Appendix, Section 6.2).
- For adverse events, the following conventions for partial or completely missing date imputations will be implemented:
    - If the start date is completely missing, the start date is imputed as the date of first dose and the adverse event is considered treatment emergent.
    - If the day and month of the start date are missing and the start year is the same as the year of first dose, the start date is imputed as the date of first dose and the adverse event is considered treatment emergent. If the start year is prior to the year of first dose, the start date is imputed as January 1 of the year and the adverse event is not considered treatment emergent.
    - If the day of the start date is missing and the start month and year is on or after the date of first dose, the start date is imputed as the date of first dose or the first day of the month of the adverse event, whichever is later, and the adverse event is considered treatment emergent. Otherwise, if the start month and year is before the month and year of first dose, then the start date is imputed as the first day of the month of the adverse event, and the adverse event is not treatment emergent.

  - If the start date imputed based on above is after the stop date, such as the stop date is before treatment onset but start date is complete missing, set the stop date as start date.

- For concomitant medications, the following conventions for partial or completely missing date imputations will be implemented:
  - If the start date and end date are completely missing, then the start date is imputed as the date of informed consent, the end date is the date of study completion or early discontinuation, and the medication would be considered both prior and concomitant.
  - Partial start dates are imputed as the first day of the month (if only missing the day) or January 1 (if missing day and month), and partial end dates are imputed as the last day of the month (if missing the day only), or December 31 (if missing the day and month).

2.4.5.1. Analyses of adverse events

The primary focus of adverse event summaries will be on TEAEs in the treatment period. Pre-treatment and post-treatment adverse events will be described separately.

For summary of adverse events, palovarotene 2.5 mg and 5.0 mg will be summarized individually as well as pooled.

Adverse event incidence tables by SOC and PT will be sorted by decreasing incidence of SOC and by decreasing incidence of PT within SOC for the pooled palovarotene group. The number (n) and percentage (%) of subjects experiencing an adverse event will be presented. Multiple occurrences of the same event in the same subject will be counted only once. The denominator for computation of percentages is the Safety Set.

The following TEAE summaries will be generated for the Safety Set.

Analysis of on-treatment treatment-emergent adverse events

- Overview of TEAEs, summarizing number and percent of subjects with
  - TEAEs
  - TEAEs by maximum severity
  - TEAEs possibly related to treatment
  - Treatment-emergent retinoid-associated AEs with maximum CTCAE rating
  - Treatment-emergent serious adverse events
  - Treatment-emergent serious adverse events possibly related to treatment
  - TEAEs leading to dose modification or interruption of study drug
  - TEAEs leading to permanent study drug discontinuation

- TEAEs leading to death
- TEAEs in the epilepsy SMQ by PT
- TEAEs in the SMQ for Torsade de pointes/QT prolongation by PT

- All TEAEs by primary SOC and PT, showing number and percent of subjects with at least one TEAE.
- All TEAEs with incidence ≥5% in the overall group and by PT.
- All TEAEs by maximum severity, presented by primary SOC and PT, showing number (%) of subjects with at least one TEAE by severity (i.e., mild, moderate or severe).
- All TEAEs by relationship, presented by primary SOC and PT, showing number (%) of subjects with at least one TEAE by relationship (i.e., not related, at least possibly related).

Analysis of on-treatment retinoid-associated TEAEs, treatment-emergent serious adverse events, TEAEs leading to dose modification or interruption of study drug, TEAEs leading to permanent study drug discontinuation, TEAEs leading to study discontinuation

- All TEAEs by primary SOC and PT, showing number and percent of subjects with at least one TEAE.
- All TEAEs by maximum severity, presented by primary SOC and PT, showing number and percent of subjects with at least one TEAE by severity (ie, mild, moderate, or severe).
- All TEAEs by relationship, presented by primary SOC and PT, showing number and percent of subjects with at least one TEAE by relationship (ie, not related, at least possibly related).
- All treatment-emergent retinoid-associated adverse events by maximum CTCAE grade, presented by primary SOC and PT, showing number and percent of subjects with at least one TEAE by CTCAE grade.

Analyses of pre-treatment adverse events

- All pre-treatment adverse events by primary SOC and PT, showing number and percent of subjects with at least one pre-treatment adverse event.

Analyses of post-treatment adverse events

- All post-treatment TEAEs by primary SOC and PT, showing number and percent of subjects with at least one post-treatment adverse event.

Listings will be provided for all pre-treatment adverse events; all on-treatment TEAEs; post-treatment TEAEs; all serious adverse events; all on-treatment TEAEs leading to dose modification or interruption of study drug; all on-treatment TEAEs leading to permanent study drug discontinuation; all on-treatment TEAEs leading to study discontinuation; and all on-treatment retinoid-associated TEAEs.

2.4.5.2. Deaths

Listings will be provided for all deaths that occur during the study.

2.4.5.3. Analyses of laboratory variables

The summary statistics of all laboratory variables (central laboratory values and changes from baseline) will be calculated for each visit. In addition, for all laboratory parameters, the mean and mean change from baseline at each scheduled visit during the treatment period will be plotted.

The incidence of PCSAs at any time (PCSA list provided in the Appendix [Section 6.2]) during the TEAE period will be summarized by biological function (laboratory, vital signs and ECGs) . Tabular summary will present any subject with a value meeting PCS criterion for each biologic function by treatment assignment. A focus will be on new-onset PCS values (ie, subjects with preexisting PCS values at baseline will not be considered to have new onset values after start of study drug). All PCS values will be presented in a separate listing as well.

Tables of shifts from baseline to worst post-baseline on study will be provided for key laboratory parameters, including alanine aminotransferase (ALT), aspartate aminotransferase (AST), bilirubin, amylase, lipase and triglycerides. Abnormal liver function test results at any time in the study will be presented using ALT, AST, and bilirubin plasma levels as follows:

1) ALT >3x upper limit of normal (ULN) or AST >3x ULN and total bilirubin >2x ULN;

2) ALT or AST >3x ULN and total bilirubin >5x ULN.

3) ALT or AST >8x ULN and total bilirubin >8x ULN.

2.4.5.4. Analyses of ECG variables

The summary statistics of all ECG variables (ECG values and changes from baseline) will be calculated for each visit. Mean heart rate, RR PR, QRS, QT, QTcF, and QTcB will be determined using centralized readings.

The incidence of ECG abnormalities at any time (abnormality list provided in Section 2.1.4.4) during the TEAE period will be summarized. New-onset ECG PCS values will be summarized by treatment assignment. Subjects with pre-existing ECG [PCS] values at baseline will not be considered to have new onset values after start of study drug. (PCSA list is provided in the Appendix, Section 6.2.)

All PCS values will be presented in a separate listing as well.

2.4.5.5. Analyses of vital sign variables

The summary statistics of all vital sign variables (vital signs values and changes from baseline) will be calculated for each visit.

The incidence of PCSAs at any time (PCSA list provided in Appendix, Section 6.2) during the TEAE period will be summarized. A focus will be on new-onset PCS values (i.e., subjects with pre-existing PCS values at baseline will not be considered to have new onset values after start of study drug).

All PCS values will be presented in a separate listing as well.

2.4.5.6. Analyses of Columbia-Suicide Severity Rating Scale variables

Subjects assessed with any type of suicidal ideations according to the C-SSRS or exhibiting any suicidal behavior during the study will be presented in listings. Tabular summary of subjects with any type of suicidal ideation or behavior will be presented. A listing will summarize any subject with Type 4 or 5 suicidal ideation or any suicidal behavior.

2.4.5.7. Analyses of knee and hand/wrist radiographs for growth plate assessment

The summary statistics of bone age (averaged across the two independent readers, as described in the Imaging Review Charter) and the difference between bone age and chronological age (values and changes from baseline) will be calculated for each visit. Number and listing of subjects with bone age advancement, defined as change in bone age greater than change in chronological age + 1 SD, will be provided.

Summary tables of growth plate abnormalities identified in radiographs will include the following:

- Number and proportion of subjects with any abnormality at baseline, and by specific abnormality (eg, growth recovery lines, sclerosis).
- Number and proportion of subjects with any new post-baseline or prior time point abnormality, and by specific abnormality.
- Number and proportion of subjects with improvement from baseline or prior time point abnormality, and by specific abnormality.

Listings of all bone age assessments (from the two independent readers and the average) will be provided. Listings of all subjects with epiphyseal growth plate abnormalities such as frayed metaphyseal edges, cupping, calcinosis, or widening of the epiphyses, sclerosis of the adjacent growing bone, under-mineralization or osteopenia of the adjacent growing bone and periosteal elevation will be provided. Listings of all subjects with new, not present at baseline or worsening post-baseline epiphyseal growth plate abnormalities will be provided. Listings of all subjects with premature growth plate closure will be provided.

2.4.5.8. Analyses of bone densitometry

Bone mineral content; BMD; percentage change from baseline BMD at the lumbar spine, hip, and mid-third radius; and BMC and aBMD z-scores at the lumbar spine and hip (values and changes from baseline) will be calculated for each visit (by the central imaging vendor). For the lumbar spine, height-adjusted aBMD z-score will be analyzed.

Figures will include changes in BMD; the percent change from baseline in BMD and BMC at the lumbar spine, hip, and mid-third radius; and BMC and aBMD z-scores at the lumbar spine and hip. For the lumbar spine, height-adjusted aBMD z-score will be analyzed. Tabular summaries will be provided for the number and proportion of subjects with study drug dose modifications and for those who discontinue due to aBMD changes, and subjects with these modifications will be

flagged in the study drug administration/accountability listing. Listings of all subjects who discontinued study treatment due to confirmed aBMD loss in the spine will be provided.

Listings of all subjects who had post baseline VCF as determined by DXA or MRI will be provided.

2.4.5.9. Analyses of osteonecrosis

Summary tables of osteonecrosis will include the following:

- Number and proportion of subjects with possible, probable, or definite osteonecrosis at baseline.
- Number and proportion of subjects with any new post-baseline osteonecrosis.
- Number and proportion of subjects with any new or worsening, or improving, post-baseline osteonecrosis.

Listings of all subjects with possible, probable, or definite osteonecrosis will be provided with new or worsening post-baseline abnormalities subjects flagged.

2.4.5.10. Analyses of linear growth

The summary statistics of linear height, linear height z-scores, and bilateral knee heights (values and changes from baseline) will be calculated for each visit. The Annualized linear height change, calculated as (Height at visit – Height at baseline)/Duration in days from baseline/365.25), will also be summarized. Figures will summarize changes in linear height z-scores.

2.4.5.11. Analyses of long bone lengths

Summary statistics of absolute and changes from baseline for bilateral long bone lengths (radius, ulna, femur, tibia, and fibula) will be calculated for each visit. Summary statistics of the ratios of the paired bones (radius and ulna; tibia and fibula) and change from baseline will calculated for each visit.

2.4.5.12. Analyses of Tanner staging

The summary statistics of Tanner staging (absolute values and changes from baseline) will be calculated for each visit. A shift table will also be produced.

2.4.6 Analysis of palatability endpoint

The 5-point hedonic face scale will be summarized using descriptive statistics, by treatment group and age group.

2.4.7 Analyses of pharmacokinetic endpoints

Pharmacokinetic analyses will be conducted using the PKS.

The following parameters will be determined where possible by model independent analysis using WinNonlin™: maximum and trough plasma drug concentrations at steady state ($C_{max,ss}$, $C_{min,ss}$, respectively), time to reach maximum steady-state plasma concentration ($T_{max,ss}$), area under the plasma concentration-time curve at steady-state from time zero to 24 hours (AUC_{0-24ss}), elimination rate constant (λ_z), elimination half-life ($t_{½z}$), and apparent total clearance of study drug from plasma after oral administration (CL/F).

Plasma concentrations and AUC will be summarized using descriptive statistics. Plasma concentrations below the limit of quantification will be set to 0.

The following steps describe how observed concentrations will be used to compute AUC if some plasma concentrations are unavailable:

- If plasma concentrations are missing at both 0 and 24 hours, then the AUC cannot be calculated.
- If the 0-hour concentration is missing, the 24-hour concentration will be used to impute the missing 0-hour concentration.
- If the 24-hour concentration is missing, the 0-hour concentration will be used to impute the missing 24-hour concentration.
- Invalid 24-hour concentrations will be replaced with the 0-hour concentrations. Twenty-four-hour concentrations will be determined to be invalid if the concentration is greater than the 10-hour concentration as it will be assumed to have been obtained after the next day dose instead of before the next day dose.
- The following concentrations must be available to calculate the AUC:
    - 0 and/or 24-hour concentration(s)
    - 10-hour concentration

3 INTERIM ANALYSIS

No formal interim analysis (to declare early efficacy or futility) will be performed, due to the early termination of the study. A modified interim analysis – to support the interim CSR to be used in the palovarotene FOP NDA – will be conducted.

4 SOFTWARE DOCUMENTATION

All summaries and statistical analyses will be generated using SAS Version 9.4 or higher, unless otherwise noted.

5 REFERENCES

1. Centers for Disease Control and Prevention. Normal Joint Range of Motion Study. Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/jointrom/index.html. Published November 29, 2010. Accessed August 26, 2017.

2. US Food and Drug Administration. *Guidance for Industry: Suicidal Ideation and Behavior: Prospective Assessment of Occurrence in Clinical Trials*. Silver Spring, MD: US Food and Drug Administration; 2012. https://www.fda.gov/drugs/guidancecomplianceregulatoryinformation/guidances/ucm315156.htm. Accessed June 7, 2017.

3. Hays, R. D., Bjorner, J., Revicki, R. A., Spritzer, K. L., & Cella, D. (2009). Development of physical and mental health summary scores from the Patient Reported Outcomes Measurement Information System (PROMIS) global items. Quality of Life Research, 18(7),873-80. (PMCID: PMC2724630) URL http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Global_Scoring_Manual.pdf

6 APPENDICES

6.1 PROMIS T-Score Conversions

The following conversion tables[3] allow a user to convert Pain Interference total raw scores into T-scores. T-score distributions are standardized such that a 50 represents the average (mean) for the US general population, and the standard deviation around that mean is 10 points. A high score always represents more of the concept being measured. Thus, a subject who has a T-score of 60 is one standard deviation better (healthier) than the general population.

Pain Interference 8a - Pediatric v2.0		
Short Form Conversion Table		
Raw Score	T-Score	SE*
8	34.0	5.6
9	38.7	4.4
10	40.6	4.2
11	42.7	3.8
12	44.3	3.7
13	45.8	3.4
14	47.1	3.3
15	48.4	3.2
16	49.5	3.2
17	50.6	3.1
18	51.7	3.1
19	52.7	3.1
20	53.7	3.0
21	54.7	3.0
22	55.7	3.0
23	56.6	3.0
24	57.6	3.0
25	58.5	3.0
26	59.5	3.0
27	60.4	3.0
28	61.4	3.0
29	62.4	3.0
30	63.4	3.0
31	64.4	3.0
32	65.4	3.1
33	66.5	3.1
34	67.6	3.2
35	68.8	3.2
36	70.1	3.3
37	71.5	3.4
38	73.2	3.7
39	75.0	3.8
40	78.0	4.3

*SE = Standard Error on T-Score

Pain Interference 8a - Parent Proxy v2.0		
Short Form Conversion Table		
Raw Score	T-Score	SE*
8	38.0	6.0
9	44.0	3.0
10	46.0	3.0
11	48.0	3.0
12	49.0	2.0
13	50.0	2.0
14	51.0	2.0
15	52.0	2.0
16	53.0	2.0
17	54.0	2.0
18	55.0	2.0
19	56.0	2.0
20	57.0	2.0
21	58.0	2.0
22	58.0	2.0
23	59.0	2.0
24	60.0	2.0
25	61.0	2.0
26	62.0	2.0
27	62.0	2.0
28	63.0	2.0
29	64.0	2.0
30	65.0	2.0
31	66.0	2.0
32	67.0	2.0
33	67.0	2.0
34	68.0	2.0
35	69.0	2.0
36	70.0	2.0
37	71.0	3.0
38	73.0	3.0
39	74.0	3.0
40	78.0	4.0

*SE = Standard Error on T-Score

6.2 Potentially Clinically Significant Abnormalities criteria

PCS criteria for laboratory parameters

Laboratory Parameter	**PCS Low**	**PCS High**
Aspartate aminotransferase (AST)	N/A	3 x ULN
Alanine aminotransferase (ALT)	N/A	3 x ULN
Amylase	N/A	3 x ULN
Lipase	N/A	3 x ULN
Total bilirubin	N/A	>2 mg/dL
Total thyroxine (T4)	<4.0 mcg/dL	>13.0 mcg/dL
Total cholesterol	N/A	>300 mg/dL
Triglycerides	N/A	>400 mg/dL
White blood cell (WBC) count	$<2.8 \times 10^3/mm^3$	$>16 \times 10^3/mm^3$
Hemoglobin	<9.5 g/dL	>19 g/dL
Hematocrit	<34%	>54%
Platelet count	$<75 \times 10^3/mm^3$	$>700 \times 10^3/mm^3$
AST, ALT and total bilirubin	N/A	AST or ALT > 3 x ULN and total bilirubin > 2 x ULN

PCS criteria for ECG parameters

Parameter	**PCS Low**	**PCS High**
PR Interval	None	1) >200 ms only OR 2) increase from baseline ≥20 ms only OR 3) >200 ms and increase from baseline ≥20 ms
QRS Interval	None	1) >100 ms only OR 2) increase from baseline ≥10 ms only OR 3) >100 ms and increase from baseline ≥10 ms
QT Interval	None	1) >500 ms only OR 2) increase from baseline ≥60 ms only OR 3) >500 ms and increase from baseline ≥60 ms
QTcF and QTcB Interval	None	1) >500 ms only OR 2) increase from baseline ≥60 ms only OR 3) >500 ms and increase from baseline ≥60 ms

PCS criteria for vital signs

Vital Sign	PCS Low	PCS High
Sitting SBP (mm Hg)		
Ages 0-12 years	<60 mm Hg OR a decrease of 20 mm Hg or more from baseline	>150 mm Hg OR an increase of 20 mm Hg or more from baseline
Ages >12 years	<86 mm Hg OR a decrease of 25 mm Hg or more from baseline	>180 mm Hg OR an increase of 25 mm Hg or more from baseline
Sitting DBP (mm Hg)		
Ages 0-12 years	<30 mm Hg OR a decrease of 15 mm Hg or more from baseline	>90 mm Hg OR an increase of 15 mm Hg or more from baseline
Ages >12 years	<48 mm Hg OR a decrease of 20 mm Hg or more from baseline	>110 mm Hg OR an increase of 20 mm Hg or more from baseline
Heart rate	1) <55 bpm OR 2) decrease of ≥20 bpm from baseline OR 3) <55 bpm AND decrease from baseline of ≥20 bpm	1) >120 bpm AND increase from baseline ≥20 bpm OR 2) >140 bpm

6.3 Reference Values for Normal Joint Range of Motion

As published by the CDC (https://www.cdc.gov/ncbddd/jointROM/), "the following table provides the reference values along with 95% confidence intervals for normal range of motion for 11 measurements taken on 5 joints. Values are provided separately by sex and age."

	Age 2–8	
Motion	**Females**	**Males**
Hip extension	26.2 (23.9 – 28.5)	28.3 (27.2 – 29.4)
Hip flexion	140.8 (139.2 – 142.4)	131.1 (129.4 – 132.8)
Knee flexion	152.6 (151.2 – 154.0)	147.8 (146.6 – 149.0)
Knee extension	5.4 (3.9 – 6.9)	1.6 (0.9 – 2.3)
Ankle dorsiflexion	24.8 (22.5 – 27.1)	22.8 (21.3 – 24.3)
Ankle plantar flexion	67.1 (64.8 – 69.4)	55.8 (54.4 – 57.2)
Shoulder flexion	178.6 (176.9 – 180.3)	177.8 (176.7 – 178.9)
Elbow flexion	152.9 (151.5 – 154.3)	151.4 (150.8 – 152.0)
Elbow extension	6.8 (5.2 – 8.4)	2.2 (0.9 – 3.5)
Elbow pronation	84.6 (82.8 – 86.4)	79.6 (78.8 – 80.4)
Elbow supination	93.7 (91.4 – 96.0)	86.4 (85.3 – 87.5)

	Age 9–19	
Motion	**Females**	**Males**
Hip extension	20.5 (18.6 – 22.4)	18.2 (16.6 – 19.8)
Hip flexion	134.9 (133.0 – 136.8)	135.2 (133.0 – 137.4)
Knee flexion	142.3 (140.8 – 143.8)	142.2 (140.4 – 144.0)
Knee extension	2.4 (1.5 – 3.3)	1.8 (0.9 – 2.7)
Ankle dorsiflexion	17.3 (15.6 – 19.0)	16.3 (14.9 – 17.7)
Ankle plantar flexion	57.3 (54.8 – 59.8)	52.8 (50.8 – 54.8)
Shoulder flexion	171.8 (169.8 – 173.8)	170.9 (169.1 – 172.7)
Elbow flexion	149.7 (148.5 – 150.9)	148.3 (146.8 – 149.8)
Elbow extension	6.4 (4.7 – 8.1)	5.3 (3.6 – 7.0)
Elbow pronation	81.2 (79.6 – 82.8)	79.8 (77.8 – 81.8)
Elbow supination	90.0 (88.0 – 92.0)	87.8 (85.7 – 89.9)

Reference: Soucie JM, Wang C, Forsyth A, Funk S, Denney M, Roach KE, Boone D, and the Hemophilia Treatment Center Network. Range of motion measurements: reference values and a database for comparison studies. Haemophilia 2010; e-pub November 11, 2010.