STATISTICAL ANALYSIS PLAN

A Phase 3, Efficacy and Safety Study of Oral Palovarotene for the Treatment of Fibrodysplasia Ossificans Progressiva (FOP)

Study PVO-1A-301 (MOVE Trial)

Sponsor: Clementia Pharmaceuticals Inc.
1000 De La Gauchetière West, Suite 1200
Montreal, QC H3B 4W5 Canada

Sponsor Representative: PI , PhD
PI , Biostatistics and Data Management

Document Date: 9 September 2020

Document Version: 1.2

STATISTICAL ANALYSIS PLAN

A Phase 3, Efficacy and Safety Study of Oral Palovarotene for the Treatment of Fibrodysplasia Ossificans Progressiva (FOP)

Study PVO-1A-301 (MOVE Trial)

Cytel Author/Date:

_____

PI , PhD
PI
PI , North America

Sponsor Approver/Date:

_____

PI , PhD
PI , Biostatistics and Data Management

_____

PI , MD
PI

TABLE OF CONTENTS

LIST OF TABLES

ABBREVIATIONS

Abbreviation	Definition
AE	adverse events
AUC	area under the concentration versus time curve
CAJIS	Cumulative Analogue Joint Involvement Scale
CI	confidence interval
C_{max}	maximum plasma concentration
C_{min}	trough plasma concentration, taken 24 hours after dose and prior to subsequent dose
C-SSRS	Columbia-Suicide Severity Rating Scale
CTCAE	Common Terminology Criteria for Adverse Events
ECG	Electrocardiogram
EP	Enrolled Population
FAS	Full Analysis Set
FDA	Food and Drug Administration
FOP	Fibrodysplasia Ossificans Progressiva
FOP-PFQ	FOP-Physical Function Questionnaire
GEE	generalized estimating equation
HO	heterotopic ossification
ICH	International Council for Harmonisation
MedDRA	Medical Dictionary for Regulatory Activities
NHS	Natural History Study
PAS	Pharmacokinetic Analysis Set
PK	pharmacokinetic(s)
PPC	Premature physeal closure
PPS	Per Protocol Set
PROMIS	Patient Reported Outcomes Measurement Information System
ROM	Range of motion
SAE	serious adverse event
SAS	Safety Analysis Set
SAP	statistical analysis plan
SD	standard deviation
T_{max}	time to maximum plasma concentration
WBCT	whole body computed tomography
wLME	weighted linear mixed effect

1 OVERVIEW AND INVESTIGATIONAL PLAN

1.1 Context for Version 1.2

This amendment is introduced after the third interim analysis. It introduces edits, including changes to existing analyses and additions of new analyses, to be included in the interim CSR for Study PVO-1A-301 (MOVE Trial) to be included in the palovarotene marketing authorization application for FOP. A list of changes is included in Section 1.8.3

1.2 Context for Version 1.1

This amendment is introduced after the second interim analysis but before the planned third interim analysis (anticipated to be held in mid-May 2020). The futility boundary was crossed at the second interim analysis, signifying that the prespecified Bayesian compound Poisson distribution model with square-root transformation indicated that the trial was unlikely to show sufficient evidence of benefit. The independent data monitoring committee (DMC) informed the sponsor of this outcome on January 15, 2020, and the efficacy assessments were subsequently unblinded to the sponsor.

Based on additional post-hoc analyses presented to the DMC on January 21, 2020, the DMC noted that using the square-root transformation of the data in the primary analysis appears to have moved the statistical conclusion from significant therapeutic benefit to showing futility of the treatment. The DMC also noted that the dilemma created by these highly disparate results precludes a confident conclusion about futility. As such, the DMC recommended that additional analyses be conducted and provided to them for consideration at the next interim analysis.

US FDA instituted a partial clinical hold on dosing of palovarotene due to premature physeal closure (PPC) for subjects <14 years of age on December 4, 2019; as of the finalization of this SAP amendment, dosing has not yet resumed for these subjects. Dosing on all other subjects not subject to US FDA's partial clinical hold was interrupted following the January 21, 2020 DMC meeting while the protocol amendment allowing the continuation of the study after the futility outcome and incorporating the additional analyses requested by the DMC was reviewed by regulatory agencies, IRBs, and investigational sites. These subjects had palovarotene dosing interrupted on or around January 24, 2020. The sponsor allowed dosing to resume as of March 26, 2020, provided each individual site had obtained Ethics Committee approval to do so and the site was able to fulfil regulatory and operational requirements. Subjects were expected to restart dosing over a span of weeks due to these logisitical considerations.

Note that Sections 1.3 (Study design), 1.4 (Objectives), 1.5 (Determination of sample size) and 1.6 (Study plan) were unchanged from the SAP V1.0, which was finalized prior to the first interim analysis. New content began in Section 1.7 (Modifications to the statistical section of the protocol).

1.3 Study Design

Study PVO-1A-301 (referred to as the MOVE Trial) is a Phase 3, multicenter, open-label study in adult and pediatric subjects with fibrodysplasia ossificans progressiva (FOP). Sources of subjects eligible for enrollment in the MOVE Trial include: (1) subjects from Study

PVO-1A-001 (natural history study [NHS]); (2) additional subjects clinically diagnosed with FOP with the R206H ACVR1 mutation or other FOP variants reported to be associated with progressive heterotopic ossification (HO) (who have not previously participated in any Clementia-sponsored study); and (3) subjects in Study PVO-1A-202 or Study PVO-1A-204 who cannot receive the chronic/flare-up regimen due to country of residence or those traveling long distances to participate in the Phase 2 trial. Approximately 90 subjects will be enrolled into the MOVE Trial (approximately 80 with the R206H mutation and no previous exposure to palovarotene, and approximately 10 with other mutations or previous participation in the Phase 2 trials) and receive chronic dosing for up to 24 months and undergo flare-up treatment should they experience an eligible flare-up confirmed by the Investigator.

This is an open-label study, with the comparator arm being subject data from the NHS. The efficacy of palovarotene treatment in the MOVE Trial, as measured by the change in new HO volume (the primary efficacy endpoint), will be compared to that observed in untreated subjects who participated in the NHS.

There are three formal interim efficacy analyses and a final analysis planned; an additional descriptive analysis occurred before the first formal interim efficacy analysis to be included in the then-planned marketing authorization filing for episodic palovarotene prevention of HO in patients with FOP. The interim efficacy analyses will be conducted to assess whether efficacy of the chronic/flare-up regimen is sufficiently demonstrated prior to the completion of the study. The study will not stop early if an interim efficacy analysis meets the efficacy criteria (pre-specified to trigger potential early submission for marketing authorization); subjects will continue treatment until completion of two years of palovarotene treatment.

A futility analysis will be conducted at the second and third interim analyses to assess whether the study should be stopped due to insufficient evidence of efficacy.

The statistical analyses and summary tabulations described in this Statistical Analysis Plan (SAP) will provide the basis for the results sections of the MOVE clinical study report.

1.4 Objectives

1.4.1 Primary Objectives

- To evaluate the efficacy of palovarotene in decreasing heterotopic ossification (HO) in adult and pediatric subjects with FOP as assessed by low-dose, whole body computed tomography (WBCT), excluding head, as compared to untreated subjects from Clementia's FOP natural history study.

- To evaluate the safety of palovarotene in adult and pediatric subjects with FOP.

1.4.2 Secondary Objectives

- To evaluate the effect of palovarotene on flare-up rate and proportion of subjects reporting at least one flare-up.

- To evaluate the effect of palovarotene on range of motion (ROM) as assessed by the Cumulative Analogue Joint Involvement Scale (CAJIS) for FOP.

- To evaluate the effect of palovarotene on physical function using age appropriate forms of the FOP Physical Function Questionnaire (FOP-PFQ).

- To evaluate the effect of palovarotene on physical and mental health using age-appropriate forms of the Patient Reported Outcomes Measurement Information System (PROMIS) Global Health Scale.

- To evaluate the pharmacokinetics of palovarotene.

1.5 Determination of Sample Size

The sample size required for the MOVE Trial was determined via simulation based on the available WBCT HO volumes from the NHS and the observed efficacy of palovarotene treatment in Phase 2. The NHS will serve as the external control group with, for the purposes of the determination of sample size, the assumption that 45 subjects will have baseline, 1-year, and 1.5-year WBCT HO volumes and 45 subjects will have baseline, 1-year, and 2-year WBCT HO volumes. It is also assumed that approximately 50% of these 90 subjects will enter the Phase 3 study to begin palovarotene treatment.

There are three interim efficacy analyses – with futility analyses at the second and third interim analyses – and a final analysis planned. Assuming a one-sided, overall type I error rate of 2.5%, the Lan-DeMets alpha-spending function with O'Brien-Fleming parameterization was used to determine stopping boundaries (see Section 2.4.1 for additional details).

Available pairs of NHS WBCT scans for 65 subjects taken one year apart and read using the original NHS single-reader process were used to tabulate the number of body regions with new HO and the mean and standard deviation (SD) of new HO volume in each body region (Table 1). The original 15 body regions from the NHS WBCT scans were mapped to approximate the 9 body regions that will be used in the MOVE Trial. The mean (SD) number of body regions with new HO per subject in approximately one year was 0.707 (1.11), i.e., there were 46 body regions with new HO in 65 subjects.

Table 1. NHS WBCT Scan Results by Body Region

Body region	Number of observations	Number of regions with new NO	Proportion of regions with new HO	Volume change given new HO in region (in thousands mm^3)	
				Mean	SD
Neck, upper spine and chest	64	12	0.19	25.83	21.06
Shoulder (left, right)	130	7	0.05	4.69	3.08
Elbow, hand and wrist (left, right)	125	1	0.01	31.27	-
Lower spine and abdomen	65	11	0.17	16.22	19.24
Hip (left, right)	130	9	0.07	81.17	89.65
Knee and distal lower extremity (left, right)	128	6	0.05	9.04	14.64

Simulations to assess power were conducted as follows:

1. A subject-level rate of body regions with new HO were randomly drawn from a gamma distribution with mean = 0.707 and SD = 1.11.

2. Using the subject-level rate, the numbers of regions with new HO for each WBCT scan were simulated using a Poisson distribution. The rate was scaled to the duration of time between each WBCT scan and the previous WBCT scan.

3. Given the number of regions with new HO, the specific body regions with new HO were randomly selected assuming equal probability across regions.

4. For each body region with new HO, the increase in HO volume was simulated from a gamma distribution parameterized with the mean and SD in Table 1.

    a. For the elbow, hand and wrist (left, right), the SD was assumed to be equal to the mean.

5. The resulting dataset was analyzed as described in Section 2.4.1 at the scheduled analysis time points.

Note that drawing subject-level rates of body regions with new HO from a gamma distribution and then drawing from Poisson distributions using these subject-level rates implies that the marginal distribution of the number of body regions with new HO follows a negative binomial distribution, which requires much less restrictive assumptions than a Poisson distribution and does not tend to result in an underestimate of the variance.

The simulated power is summarized in Table 2 given the specified testing sequence, 80 subjects enrolled into the MOVE Trial with the R206H mutation and no previous palovarotene experience (including approximately 45 subjects from the NHS) and the NHS serving as the external control group with 45 subjects with baseline, 1-year and 1.5-year WBCT HO volumes

and 45 subjects with baseline, 1-year and 2-year WBCT HO volumes; it is assumed that the full NHS dataset will be available at the first interim analysis.

Table 2. Power of MOVE Study Primary Efficacy Analysis Assuming 80 Subjects Enrolled with the R206H Mutation and No Previous Palovarotene Exposure

Scenario		Cumulative Probability of Success			
% Reduction HO Volume Conditional on New HO in a Body Region	% Reduction in Number of Regions with New HO	Interim #1	Interim #2	Interim # 3	Final
0%	0%	0.006	0.010	0.013	0.021
50%	30%	0.507	0.789	0.886	0.922

The primary efficacy analysis comparing the annualized change in new HO volume between subjects treated with palovarotene and untreated subjects has an overall probability of study success of 0.92 if palovarotene treatment reduces the number of regions with new HO by 30% and reduces the volume of new HO conditional on new HO in a per body region by 50%. With this treatment effect, the probability of declaring study success is 0.51, 0.79, and 0.89 at the first, second, and third interim analyses, respectively. Table 2 shows that the type I error rate is maintained under the null hypothesis of no palovarotene treatment effect.

Note that Protocol Amendment 2 increased the number of subjects enrolled from approximately 90 to a maximum of 110 (up to 99 with the R206H mutation and no previous exposure to palovarotene; up to 11 with other mutations and previous participation in the Phase 2 trials). This increase in the number of enrolled subjects, implemented late in enrollment to allow sites to screen already-scheduled patients and enroll those found to be eligible, does not substantially change the power of the trial or have a significant impact on the primary or secondary statistical analyses.

Additional details of the operating characteristics of the Bayesian analysis including its chance of producing erroneous conclusions and the reliability of treatment effect estimates will be included in a separate simulation report, which will be included in the interim clinical study report as an appendix.

1.6 Study Plan

The schedule of the safety, efficacy, and pharmacokinetic (PK) assessments can be found in the Protocol in Table 1 and Table 2.

1.7 Modifications to the Statistical Section of the Protocol

1.7.1 Modifications to the Statistical Section of Protocol Amendment #1

The method for the primary efficacy analysis (a Bayesian analysis of a compound Poisson distribution) described in Protocol Amendment 1 and this SAP is different from the method included in the original protocol dated 10 July 2017 (weighted linear mixed effects model). The

method and accompanying power simulations in the original protocol were based on a March 2017 evaluation of the available NHS 12-month WBCT data. In a subsequent dataset, increased variability in the change from baseline in new HO resulted in a substantive power decrease for the weighted linear mixed effects model. Therefore, additional modeling and simulation work was undertaken to assess whether alternative models or transformations would be more appropriate methods of analyzing the data, and thus maintain the power of the MOVE Trial. This resulted in changes to the primary efficacy analysis method as well as the simulations underlying sample size determination. Additionally, the Lan-DeMets alpha-spending function with Pocock parameterization was revised to O'Brien-Fleming parameterization based on FDA scientific advice.

1.7.2 Modifications to the Statistical Section of Protocol Amendment #2

Rather than at successive 6-month intervals after the first interim analysis, the timing of the second and third interim analyses were clarified; they were to be performed when all subjects in the Principal EP have completed (i.e., have WBCT data) 12 months and then 18 months of follow-up, respectively.

1.7.3 Modifications to the Statistical Section of Protocol Amendment #3

Protocol Amendment 3 added a 2-year extension to the study; the first 24 months, i.e., the original study is designated Part A, and the 2-year extension is designated Part B.

Note that the current scope of this document remains Part A. The SAP may be amended again to describe analyses conducted on data collected in Part B.

1.7.4 Modifications to the Statistical Section of Protocol Amendment #4

Protocol Amendment 4 followed the second interim analysis at which the futility boundary was crossed, signifying that the prespecified Bayesian compound Poisson distribution model with square-root transformation indicated that the trial was unlikely to show sufficient evidence of benefit. The DMC informed the sponsor of this outcome on January 15, 2020, and the efficacy assessments were subsequently unblinded to the sponsor.

Based on additional post-hoc analyses presented to the DMC on January 21, 2020, the DMC noted that using the square-root transformation of the data in the primary analysis appears to have moved the statistical conclusion from significant therapeutic benefit to showing futility of the treatment. The DMC also noted that the dilemma created by these highly disparate results precludes a confident conclusion about futility. As such, the DMC recommended that additional analyses be conducted and provided to them for consideration at the next (ie, third) interim analysis. Therefore, the study continued and the protocol was amended to allow for the continuation and to include these analyses requested by the DMC.

1.8 Statistical Modifications to the Statistical Analysis Plan

1.8.1 Statistical Modifications to the Statistical Analysis Plan, V1.0

The primary and secondary efficacy analyses described in Protocol Amendment 1 do not include explanatory covariates. Because recent studies – Clementia's NHS as well as other studies – suggest that age may be important for flare-up rate and the formation of new HO, age (<18, ≥18 years) and additional categorical variable sex (Male, Female) is included in the primary and secondary efficacy analyses.

The analysis of the secondary endpoints proportion of subjects with any new HO and number of body regions with new HO will both be conducted using the Bayesian compound Poisson model used in the primary efficacy analysis; in fact, the inference is identical for these two secondary endpoints. Within the compound Poisson model, $\theta_{1,1}<1$ implies both a lower proportion of subjects with any new HO and a smaller number of body regions with new HO for palovarotene-treated subjects.

1.8.2 Statistical Modifications to the Statistical Analysis Plan, V1.1

SAP Version 1.1 introduces methods to address the prolonged interruptions in subject dosing in MOVE due to two events:

1) US FDA instituted a partial clinical hold on dosing of palovarotene due to premature physeal closure (PPC) for subjects <14 years of age. This hold was effective December 4, 2019; as of the finalization of this SAP amendment, dosing has not yet resumed for these subjects.
2) Dosing on all subjects not subject to US FDA's partial clinical hold was interrupted following the January 21, 2020 DMC meeting while the protocol amendment allowing the continuation of the study after the futility outcome and incorporating the additional analyses requested by the DMC was reviewed by regulatory agencies, IRBs, and investigational sites. These subjects had palovarotene dosing interrupted on or around January 24, 2020. Following clearance received from several regulatory agencies including FDA to reinitiate dosing in subjects age 14 and above, the sponsor allowed dosing to resume as of March 26, 2020, provided each individual site obtains Ethics Committee approval to do so and the site is able to fulfil regulatory and operational requirements. Subjects were expected to restart dosing over a span of weeks due to these logisitical considerations.

Specifically, analyses of all efficacy endpoints will include assessments collected on or before these interruptions, unless otherwise specified. (Assessments were to continue as planned during the interruptions.) Analyses in which all timepoints are included, irrespective of dosing interruption, will also be conducted for the primary analysis (and additional analysis without the square-root transformation) and secondary analyses to aid in the interpretation of the efficacy.

The sponsor also notes that after the DMC communicated to the sponsor that the pre-specified futility boundary had been crossed, the sponsor became unblinded to all study data. With these

considerations, all subsequent analyses of efficacy data will be supplied without applying thresholds for success or futility, as these thresholds are no longer relevant.

SAP Version 1.1 also

- Provides a detailed description of the wLME analysis designated as the primary efficacy analysis in the original clinical protocol and reintroduced as a supportive analysis in Protocol Amendment 4.
- Introduces several age (years) by sex partitions for use in subgroup analyses, including <8 for Females and <10 for Males vs. ≥8 for Females and ≥10 for Males; and <13 for Females and <15 for Males vs. ≥13 for Females and ≥15 for Males. These are intended to inform the benefit/risk assessment for younger subjects at risk of PPC.
- Adds a subgroup of Asian subjects (to support interactions with global regulatory authorities).
- Describes a sensitivity analysis in which the impact of the difference in WBCT visit schedules between MOVE and NHS is assessed by analyzing the data as if the WBCT assessments in MOVE were conducted annually rather than biannually. This is intended to investigate whether the differences in visit schedules introduced bias, with and without use of the square-root transformation.
- Defines exploratory endpoints of incidence and volume of "catastrophic HO" based on the observation – made after unblinding of the second interim analysis – that a minority of subjects develop large amounts of HO in a year; post-hoc cut-offs of at least 50,000 mm^3 and 30,000 mm^3 of new HO are used.
- Describes the analyses to be conducted at the third interim analysis, the results of which will be provided to the DMC to assess efficacy and make a benefit:risk assessment. No predefined efficacy rules for success or futility will be used.

1.8.3 Statistical Modifications to the Statistical Analysis Plan, V1.2

SAP Version 1.2 introduces the following changes:

- Clarifies that drawing subject-level rates of body regions with new HO from a gamma distribution and then drawing from Poisson distributions using these subject-level rates implies that the marginal distribution of the number of body regions with new HO follows a negative binomial distribution, which requires much less restrictive assumptions than a Poisson distribution and does not tend to result in an underestimate of the variance.
- Notes that additional details of the operating characteristics of the Bayesian compound Poisson analysis, including its chance of producing erroneous conclusions and the reliability of treatment effect estimates, will be included in a separate simulation report.
- Clarifies the secondary and exploratory endpoint definitions to add timepoints at which the endpoints will be compared between NHS and MOVE.
- Clarifies the definitions of treatment-emergent AEs and post-treatment AEs.
- Removes summaries of retinoid-associated AEs, as they already appear in the AE tables.
- Clarifies that a listing of traumas leading to flare-up will be generated.

- Adds summaries of imaging performance, i.e., intra-reader agreement and inter-reader agreement, and quality metrics.
- Adds summaries of dosing compliance.
- Clarifies the wLME analysis method by adding pseudocode.
- Clarifies the methods used to prepare the dataset for application of the Bayesian compound Poisson analyses and square-root transformation, including the methods used to 'zero out' negative changes in HO volume.
- Clarifies the purpose, timing and conduct of the sensitivity analyses.
- Updates the list of subgroups to be studied to be consistent with feedback from FDA.
- Adds analyses for the HO-based secondary analyses and simplifies the analyses of the flare-up-based secondary endpoints.
- Adds summaries of growth velocity.
- Summarizes duration of flare-up treatment and duration of non-flare-up treatment, which are used in the summary of compliance.

2 STATISTICAL AND ANALYTICAL PROCEDURES

Data analyses will be performed for the MOVE Trial study period. Where applicable, the analyses will incorporate data from the NHS for comparative purposes. The analysis results will be presented by two treatment groups (palovarotene and untreated). The list of analysis endpoints described below is specific to the MOVE Trial, unless otherwise specified. The incorporation of NHS data is described in Section 2.2, Section 2.3, and Section 2.4.

2.1 Analysis Endpoints

2.1.1 Demographic and Baseline Characteristics

Demographic Characteristics

The following demographic characteristics will be summarized:

- 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)
- Asian (Yes, No)
- Age (years)
- Age group (<12, ≥12, <14, ≥14, <17, <18, ≥18 years, ≥12 - <17 years)
- Age by sex (<8 for Females and <10 for Males, ≥8 for Females and ≥10 for Males; ≥8 - <14 for Females and ≥10 - <14 for Males)

Medical or Surgical History

This information will be coded using the most current version of the Medical Dictionary for Regulatory Activities (MedDRA). FOP history will also be summarized.

FOP History

The following FOP history will be summarized:

- Age at diagnosis (years)
- Time since FOP diagnosis (years) (Until enrollment, i.e., informed consent)
- Great toe malformations (yes/no)
    - If yes, age when noted (years)
    - Surgically repaired? (yes/no)
- Other associated clinical findings (frequency table for findings, e.g., cervical spine malformations, hearing loss, etc.)
- Osteochondromas (yes/no, with frequency table for location, e.g., tibia, femur, etc.)
- How was the diagnosis made? (frequency table for method, i.e., clinical symptoms, genotyping, imaging w/ modality)
- Were there any misdiagnoses before proper diagnosis was made? (yes/no, with frequency table for misdiagnosis, e.g., Aggressive Juvenile Fibromatosis, Lymphedema, Soft Tissue Sarcoma, Other (with frequency table for free text entries)
- Any family history of FOP? (yes/no)
    - If yes, family member with disease (Mother, Father, Both, Unknown)
    - Other family members with FOP symptoms? (Brother, Sister, Both, Unknown, None)
    - How was diagnosis made in family member? (frequency table for diagnostic, i.e., Clinical Symptoms, Genotyping, Biopsy, Unknown)

Flare-up History

The following flare-up history will be summarized:

- Subject with any history of a flare-up (yes/no)
- Age at 1st flare-up (years)

- Time since 1st flare-up (years)
- Location of first flare-up
- Number of flare-ups within the past 12 months
- Time since last flare-up (months)
- Location of last flare-up
- Symptoms of last flare-up
- Cause of last flare-up
- Last flare-up bone formation (yes/no)
- Last flare-up movement restriction

2.1.2 Prior or Concomitant Medications

All medications taken within 30 days before study drug initiation and until the end of the study are to be reported in the case report form (CRF) pages.

All medications captured in the MOVE study will be coded using the using WHODrug Global; for NHS, all medications not appearing on the prespecified list were also coded using this dictionary.

- Prior medications are those the subject used prior to first dose of study drug. 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 by the subject after first dose of study drug, during the treatment-emergent period (Section 2.1.3). Concomitant medications first received after first dose of study drug will also be referred to as new onset medications.
- New onset medications are concomitant medications first received after the first dose of study drug.

A given medication can be classified as a prior medication, as a concomitant medication, and as a new onset medication. Note that these categories are not mutually exclusive. A medication initiated prior to first dose of study drug and continued during the treatment-emergent period will be prior and concomitant.

2.1.3 Efficacy Endpoints

Primary Efficacy Endpoint

The primary efficacy endpoint is the annualized change in new HO volume as assessed by low-dose, WBCT, excluding head. The annualized change in new HO volume will be compared between treated and untreated subjects. WBCT scans are performed at baseline, Month 12, Month 24 and Month 36 in the NHS and at baseline, Month 6, Month 12, Month 18 and Month 24 in the MOVE Trial. WBCT scans may also be performed at the time of early study discontinuation.

As described in the MOVE Trial and NHS Independent Review Charter, HO volume is assessed by two independent readers blinded to study and timepoint, with adjudication by a third independent reader. HO volume is measured by body region with the baseline scan. The following 9 anatomical regions will be used to describe HO location:

- Right chest, neck, shoulder through mid-humerus
- Left chest, neck, shoulder through mid-humerus
- Mid Torso: torso below shoulder regions and above iliac crests
- Right arm: mid-humerus through hand
- Left arm: mid-humerus through hand
- Right hip: iliac crest, hip through mid-femur
- Left hip: iliac crest, hip through mid-femur
- Right lower leg: mid-femur through foot
- Left lower leg: mid-femur through foot

In post-baseline scans, the presence of new HO within each body region and, if present, the HO volume within the body region, are measured. The primary efficacy analysis described in Section 2.4.1 models the number of body regions with new HO compared to the previous scan and the volume of new HO within each body region with new HO compared to the previous scan as the unit of measurement.

Note that the primary endpoint includes annualized HO based on the new HO observed up to Month 24 in MOVE and up to Month 36 in NHS. The sponsor believes that the new HO observed between Month 24 and Month 36 in NHS is relevant and that annualization is sufficient to protect against bias. Sensitivity analyses to assess impact of differences in length of follow-up between NHS and MOVE are described in Section 2.4.4.

The primary efficacy endpoint, as well as the secondary endpoints that depend on the efficacy imaging, will not be summarized in the descriptive analysis to be conducted before the first formal interim efficacy analysis, due to the ongoing nature of the blinded reads.

Secondary Efficacy Endpoints

The following secondary efficacy endpoints will be assessed:

- The proportion of subjects with any new HO at Month 12 (key secondary endpoint).
- The number of body regions with new HO at Month 12.
- The proportion of subjects reporting flare-ups at Month 12.
- The flare-up rate per subject-month exposure through Month 24.

Note that only MOVE flare-ups with ≥2 symptoms will be included in the derivation of the proportion of subjects reporting flare-ups and flare-up rate per subject-month exposure. This will better align the MOVE derivation with the definition of flare-up in the NHS, which required >1 symptom.

The secondary endpoints that depend on the efficacy imaging, i.e., the proportion of subjects with any new HO and the number of body regions with new HO, will not be summarized in the descriptive analysis to be conducted before the first formal interim efficacy analysis, due to the ongoing nature of the blinded reads.

Exploratory Efficacy Endpoints

The following exploratory efficacy variables will be assessed:

- Change from baseline in ROM assessed by CAJIS through Month 24 in MOVE and NHS.
- Change from baseline in physical function through Month 24 in MOVE and NHS, using age-appropriate forms of the FOP-PFQ.
- Change from baseline in physical and mental function for subjects ≥15 years old and mental function for subjects <15 years old using age appropriate forms of the PROMIS Global Health Scale, through Month 24 in MOVE and NHS.
- Incidence and volume of catastrophic HO per year, i.e., new HO >50,000 mm^3; and new HO >30,000 mm^3 at Month 12 and Month 24.

ROM Assessed by CAJIS

Range of motion will be assessed using the CAJIS. The assessments will be performed on 12 joints (shoulder, elbow, wrist, hip, knee, and ankle on both the right and left sides), and three

body regions (jaw, cervical spine [neck], and thoracic/lumbar spine). Each joint/region will be scored as: 0 = uninvolved; 1 = partially involved; and 2 = ankylosed/completely involved.

The CAJIS Total score will be calculated as the sum of the scores of all joints/regions and ranges from 0 (no involvement) to 30 (maximally involved).

The CAJIS Upper Extremities subscore will be calculated as the sum of the scores from six joints (shoulder, elbow, and wrist on both the right and left sides) and one region (cervical spine [neck]) and ranges from 0 (no involvement) to 14 (maximally involved).

The CAJIS Mobility subscore (lower extremity) will be calculated as the sum of scores from six joints (hip, knee, and ankle on both the right and left sides) and ranges from 0 (no involvement) to 12 (maximally involved).

FOP-Physical Function Questionnaire (FOP-PFQ)

Age-appropriate forms of the FOP-PFQ will be administered to subjects. The adult form will be administered to subjects ≥15 years old. For subjects between 8 and 14 years of age, both the pediatric self-completed and the pediatric proxy-completed forms will be administered. The proxy-completed form will be used for analyses unless only the self-completed form is available.

The FOP-PFQ consists of 28 questions on the adult form and 26 questions on the pediatric form scored on a scale from 1 to 5, with lower scores indicating more difficulty.

The total score will be calculated as:

- The sum of the scores from each question and ranges from 28x1 = 28 to 28x5 = 140 for the adult form.
- The sum of the scores from each question and ranges from 26x1 = 26 to 26x5 = 130 for the pediatric form.

If a subject is missing some (but not more than 20%) of the contributing question scores, the total score will be calculated as the average observed score multiplied by the number of expected question scores. For example, the total score would be calculated as the average of the non-missing scores x 28 for the adult form.

As the analysis for FOP-PFQ will be performed across all subjects (adult and pediatric) and the number of contributing questions differs, the scores will be transformed to reflect a percentage of worst score. The percentage of worst score ranges from 0% to 100% with 0% indicating the best possible function and 100% indicating the worst possible function. Table 3 illustrates some sample derivations of the percentage of worst score.

Table 3. Sample Derivations for the Percentage of Worst Score for the FOP-PFQ

Sample Subject	Observed FOP-PFQ Score	Lowest Possible Score	Highest Possible Score	Range of Possible Scores	Distance = Highest - Observed	Distance/ Range	Percentage of Worst Score
1	45	15	75	60	30	0.500	50.0%
2	40	15	75	60	35	0.583	58.3%
3	35	15	75	60	40	0.667	66.7%
4	30	15	75	60	45	0.750	75.0%
5	25	15	75	60	50	0.833	83.3%

PROMIS Global Health Scale

An age-appropriate form of the PROMIS Global Health short form will be provided to subjects (or parent proxy). The adult form will be provided to subjects 15 years and older. For subjects between 8 and 14 years of age, both the pediatric self-completed and the pediatric proxy-completed forms will be provided. The proxy-completed form will be used for analysis unless only the self-completed form is available.

The PROMIS Global Health short form consists of 10 questions on the adult form and nine questions on the pediatric form scored on varying scales.

For the adult form, Global Physical Health and Global Mental Health scores will be calculated. The Global Physical Health score will be calculated as the sum of scores from Questions 3, 6, 7, and 8, and will range from 4 (worse health) to 20 (better health). The Global Mental Health score will be calculated as the sum of scores from Questions 2, 4, 5, and 10, and will range from 4 (worse health) to 20 (better health). In the calculation of the Global Physical Health and Global Mental Health scores, the following questions will be rescaled as shown in Table 4.

Table 4. Rescaled PROMIS Global Health Scale Scores

Question(s)	Score	Rescaled Score
7	0	5
	1-3	4
	4-6	3
	7-9	2
	10	1
8 and 10	1	5
	2	4
	3	3
	4	2
	5	1

If a subject is missing any of the contributing scores, the Global Physical Health or Global Mental Health score will not be calculated.

For the pediatric form, only a total score will be calculated as the sum of scores from the first seven questions and will range from 7 (worse health) to 35 (better health). If a subject is missing some (but not more than three) of the contributing question scores, the total score will be calculated as the average observed score multiplied by the number of expected question scores. For example, the total score would be calculated as the average of the non-missing scores x 7.

Global Physical Health and Global Mental Health scores and total scores will be converted to T-scores for analysis as described in Appendix 6.1. A T-score of 50 is normal and increments of 10 ± 1 standard deviation away from the norm. A T-score <50 indicates worse health, while a T-score >50 indicates better health. Note that higher values indicate better health.

2.1.4 Safety Endpoints

Safety evaluations will include adverse event (AE) and serious AE (SAE) reporting, electrocardiograms, vital signs (temperature, respiratory rate, blood pressure, and heart rate), physical examination, body weight/BMI , knee and linear height, laboratory parameters (hematology, biochemistry, and urinalysis), urine pregnancy tests for FOCBP, radiographic imaging and whole body computed tomography (WBCT), concomitant medication reporting, and the age-appropriate C-SSRS.

Adverse Event Variables

Adverse Event Observation Periods

- A Pre-treatment AE is an AE with an onset date before the first dose of study drug.
- TEAEs are AEs with (first dose date of study drug) ≤ (AE start date) ≤ (last dose date of study drug +7 days for all AEs, i.e., non-serious and serious AEs).
    - A chronic TEAE is a TEAE with an onset date during chronic treatment
    - A flare-up TEAE is a TEAE with an onset date during flare-up treatment
- Post-treatment AEs are AEs with a start date > last dose date + 7 days for All AEs (non-serious and serious AEs).

Note that subjects may have multiple independent flare-up treatment cycles, each potentially with associated AEs, within the study.

For the efficacy interim analysis and final analysis, treatment emergence for chronic or flare-up treatment is determined based on the subject's dosing diary. For all other pre-lock analyses, treatment emergence is determined using the subject's planned dosing log if insufficient diary data is available, with the date of the extract of data for the interim analysis used as the end date for TEAE assignment purposes.

Any AE with a missing start date is considered TEAE. For TEAEs that have a partial date where the day is missing, if the month is the same month as a flare-up treatment was given, then the TEAE will be defined as a TEAE associated with flare-up dosing.

All AEs (including SAEs) will be coded using current version 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 AEs (including SAEs) will be recorded from the time of signed informed consent until the end of the study.

Deaths

Deaths after the first dose of study drug are either on-treatment or post-treatment:

- Deaths on-treatment: deaths occurring after the first dose of study drug and up to 7 days after the last dose of study drug
- Death post-treatment: deaths occurring more than 7 days after the last dose of study drug until 30 days after last dose of study drug

Laboratory Safety Variables

Clinical laboratory data will include hematology, biochemistry (including lipids and serum pregnancy test (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 and every 6 months (Months 6, 12, 18, and 24). For subjects receiving flare-up treatment, blood and urine samples will be collected every 12 weeks during safety assessments until the end of the flare-up cycle.

The laboratory parameters will be classified as listed in Table 5.

Table 5. Clinical Laboratory Parameters

Biochemistry:	
Sodium	Globulin
Potassium	Alkaline phosphatase (ALP)
Chloride	Aspartate aminotransferase (AST)
Bicarbonate	Alanine aminotransferase (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	
Lipid Profile:	
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[1]:	
pH	Blood (free hemoglobin)
Protein	Nitrite
Glucose	Urobilinogen
Ketones	Specific gravity
Bilirubin	Color & appearance

[1] If results are abnormal, then a microscopic evaluation will be completed.

Electrocardiogram Variables

A 12-lead electrocardiogram (ECG) will be performed at Screening and every 6 months (Months 6, 12, 18, and 24). Mean heart rate, RR, PR, QRS, QT, QTcF, and QTcB will be determined using centralized readings.

The abnormal ECG categories are defined as follows:

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
- 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

Vital Signs Variables

Vital sign variables include heart rate, systolic blood pressure, diastolic blood pressure, respiratory rate and temperature. These assessments will be performed at Screening, every 3 months (Months 3, 6, 9, 12, 15, 18, 21 and 24), on Flare-up Cycle Safety Day 1 and then every 12 weeks thereafter until treatment of the last flare-up or traumatic event within a flare-up cycle is completed.

Columbia-Suicide Severity Rating Scale

In accordance with the Guidance for Industry: Suicidal Ideation and Behavior: Prospective Assessment of Occurrence in Clinical Trials, 2012, all subjects ≥8 years old who are receiving palovarotene will be assessed for suicidal ideation and behavior using the C-SSRS at Screening and every 3 months (Months 6, 9, 12, 15, 18, 21, and 24); and as part of every Flare-Up Cycle Safety Assessment. The adult form will be used for subjects ≥12 years old and the pediatric form will be used for subjects 8 to 11 years old.

Knee and Hand/Wrist Radiographs and WBCTs for Assessment of Bilateral Hand/Wrist and Knee Epiphyseal Growth Plate

All subjects under the age of 18 years enrolled into the MOVE Trial will undergo hand/wrist radiographs (posterior/anterior (PA) view, preferably on the left side) and knee radiographs (AP view) at baseline and every 6 months (Months 6, 12, 18, and 24) to determine bone age (hand/wrist only) and skeletal maturity. Knee and hand/wrist radiograph assessments will be performed at Months 3, 9, 15, 21, etc. in those subjects who (1) received flare up dosing in the period of time since their last assessment; and (2) had not achieved 100% skeletally maturity on their last assessment. Once a subject has achieved 100% skeletal maturity (i.e., epiphyseal closure of all assessed growth plates) as determined by the knee and hand/wrist radiographs, further radiographs will no longer be necessary. Determination of epiphyseal closure requires consensus, with differences in the two assessments undergoing adjudication.

Although the WBCT scans obtained in the NHS and MOVE Trial are used primarily to quantify change in HO volume, they will also be used to assess for potential adverse effects on bilateral hand/wrist and left knee growth plates as well as long bone growth of growing children.

All radiographs and WBCT scans will be assessed by two independent musculoskeletal radiologists at a central imaging laboratory using standardized procedures to ensure consistent and accurate assessment. Bone age will be the average of the two independently assessed bone ages.

The following parameters will be assessed:

- Hand/wrist and knee radiograph images
    - Epiphyseal growth plate abnormalities assessment. Abnormalities to be reported include:
        - 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)
    - Epiphyseal growth plate closure assessment
    - Bone age (years, months), if applicable (hand/wrist only)

- Hand/wrist and knee WBCT safety images
    - Epiphyseal growth plate abnormalities assessment

Standardized Stadiometry and Knee Height for Assessments of Linear Growth

Subjects under the age of 18 years enrolled with open epiphyseal growth plates will have linear growth assessments (in triplicate) by stadiometer at Screening and every 6 months (Months 6, 12, 18, and 24). Measurement of knee height using a standard caliper will also be performed (in triplicate) at Screening and every 6 months (Months 6, 12, 18, and 24). Once a subject is 18 years old, triplicate linear growth and knee height measurement will no longer be required.

WBCT Measurements of Femur and Tibia Lengths

Although the WBCT scans obtained in the NHS and MOVE Trial are used primarily to quantify change in HO volume, they will also be used to assess the length of the femur and tibia bilaterally on all subjects aged 17 years and below.

Evaluation of Bilateral Hips by WBCT Safety Read

Although the WBCT scans obtained in the NHS and MOVE Trial are used primarily to quantify change in HO volume, they will also be used to assess potential adverse effects on the femoral head related to avascular necrosis of the hip (AVN).

Two independent radiology reviewers will qualitatively assess the hip joints on WBCT scan to determine whether there is evidence of AVN. Discrepancies in the reads will be adjudicated. Each hip will be assessed for presence or absence of AVN and, if present, the degree of change from the baseline scan and previous scan.

Trauma Leading to Flare-up

In MOVE, subjects can initiate flare-up dosing for substantial high-risk traumatic events likely to lead to a flare up. A listing of traumatic events and whether the trauma resulted in a confirmed flare-up will be generated.

2.1.5 Imaging Performance and Quality Metrics

The imaging Independent Review Charter (V7.1) describes the process for reading efficacy and safety imaging in the MOVE and NHS studies. It also describes an exercise comprised of secondary reviews, i.e., rereads of a subset of cases by the independent reviewers, of low-dose WBCT efficacy, radiograph and low-dose WBCT safety review types. The goal of the exercise is to characterize the intra-reader agreement and the inter-reader agreement for each of these review types.

Intra-Reader Agreement

The imaging vendor will randomly select sixteen study subjects for WBCT efficacy, eight study subjects for WBCT safety and four study subjects for radiograph read types, with a minimum of two timepoints each, for re-evaluation by the same reviewer at least 3 weeks after the initial radiology review. Discrepancies identified within the completed secondary review cases will be reviewed by the vendor as described below:

Discrepancies are defined as below:

- Low-dose WBCT Efficacy Review Any difference in post-TP1 scans as follows:
    - disagreement on the presence/absence of new HO in any region in the post-TP1 scan compared to the TP1 Consensus scan
    - [|diff change|/(mean (change)) >20% AND |diff change|>10000] OR |diff change|>25000, where diff change is the difference in the change in new HO compared to the TP1 consensus scan between secondary read and the original read and mean (change) is the average of the change in new HO volume.
- Radiograph Review:

  Any difference as follows:
    - any difference in the assessment of presence of epiphyseal abnormalities in either the knee and/or hand/wrist (i.e., No vs. Not Evaluable vs. Yes)
    - any difference in the assessment of any epiphyseal abnormality change from TP1 and/or from prior timepoint in either the knee and/or the hand/wrist (i.e., not present vs. no change vs. improved vs. worse vs. unknown)
    - any difference in selected epiphyseal abnormality(ies) per location
    - any difference in epiphyseal growth plate closure assessment (i.e., no (closed) vs. not evaluable vs. yes, partially closed vs. yes, open for any location (distal radius, distal ulna, proximal tibia, or distal femur, as applicable))
    - any difference in skeletal maturity assessment
- Low-dose WBCT Safety Review:

  Any difference as follows:
    - any difference in the assessment of presence of femoral head abnormalities (i.e., no vs. not evaluable vs. yes)
    - any difference in the presence of hand/wrist and/or knee epiphyseal abnormalities (i.e., no vs. not evaluable vs. yes)

- any difference in the assessment of any epiphyseal abnormality change from TP1 and/or from prior timepoint in either the knee and/or hand/wrist (i.e., not present vs. no change vs. improved vs. worse vs. unknown)

- any difference in selected abnormality(ies) per location.

Given the granularity of the discrepancy evaluation, discrepancies of 30% or more are expected to occur for each review type.

Inter-Reader Agreement

To determine inter-reader agreement, the radiologists participating in primary radiology reviews will read a subset of eight study subjects for WBCT efficacy, eight study subjects for WBCT safety and four study subjects for radiograph read types, with a minimum of two timepoints each, randomly selected by Parexel's image management and review system at least 3 weeks after the initial radiology review. Discrepancies identified within the completed secondary review cases will be reviewed by Parexel as described below. Discrepancies are defined in IRC Section 14.1.1 (Intra-Reader Agreement).

If a secondary review results in discrepancies for more than 30% of the timepoints for a reviewer for each review type, Parexel medical personnel will review the discrepancies to assess whether preventative actions as described in IRC Section 14.2 (Quality Monitoring) are required.

2.1.6 Pharmacokinetic Endpoints

Pharmacokinetics of palovarotene dosing will be assessed at the first 3-month safety assessment during chronic-based treatment; if samples cannot be obtained during the first 3-month safety assessment, then they can be obtained during any subsequent 3-month safety visit. Pharmacokinetics will also be assessed twice during flare-up dosing: once during the 20 mg regimen at any time between Study Days 4 to 28, and once during the 10 mg regimen at any time between Study Days 32 and 84, for the first treated flare-up only. If this is not possible for the first treated flare-up, pharmacokinetic blood samples can be obtained during any subsequent flare-up dosing. Pharmacokinetic blood samples will be collected at pre-dose and 3, 6, 10, and 24 hours post-dose.

The following parameters will be determined where possible by model independent analysis using WinNonlin™: $C_{max,ss}$, $C_{min,ss}$, $T_{max,ss}$, $AUC_{0\text{-}24ss}$, λ_z, $t_{½z}$, and CL/F.

The determination of palovarotene plasma concentrations will be performed using a validated LC-MS/MS method.

2.1.7 Compliance

Compliance is defined as (duration of treatment/expected duration of treatment). Calculation of duration of treatment is described in Section 2.4.8 and Section 2.4.9.

The expected duration of treatment is calculated as following:

- For non-flare-up treatment, expected duration of treatment is calculated as (last non-flare-up dose date – first non-flare-up dose date) – (sum of (all flare-up cycle dosing periods) + 1.
- For flare-up treatment, expected duration of treatment is calculated as total of expected flare-up cycle dosing periods across all flare-up cycles. Each expected flare-up cycle dosing period is defined as (last flare-up cycle dose date – first flare-up cycle dose date + 1).

Compliance will be summarized for the Principal FAS with descriptive statistics. Compliance in category (<80%, ≥80%) will also be summarized for counts and percentages of subjects.

2.2 Disposition of Subjects

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

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

Enrolled subjects include all subjects who sign the informed consent and meet all eligibility criteria of the MOVE Trial.

For subjects in the MOVE Trial, 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
- Enrolled subjects
- Enrolled and treated subjects including by whether the subjects crossed over from the NHS
- Subjects who permanently discontinued study drug by reason for discontinuing study drug;
- Subjects who completed the MOVE Trial
- Subjects who discontinued the MOVE Trial by reason for discontinuing study

For all categories of subjects (except for the screened category), percentages will be calculated using the number of enrolled subjects as the denominator. Reasons for study drug discontinuation will be presented in tables giving numbers and percentages. The NHS subject disposition will be summarized using the categories above, where applicable.

All critical or major deviations potentially impacting efficacy analyses and other major/critical deviations will be summarized in tables giving numbers and percentages of deviations by treatment group.

Additionally, the analysis populations will be summarized in a table by subject counts based on the enrolled populations (see definitions in Section 2.3).

2.3 Analysis Populations

The MOVE Trial subjects will be grouped into the following populations for analysis:

- The Principal Enrolled Population (Principal EP) includes all subjects with the R206H ACVR1 mutation who have not previously been treated with palovarotene and who sign the informed consent form and meet all eligibility criteria of the MOVE Trial. For the NHS, the Principal Enrolled Population will include all subjects enrolled.

- The Principal Full Analysis Set (Principal FAS) includes all enrolled subjects in the Principal EP who have a baseline HO volume measurement and at least one post-baseline HO volume measurement in the MOVE trial. For efficacy comparisons to the NHS, the Principal FAS will also include subjects enrolled in the NHS with available baseline and at least one post-baseline HO volume measurement.

- The Principal Per-Protocol Set (Principal PPS) is a subset of the Principal FAS including subjects with no major protocol deviations that are expected to interfere with assessments of the primary endpoint, and with at least 80% compliance to the study drug regimen. For efficacy comparisons to the NHS, the Principal PPS will also include subjects in the NHS with available baseline and at least one post-baseline HO volume measurement and with no major protocol deviation that are expected to interfere with assessments of the primary endpoint.

- The Principal Safety Set (Principal SS) includes all enrolled subjects with the R206H ACVR1 mutation receiving at least one dose of palovarotene in the current study. For safety comparisons to the NHS, the Principal SS will also include subjects enrolled in the NHS with available post-baseline follow-up.

- The Principal Pharmacokinetic Set (Principal PS) includes all enrolled subjects with the R206H ACVR1 mutation receiving at least one dose of palovarotene and providing evaluable pharmacokinetic data in the current study.

Subjects who do not have the R206H ACVR1 mutation or who have received previous treatment with palovarotene will comprise the Supplementary EP, the Supplementary FAS, the Supplementary SS, and the Supplementary PS, with these populations defined analogously as above for the subjects with the R206H ACVR1 mutation. (Due to the limited sample size, there is no Supplementary PPS.) The Principal populations will be used for all analyses described in this SAP. The Supplemental populations will be supportive and summarized separately and, unless otherwise specified, will not be used in hypothesis testing.

2.4 Statistical Methods

2.4.1 Demographic and Baseline Characteristics

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

Parameters will be summarized for the Principal FAS, PPS, and SS and the Supplementary FAS and SS. The respective NHS populations will also be presented, where applicable.

Parameters described in Section 2.1.1 will be summarized by treatment group and overall 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 MOVE analysis populations. P-values for demographic and baseline characteristics will not be calculated.

No specific description of the safety parameters will be provided at baseline. If relevant, the baseline values will be described along with each safety analysis.

2.4.2 Prior or New Onset Medication

The prior and new onset medications will be summarized based on the enrolled populations as follows:

- Prior medications (excluding systemic glucocorticosteroids) in the MOVE study discontinued prior to the first dose of study drug
- Prior medications (excluding corticosteroids) in the MOVE study ongoing at the first dose of study drug
- New onset medications in the MOVE study (excluding corticosteroids)
- Concomitant medications are any treatments received by the subject after first dose of study drug, and include prior ongoing medications and new onset medications received after first dose of study drug

The medications will be summarized according to the WHODrug Global by Anatomic Therapeutic Class (ATC) and preferred term in the MOVE study. The ATC Class 3 name will be presented if available, and if not available the ATC class 2 name will be presented.

Systemic gluococorticosteroids will be summarized analogously. New onset medications (excluding corticosteroids) and new onset corticosteroids will be summarized by whether the medication was started during chronic treatment or flare-up treatment.

In addition, the NHS only recorded usage of a specific list of medications (Table 6). A summary table presenting the new onset medications in both the MOVE and NHS studies will be presented. In that particular table, only medications that are listed in Table 6 will be included.

Table 6. NHS List of Medications

Prednisone	Montelukast	Narcotics
Other oral steroids	Cromolyn	Calcium
Celecoxib	Etidronate	Vitamin D
Ibuprofen	Pamidronate	Other
Indomethacin	Zoledronate	
Naproxen	Other bisphosphonates	
Other non-steroidal anti-inflammatory (NSAID) medications	Systemic retinoids	
Topical NSAIDs	Thalidomide	
Acetaminophen or similar	Antihistamine	

The tables for prior and new onset medications will be sorted by decreasing frequency of ATC followed by the generic names based on the overall incidence in the MOVE analysis populations. The table summarizing the medications listed in Table 6 will be sorted by decreasing frequency of the drug name listed in the MOVE analysis population.

2.4.3 Study Drug Exposure

The extent of study drug exposure will be assessed and summarized based on the safety populations.

The extent of study drug exposure is defined as: last study drug date – first study drug date + 1. This calculation is regardless of unplanned study drug interruptions.

The duration of chronic dosing, the number of subjects with at least one chronic dose reduction, the number of dose reductions by type (e.g., 5 to 2.5 mg), the number of subjects with interrupted study drug during chronic dosing and the number of subjects discontinuing study drug during chronic dosing will be summarized.

The number of subjects with flare-up dosing, the duration of flare-up dosing overall and by high or low flare-up-based dosing, the number of subjects with at least one flare-up dosing reduction by type, the number of subjects with interrupted study drug during flare-up dosing, the number of subjects discontinuing study drug during flare-up dosing will be summarized.

For chronic and flare-up dosing, a dose reduction is defined as sustained dosing under the expected dose, i.e., underdosing for at least 2 consecutive doses.

2.4.4 Analyses of Efficacy Endpoints

The efficacy endpoints will be analyzed using the Principal FAS. Analyses using the Principal PPS will be performed as noted. Listings will be generated for the Supplementary FAS.

Note that no statistical inference is planned to be included in the descriptive analysis to be conducted before the first formal interim efficacy analysis; the inference described in this section will be conducted for the three interim efficacy analyses and the final analysis.

In order to address the prolonged interruptions in dosing in subjects <14 years old upon US FDA's institution of a partial clinical hold due to premature physeal closure (PPC) on December 4, 2019 and interruptions in dosing for the remaining subjects on January 24, 2020 due to the crossing of the futility boundary at the second interim analysis, analyses of all efficacy endpoints will include assessments collected on or before these interruptions, unless otherwise specified. (Assessments were to continue as planned during the interruptions.) Analyses in which all timepoints are included, irrespective of dosing interruption, will also be conducted for the primary analysis (and additional analysis without the square-root transformation) and secondary analyses to aid in the interpretation of the efficacy results.

The sponsor also notes that because the futility boundary was crossed at the second interim analysis, all subsequent analyses of efficacy data will be supplied without applying thresholds for success or futility, as these thresholds are no longer relevant.

Analyses of Primary Efficacy Endpoint

Primary Analysis

The primary efficacy endpoint is the annualized change in new HO volume (as assessed by WBCT). The change in new HO volume is calculated by summing the increase in HO volume across all body regions for which new HO has occurred where the increase in HO volume per region is defined as the square-root of the volumetric increase in that region. The square-root transformation is used to reduce the influence of outliers. Transformation of volumetric changes has been used in related contexts. For example, a log transformation has been used to model volumetric increases in the kidneys of subjects with autosomal dominant polycystic kidney disease[1,2]. The change in new HO volume is modeled using a Bayesian compound Poisson distribution. The Bayesian compound Poisson distribution assumes that the change in new HO volume can be modeled as a compound distribution of the number of body regions with new HO, K, and the new HO volume per region where new HO has occurred, Z. The number of body regions with new HO in subject i for WBCT scan j with duration w_{ij} (the time between scan j and the previous scan) is distributed as

$$K_{ij} \sim Pois\left(\lambda_{i,j} * w_{ij} * \theta_{1,t(ij)}\right);$$

$$\log(\lambda_{i,j}) = \log(\lambda_i) + \beta_1 X_{1,i} + \beta_2 X_{2,i,j}$$

The subject-level rate λ_i follows a gamma distribution and accounts for potential correlation in measurements from the same subject; $t(ij)$ is an indicator function equal to 1 if the subject was

on treatment at the time of the jth WBCT scan and 0 if the subject was not on treatment. Therefore, the marginal distribution of the number of body regions with new HO follows a negative binomial distribution, which requires much less restrictive assumptions than a Poisson distribution and does not tend to result in an underestimate of the variance.

Letting $\theta_{1,0} = 1$, the variable $\theta_{1,1}$ is the multiplicative effect of palovarotene treatment on the rate of body regions with new HO. Covariates are included in the analysis of the number of body regions with new HO to adjust for potential explained differences in the rate of new HO based on the subject's sex, $X_{1,i}$, and age at time of scan, $X_{2,i,j}$. Age at the time of scan is represented as a factor variable with 0 for <18 years old and 1 for ≥18 years old. Covariate effects $\exp(\beta_1)$ and $\exp(\beta_2)$ are multiplicative effects of sex and age on the rate of body regions with new HO.

The new HO volume (square-root of the volumetric increase in that region) in region r, $\sqrt{Z_{ijr}}$, where new HO has occurred for subject i in scan j is assumed to be distributed as:

$$\sqrt{Z_{ijr}}/1000 \sim N\left(\alpha * \alpha_r * \theta_{2,t(ij)}, \frac{\alpha_r^2}{\tau_{t(ij)}}\right).$$

The scale of new HO volume is modeled in the thousands. Letting $\theta_{2,0} = 1$, the variable $\theta_{2,1}$ is the multiplicative effect of palovarotene treatment on the new HO volume conditional on new HO occurring. The α_r are region-specific variables that contribute to the mean of new HO and the variance;the restrictions α_r=$\alpha_{r'}$ where r and r' are a left and right region pair are included, i.e., right chest and left chest, right arm and left arm, right hip and left hip, and right lower leg and left lower leg. The precision variables, τ_0 and τ_1, introduce flexibility by allowing variability to differ between new HO volume in treated subjects and untreated subjects.

The prior distributions for the variables in the Bayesian compound Poisson distribution are the following:

$$\begin{array}{ccc} \lambda_i & \sim & Gamma(a, \text{scale} = b) \\ \beta_1, \beta_2 & \sim & N(0, 2^2) \\ a, b & \sim & Gamma(1,1) \\ \alpha & \sim & N(0, 1) \\ \alpha_r & \sim & Unif(0, 4) \\ \tau_0, \tau_1 & \sim & Gamma(1, 0.01) \\ \theta_{1,1}, \theta_{2,1} & \sim & Unif(0, 2). \end{array}$$

All gamma distributions are parameterized as the shape and rate, except as noted for λ_i. The primary efficacy analysis comparing the annualized change in new HO volume between subjects treated with palovarotene and untreated subjects is performed by calculating the ratio of the annual mean change in HO volume in palovarotene treated subjects to untreated subjects using the Principal FAS and assuming missing at random.

In this SAP, new HO is deemed to be present in a body region if the volume of net new HO is greater than zero, i.e., the change in volume of HO in a region between two timepoints is

positive. (Volumes of individual lesions are not assessed; the net volume of HO in a body region is reported.). Any regions with amounts of new HO<0 mm^3, i.e., reductions in the volume of HO over time (which can happen due to bone remodeling or measurement variability), will be represented as having no new HO and the volumes will be set to zero in the analysis.

Missing timepoints for subjects, i.e., assessments not conducted, assessments conducted but not evaluable for the presence of HO or volume of HO, or assessments for which HO volume for at least 50% of body regions is not evaluable or not available, will not be imputed. Body regions with non-evaluable HO volume at a timepoint will be represented as having no new HO and the volumes will be set to zero in the analysis.

As described in the MOVE and NHS Independent Review Charter, adjudication is by timepoint. When adjudicated, analysis will use the adjudication assessment of volume of new HO. When not adjudicated, the average of the two readers' amounts of HO for each body region will be used, assigning a volume of 0 mm^3 if the reader does not measure due to the qualitative assessment of no new HO.

Using the Bayesian compound Poisson model described above, the efficacy ratio γ is calculated as the treatment effect on the mean number of body regions with new HO, $\theta_{1,1}$, multiplied by the treatment effect on new HO volume conditional on new HO occurring, $\theta_{2,1}$, expressed as $\gamma = \theta_{1,1} * \theta_{2,1}$. Random samples generated via Gibbs sampling from the posterior distribution of $\theta_{1,1}$ and $\theta_{2,1}$ will be used to compute the posterior probability that $\gamma < 1$ to determine statistical significance.

Additional Analyses

The primary efficacy analysis described above will be repeated without the square-root transformation. As noted by the DMC after the second interim analysis, using the square-root transformation of the data in the primary analysis was impactful in that it appeared to have moved the statistical conclusion from significant therapeutic benefit to showing futility of the treatment.

Furthermore, as is explained later in this section, the square-root transformation together with the more frequent WBCT efficacy assessments in MOVE relative to NHS seem to introduce a bias against palovarotene in the primary analysis.

Also, with its square-root transformation reducing the influence of subjects with unusually large amounts of new HO, the primary analysis is expected to be less powerful relative to the analysis without the square-root transformation if an effect of palovarotene is to reduce the chance of catastrophic HO as is suggested in the results observed thus far in MOVE (and was also observed in post-hoc analyses of flare-up-based imaging endpoints in earlier studies).

For these reasons, while acknowledging that it is post-hoc, the sponsor believes that the Bayesian compound Poisson analysis without the square-root transformation is the more appropriate analysis of the primary endpoint, relative to the Bayesian compound Poisson analysis with the square-root transformation. The Bayesian analysis without square root transformation is referred to as the analysis of primary interest.

When performed without the square-root transformation, the new HO volume in region r, Z_{ijr}, where new HO has occurred for subject i in scan j is assumed to be distributed as:

$$Z_{ijr}/1000 \sim N\left(\alpha * \alpha_r * \theta_{2,t(ij)}, \frac{\alpha_r^2}{\tau_{t(ij)}}\right).$$

The prior distributions for the variables in this Bayesian compound Poisson distribution are the following:

$$\begin{array}{ccc} \lambda_i & \sim & Gamma(a, \text{scale} = b) \\ \beta_1, \beta_2 & \sim & N(0, 2^2) \\ a, b & \sim & Gamma(1,1) \\ \alpha & \sim & N(0, 10^2) \\ \alpha_r & \sim & Unif(0, 4) \\ \tau_0, \tau_1 & \sim & Gamma(1,100) \\ \theta_{1,1}, \theta_{2,1} & \sim & Unif(0, 2). \end{array}$$

The Bayesian compound Poisson model will be fitted using the R statistical computing language and environment. The R programs will be run using v3.5.0 or later; packages used, including version number, will be identified in the CSR.

Additional supportive analyses including the weighted mixed linear effect (wMLE) analysis with and without the square-root transformation will also be performed. The wLME analysis without the square-root transformation was previously the primary efficacy analysis, before the Bayesian compound Poisson analysis was added in Amendment 1 of the protocol. This wLME analysis without square-root transformation is referred to as the alternative analysis of primary interest.

A subject-level random effect will be used to account for the correlation among repeated measures on the same subject as subjects may contribute follow-up from both the NHS and the current study. Baseline HO volume divided by age will be the only covariate included in the model, in addition to the factor identifying study of origin. The LME model will be fitted using only a subject's observations associated with the longest follow-up in the NHS and the current study with weights used to account for the different lengths of observed subject follow-up. Missing at random will be assumed for the primary analysis and no imputation of missing data performed.

The wLME analyses will be run using SAS 9.4 or higher. The code will use the following structure:

```
proc mixed data=data;
                class treat usubjid;
                model annrate newho = treat basetot_bage /s;
                random usubjid;
                lsmeans treat/pdiff=control('0') cl;
                weight fup;
run;
```

In this code, treat is a binary indicator for study (1=MOVE, 0=NHS); usubjid is a unique identifier of subjects (used across all FOP studies); annrate_newho is the annualized rate of new HO; and fup is follow-up in years.

The Bayesian compound Poisson analysis and wLME analysis without square-root transformation and including all assessments on or before December 4, 2019, for subjects <14 years old on the date of their assessment and assessments on or before January 24, 2020, for all other subjects, will be repeated using the Principal PPS after the third interim analysis.

Descriptive statistics summarizing the total HO by study visit, change in HO volume by study visit, and change in HO volume from baseline by study visit will be provided. Tables summarizing annualized new HO will also be produced. Summaries will also be provided by body region. Summaries of new HO coincident with flare-ups and not coincident with flare-ups will also be generated. Summaries of annualized mean New HO, conditional on new HO volume>0 mm3, will be created. HO listings will also be generated.

Missingness, including frequency of WBCT not collected and regions that are not evaluable, will be summarized. Frequency of negative new HO, i.e., apparent reductions in HO overall and by body region (defined below), will also be summarized.

Handling of Reductions in HO

Please note that, as referenced above, the Bayesian compound Poisson models with and without square-root transformation assume non-negative changes in HO volume for each subject for each region, and any apparent reductions in HO volume over time (i.e., negative new HO volumes) are "zeroed out", i.e., replaced with zero. (All analyses of square-root transformed new HO volumes require non-negative new HO volume in order for the square-root transformation to be applied.) In contrast, because wLME is a regression method, it is less specialized and can accommodate the negative values when applied to non-transformed data.

However, in order to understand the relative contributions to the results of 1) zeroing out the negative changes; 2) square-root transformation of the zeroed out new HO volumes; 3) using a body-region or overall whole body based analysis; 4) using data on intervals between WBCTs, not just the first and last timepoint; 5) Bayesian compound Poisson analysis; we will run five wLME "series":

1. Raw, untransformed data (i.e., including all negative values)
2. Zero out overall whole-body new HO without square-root transformation
3. Zero out overall whole-body new HO with square-root transformation
4. Zero out by region [by visit] without square-root transformation
5. Zero out by region [by visit] with square-root transformation

For series 2 and 3, the sum of new HO at the last timepoint, i.e., changes from baseline including negatives, is added across body regions. If negative, the sum is replaced with zero. For series 3, the sum is square-root transformed. Then the sum – or square-root of the sum – is divided by time in years, resulting in the annualized new HO.

For series 4 and 5, negative changes in HO volume by region (from previous visit) are replaced with 0. For series 5, the square-root of each change is taken. These numbers are then added across timepoints and body regions and their sums are divided by time in years, resulting in annualized new HO. This method for constructing annualized new HO, with and without square-root transformation, yields wLME analyses that are the closest analogs to the Bayesian compound Poisson analyses.

Note that in applying the square-root transformation to zeroed out new HO for intervals, wLME using series 5 should be subject to the same bias as is attributed to the Bayesian compound Poisson primary analysis illustrated later in this section.

Control of Type I Error

Ninety-nine palovarotene-naïve subjects with the R206H ACVR1 mutation were enrolled into the current trial and are to be followed for 2 years. There were to be three early interim efficacy analyses and one final analysis. Futility was declared at the second interim analysis, but the study continued. The final analysis for Part A will occur when all subjects have completed 2-year follow-up.

The first interim analysis was to take place when approximately 35 MOVE subjects complete 1 year of follow-up with the second and third interim analyses occurring when all subjects in the Principal EP have the opportunity to have complete (i.e., have WBCT data) 12 months and then 18 months of follow-up, respectively. It was assumed that the full NHS dataset will be available at the first interim analysis. The amount of statistical information at each interim analysis can be approximated by the sum of the follow-up in each study divided by the total expected. The total follow-up expected across the NHS and the MOVE Trial is 317.5 years (NHS: 45*1.5 years + 45*2.0 years; MOVE: 80*2.0 years). Assuming eight subjects enroll in the MOVE study per month, the percentage of subject follow-up for each analysis is summarized in Table 7 where values with an asterisk indicate approximate information based on expected accrual rates and total planned accrual in the Principal EP.

Table 7. Observed MOVE Follow-up at Each Planned Analysis

Analysis	Observed MOVE follow-up	Percentage of follow-up available
Interim #1	35 subjects with 12 months *45 subjects with 6 months	66%
Interim #2	*80 enrolled subjects with 12 months *35 subjects with 18 months	80%
Interim #3	*80 enrolled subjects with 18 months *35 subjects with 24 months	92%
Final	80 enrolled subjects with 24 months	100%

Using these percentages, the alpha level threshold used to determine treatment effect significance at each analysis was derived using the Lan-DeMets alpha-spending function with O'Brien-Fleming parameterization and assuming a one-sided, overall type I error rate of 2.5%.

The one-sided significance thresholds are 0.0058, 0.0103, 0.0156, and 0.0190 for the first, second and third interim analyses and the final analysis, respectively, in the scenario above.

The one-sided significance thresholds for the Lan-DeMets alpha spending function with an O'Brien-Fleming boundary will be recalculated at the time of each interim analysis based on the estimate of the information fraction. The estimate of the information fraction is a fraction in which the numerator is the sum of the time between the baseline WBCT assessments and the last timepoint for each subject for NHS and MOVE assessments; the denominator is the sum of the time between the baseline WBCT assessments and the expected last timepoint (at the time of the final analysis) for each subject for NHS and MOVE assessments.

At the final analysis, the information fractions at the previously conducted interim analyses will be re-calculated using the observed information, i.e., the sum of the time between the baseline WBCT assessments and the last timepoint for each subject for NHS and MOVE assessments. The critical p-value for the final analysis will be set such that the remaining alpha is spent. The information fraction at the final analysis will be set to 100%.

Study success was to be declared if the posterior probability that palovarotene reduces the change in new HO volume is greater than 1 minus the one-sided significance threshold. For example, the posterior probability that palovarotene reduces the annualized change in new HO volume must be greater than 0.9810 to declare study success at the final analysis in the scenario above. Because the futility boundary was crossed at the second interim analysis, i.e., the posterior probability that $\gamma < 0.7$ (at least a 30% reduction in annualized new HO volume on the square-root scale) was less than 5%, all subsequent analyses of efficacy data will be supplied without applying thresholds for success or futility, as these thresholds are no longer relevant.

Sensitivity Analyses

The following sensitivity analyses will be conducted after the third interim analysis to examine the robustness of the Bayesian compound Poisson analysis without the square-root transformation, unless otherwise specified. All assessments on or before December 4, 2019, for subjects <14 years old on the date of their assessment and assessments on or before January 24, 2020, for all other subjects, will be used. Note that, given the consistency of results between the Bayesian compound Poisson and wLME analyses observed at the second and third interim analyses, the sensitivity analyses below will utilize both methodologies, as described. The results may be summarized jointly in forest plots, as appropriate.

To assess impact of differences between NHS and MOVE:

- Adjustment for additional baseline covariates: This sensitivity analysis will extend the primary efficacy analysis to include additional baseline covariates. These will include baseline HO volume divided by age, age, sex, months since last flare-up, and CAJIS. The palovarotene treatment effect will be estimated using the Bayesian compound Poisson distribution with all covariates included. A similar wLME will also be performed.
- Propensity score analysis: The propensity score, defined as the conditional probability of being treated given the baseline covariates, can be used to balance the baseline covariates

across treatment groups.[3] Therefore, it can reduce potential bias due to imbalances in the distributions of demographic and disease characteristics between the NHS and MOVE Trial. The propensity scores will be derived from a logistic regression model with the following covariates: baseline HO volume divided by age, age, sex, months since last flare-up, and CAJIS. Each subject's propensity score will be estimated from the logistic regression model, and subjects will be grouped by quartiles of their propensity scores. Propensity score quartile will be used as the only covariate in the Bayesian compound Poisson analysis. wLME will be conducted for each quartile and overall using the 5-level factor as a covariate (in addition to treatment).

- Restriction to subjects contributing follow-up for both the NHS and MOVE Trial: This analysis will use the wLME analysis method on the subset of the MOVE subjects who transitioned from the NHS and have annualized rates of new HO for both studies. Baseline HO volume divided by age will be the only covariate included in the model, in addition to the factor identifying the study of origin.

To assess impact of differences in length of follow-up between NHS and MOVE:

- Analysis restricting subjects to specific durations of follow-up: The WBCT dataset will be restricted to observations associated up to 1-year – interpreted as assessments within the first 15 months of study participation (i.e., midway between Month 12 and Month 18 for MOVE) – and, separately, with up to 2-years of follow-up – interpreted as assessments within the first 27 months of study participation. The analysis method will be wLME.

To assess impact of missing data in MOVE:

- Missing data sensitivity analysis using a tipping point approach: Subjects who discontinue the MOVE study will have volume of new HO imputed using the wLME-modeled treatment effect. The multiple imputations will be scaled to induce 0%, 25%, 50%, 75%, and 100% reductions in the palovarotene treatment effect. The 0% corresponds to subjects who discontinue the MOVE Trial having the same distribution for number of body regions with new HO and change in HO volume per region as the MOVE subjects who did not discontinue; and 100% corresponds to subjects who discontinued the MOVE Trial having the same distributions as the NHS subjects. Missing data will be imputed for the visits in MOVE that were either missed or would have been missed due to study discontinuation prior to December 4, 2019 / January 24, 2020. Analyses in which subjects who discontinue the NHS have volume of new HO imputed may also be considered, while noting that the majority of subjects who discontinued NHS did so to enroll in an interventional study, e.g., MOVE or less frequently, Study PVO-1A-201.

To assess robustness of efficacy signal against choice of wLME as post-hoc analysis:

- Analysis using generalized estimating equations (GEE): A GEE model will be fit to the annualized change in new HO and the empirical variance estimator will be used for inference. Weights will be used to account for the different lengths of observed subject follow-up. Baseline HO volume divided by age will be included in the model as a

covariate and a compound-symmetric working correlation matrix will be used. The GEE model takes a different approach to account for within-subject correlation compared to the primary analysis.

To assess consistency of efficacy signal across subgroups:

- Subgroup analyses: The subgroup analyses will also be performed for all efficacy endpoints to examine the consistency of the estimated treatment effect by sex (Male, Female), Asian ancestry (Yes, No), and age (≥12, ≥14, ≥18, ≥12 – <17 years) and age by sex (<8 years for Females or <10 years for Males; ≥8 years for Females or ≥10 years for Males; ≥8 – <14 for Females or ≥10 – <14 years for Males).

  The inferential analysis method used in the subgroup analyses will be Bayesian compound Poisson analysis and wLME for new HO volume by age and age by sex, and for wLME for all other endpoints and factors, with the exception of the above contingency table analysis of proportions of subjects with new HO and flare-ups; this analysis will employ Fisher's exact test.

To assess impact of the difference in WBCT visit schedules between MOVE and NHS:

- The impact of the difference in WBCT visit schedules between MOVE and NHS is assessed by analyzing the data as if the WBCT assessments in MOVE were conducted annually rather than biannually. This is intended to investigate whether the differences in visit schedules introduced bias, with and without use of the square-root transformation. Estimates of $\gamma, \theta_{1,1}, \theta_{2,1}$ will be compared between Bayesian compound Poisson analyses, with and without square-root transformation, with the observed biannual MOVE WBCT and with the dataset collapsed to simulate annual MOVE WBCT assessments.

  It must be noted, however, that this analysis does not eliminate the effect of the more frequent assessments in MOVE; as described in the MOVE and NHS Independent Review Charter, radiologists only measured the amount of HO if they first determined there was new HO relative to the previous timepoint. If there was no new HO, the volume from the previous assessment was carried forward. Because the read process for the 12-month timepoint used the 6-month assessment and the read process for the 24-month assessment used the 18-month assessment, this analysis can investigate only the bias introduced by the differences in visit schedules into the analysis, not the differences in visit schedules per se.

  In any case, the sponsor believes that the differences in visit schedules could result in biases of the event rate treatment effect $\theta_{1,1}$ against palovarotene and growth rate treatment effect $\theta_{2,1}$ in favor of palovarotene but that, in the absence of the square-root transformation, these biases are offset, yielding an unbiased estimate of $\gamma = \theta_{1,1} * \theta_{2,1}$.

  With the square-root transformation, however, the bias against palovarotene in the event rate treatment effect $\theta_{1,1}$ is unchanged, while the bias on growth rate treatment effect $\theta_{2,1}$ in favor of palovarotene is reduced, yielding a bias in γ against palovarotene.

Consider identical growth events in NHS and MOVE of 8000 mm^3 that is split across two six-month intervals in the MOVE study as 4000 mm^3 in the first 6-month interval and 4000 mm^3 in the next 6-month interval. This registers as two events in one year on MOVE, but just one event in NHS. Similarly, the growth rate, conditional on there being an event, is 2X higher in NHS than MOVE, i.e., 8000/4000=2, in the absence of a square-root transformation. With a square-root transformation, however, the growth rate per event is only 1.4X higher in NHS than MOVE, i.e., $\sqrt{8000}/\sqrt{4000} \sim 1.41$.

A similar comparison of analyses will be performed with wLME, where negative values are zeroed out by region and then square-root transformed, because the bias illustrated above from the square-root transformation should also affect this annualized HO endpoint, which is constructed using new HO observed at each visit.

Analyses of Secondary Efficacy Endpoints

Proportion of Subjects with Any New HO (Key Secondary Endpoint) and Number of Body Regions with New HO

The proportion of subjects with any net new HO (at Month 12) – and number of body regions with any net new HO (at Month 12) – will be compared between the MOVE and NHS groups using the Bayesian compound Poisson model used in the primary efficacy analysis. (New HO is deemed to be present in a body region qualitatively if the volume of net new HO is greater than zero, i.e., the change in volume of HO in a region between two timepoints is positive when new HO is identified qualititatively) Gibbs sampling from the posterior distribution will be generated to compute the posterior distribution of $\theta_{1,1}$, the multiplicative effect of palovarotene treatment on the rate of body regions with new HO. Because, within the compound Poisson model, $\theta_{1,1}<1$ implies both a lower proportion of subjects with any new HO and a smaller number of body regions with new HO for palovarotene-treated subjects, the inference is identical for these two secondary endpoints. The probability of this ratio being less than 1 must be greater than 0.975 to declare statistical significance.

Inference on $\theta_{1,1}$ from the sensitivity analysis analyzing the data as if the WBCT assessments in MOVE were conducted annually rather than biannually will also be relevant, given the concerns about bias in estimating $\theta_{1,1}$ even in the absence of square root transformation will be summarized.

Tables will be provided summarizing the number and percentage of subjects with any net new HO by study visit and overall. In these tables, Month 12 incidence will be compared using an exact test for the difference in proportions.

Tables will also be provided summarizing the number of body regions with new HO by study visit and overall. The body regions with new HO will also be summarized by study visit and overall. In these tables, Month 12 counts will be compared using a negative binomial regression.

Flare-up Rate Per Subject-Month Exposure

The flare-up rate per subject-month exposure will be compared using a negative binomial regression on the number of flare-ups. 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.

In the NHS study, events were considered flare-ups when two or more flare-up signs were reported. This is different from the MOVE study, which defines a flare-up as an event with one or more flare-up sign. The analysis of the proportion of subjects reporting flare-ups will restrict the MOVE flare-ups to those events that match the NHS study definition.

Descriptive tables will be provided summarizing the number of flare-ups and the rate of flare-ups by study visit, by age group and overall.

Proportion of Subjects Reporting Flare-ups

The proportion of subjects reporting flare-ups will be compared between untreated and treated subjects using an exact test for the difference in proportions.

As with flare-up rate analysis, this analysis will be performed restricting the MOVE flare-ups to those events that match the NHS study definition.

Descriptive tables will be provided summarizing the number and percentage of subjects with any flare-up by study visit, by age group, and overall.

Analyses of Exploratory Endpoints

The exploratory efficacy endpoints, including change from baseline in CAJIS Total score, CAJIS Upper Extremities subscore and CAJIS Mobility subscore; FOP-PFQ total score; and PROMIS Global Physical Health and Global Mental Health T-scores; will be summarized by visit using descriptive statistics for the Principal FAS and presented for the Supplementary FAS in listings.

Additional exploratory endpoints of incidence and volume of "catastrophic HO" will also be described and compared between MOVE and NHS for the Principal FAS. It has been observed – after unblinding at the second interim analysis – that a minority of subjects develop large amounts of HO in a year; the incidence of subjects with at least 50,000 mm^3 and 30,000 mm^3 of new HO across all body regions in each year of study will be summarized.

2.4.5 Analysis of Safety Endpoints

Safety analyses will be provided by chronic, flare-up dosing and overall, unless otherwise specified.

General Rules

All safety analyses will be performed on the safety populations defined in Section 2.3, unless otherwise specified, using the following rules:

- The baseline value when summarizing overall visits is defined as the last available value prior to the first chronic dose of study drug.

- The baseline value when summarizing over flare-up visits only is defined as the last available value prior to or on the date of first dose of flare-up treatment within the particular cycle. For the NHS study, the baseline value when summarizing over flare-up visits is defined as the value on the Day 1 visit of the corresponding flare-up.

- The potentially clinically significant (PCS) values for pre specified lab, vital sign and ECG variables are values according to predefined criteria/thresholds (see Appendix 6.2).

- For AEs, 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 AE is considered treatment emergent and assigned to the chronic dosing period.

    - 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 AE is considered treatment emergent and assigned to the chronic dosing period. 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 AE is not considered treatment emergent.

    - If the day of the start date is missing and the start month and year is within a period of time within a flare-up treatment, the start date is imputed as the date of the flare-up treatment in the same month as the AE, and the AE is considered treatment emergent and assigned to the flare-up dosing period. Otherwise, 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 AE, whichever is later, and the AE is considered treatment emergent and assigned to the chronic dosing period. 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 AE, and the AE is not treatment emergent.

    - Partial AE end dates will be imputed as the last day of the month or year for missing end dates. Completely missing end dates will be imputed as the date of last known treatment exposure for the subject.

- 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 last assessment for the subject, and the concomitant medication would be considered both prior and concomitant and assigned to both the chronic dosing period and flare-up period (if applicable to the subject).

- 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). Prior Medications would then be defined as any medications that started prior to the date of first dose. Concomitant medications for chronic dosing would be defined as medications taken at any point in time during the chronic treatment period. Concomitant medications for flare-up dosing would be defined as medications taken at any point in time during any flare up cycle, per imputed or non-imputed dates.

- Lab data identified as having a "Flare X" visit name will be assigned to a flare-up cycle based on the date of the lab data and the visit date window as in the following algorithm:
  - If an observation has a designation of "Flare X" for the visit, then the date of the observation will be compared to the dates of dosing for flare-up cycles and weeks.
  - If the observation falls within 14 days of a known flare-up dose cycle and week date, then it is assigned to that particular flare-up cycle and week.

Summaries of WBCT, C-SSRS (suicide ideation/behavoir), laboratory data, vital signs, linear/knee height, weight/BMI , electrocardiograms, and prior/concomitant medications will include the NHS data, where applicable.

Analyses of Adverse Events

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

Adverse event incidence tables by SOC and PT will be sorted by decreasing frequency incidence of SOC and by decreasing incidence of PT within SOC. The number (n) and percentage (%) of subjects experiencing an AE will be presented by treatment (chronic dosing, flare-up dosing and total palovarotene [overall]) . Multiple occurrences of the same event in the same subject will be counted only once. The denominator for computation of percentages is the safety population for the chronic dosing and overall group and the number of subjects in the safety population with flare-up dosing for the flare-up dosing group. TEAEs by Pediatric Subjects (<18 years) and Adult Subjects (>=18 years) are summarized

The following TEAE summaries will be generated for the safety populations:

Analysis of Treatment-Emergent Adverse Events

- Overview of TEAEs, summarizing number (%) of subjects with
  - TEAEs
  - TEAEs possibly related to treatment
  - TEAEs by maximal severity
  - Treatment-emergent SAEs
  - Treatment-emergent SAEs possibly related to treatment
  - TEAEs leading to dose modification or interruption of study drug
  - TEAEs leading to permanent study drug and study discontinuation

  - TEAEs leading to death

- All TEAEs by primary SOC and PT, showing number (%) of subjects with at least one TEAE
- All TEAEs with incidence ≥ 5% in the overall group by primary SOC and PT.
- All TEAEs by maximal 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)
- All TEAEs leading to Dose Modification of Study Drug
- All TEAEs leading to Dose Interruption
- All TEAEs leading to Permanent Discontinuation of Study Drug
- All TEAEs leading to Study Discontinuation

Analyses of Pre-Treatment AEs

- All pre-treatment AEs by primary SOC and PT, showing number (%) of subjects with at least one pre-treatment AE

Analyses of Post-Treatment AEs

- All post-treatment AEs by primary SOC and PT, showing number (%) of subjects with at least one post-treatment AE
  - Post-treatment SAEs
  - Post-treatment SAEs possibly related to treatment
  - Post-treatment AEs by maximal severity
  - Post-treatment AEs leading to permanent study drug and study discontinuation
  - Post-treatment AEs leading to death

Listing will be provided for all pre-treatment AEs, all TEAEs, all SAEs, all TEAEs leading to dose modification or interruption of study drug, all TEAEs leading to permanent study drug discontinuation, all deaths, and all post-treatment AE categories noted above.

Analyses of Laboratory Variables

The summary statistics of all laboratory variables, including raw values and changes from baseline, will be calculated for each visit. Central lab values will be summarized; if local labs exist but no central labs are available, then local labs will be summarized. Summary tables will be provided for all treatment visits, both chronic and flare-up visits. For these tables, the baseline

value for the MOVE trial is the last available value prior to or on the date of first dose of chronic treatment. For the NHS study, the baseline value is the value taken at the Day 1 visit.

Summary tables will also be provided for flare-up treatment visits only using date of flare-up dosing to derive change from baseline. For flare-up treatment, subjects may contribute more than one observation for summaries due to multiple cycles of flare-up treatment. Each cycle of flare-up treatment will be summarized separately. For these tables, the baseline value for the MOVE trial is the last available value prior to or on the date of first dose of flare-up treatment within the cycle. For the NHS study, the baseline value is taken at Day 1 of each individual flare-up. In addition, for all laboratory parameters, the mean and mean change from baseline at all chronic visits and all flare-up cycle safety visits will be plotted separately.

The incidence of potentially clinically significant (PCSs) at any time (see Appendix 6.2 for the criteria defining potentially clinically significant (PCS) for clinical laboratory parameters) during the treatment period will be summarized. One summary table will be presented for all chronic visits (non-flare-up visits in the NHS) and for flare-up treatment visits. For derivation purposes for chronic visits, the baseline value for the MOVE trial is the last available value prior to or on the date of first dose of chronic treatment. For the NHS study, the baseline value is the value taken on the Day 1 visit. For flare-up summaries, the baseline value for the MOVE trial is the last available value prior to or on the date of first dose of flare-up treatment with the cycle. For the NHS study, the baseline value is the value taken on the Day 1 visit of the corresponding cycle.

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. Pre-treatment was defined as the last value prior to initiation of chronic treatment.For flare-up treatment, PCS will be summarized at the subject-level and the flare-up level. Flare-up analyses may include subjects multiple times as a subject may have more than one flare-up treatment cycle within the study, whereas the subject-level analyses will include each subject once, summarizing at least one post-baseline PCS abnormality across all flare-up cycles.

All PCS values will be presented in a subject-based separate listing as well including baseline values, PCS values and follow-up values.

Pregnancy test results for females of child-bearing potential were provided as subject listings.

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.

Absolute QTc interval prolongation QTcF, and QTcB will be summarized at baseline and post-baseline, using counts of subjects and the following criteria:

- QTc interval >450 msec
- QTc interval >480 msec
- QTc interval >500 msec

- QTc interval increases from baseline >30 msec
- QTc interval increases from baseline >60 msec

The incidence of ECG abnormalities at any time (see Section 2.1.4 for the abnormality list) during the treatment period will be summarized by visit and category (rhythm, conduction, morphology/chamber enlargement, axis deviation, myocardial infarction, ST Segment/T waves/U waves).

New onset ECG PCS values will be summarized for chronic treatment visits. Subjects with pre-existing ECG PCS values at baseline will not be considered to have new onset values after start of study drug. PCS will be summarized at the subject-level. For flare-up treatment, PCS will be summarized at the subject-level. The PCS list is provided in Appendix 6.2.

All PCS values will be presented in a subject-based separate listing as well including baseline values, PCS values and follow-up values.

All ECG abnormalities (rhythm, conduction, morphology/chamber enlargement, axis deviation, myocardial infarction, ST Segment/T waves/U waves), with specific abnormality will be provided in a by subject listing.

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. Summary tables will be provided for all treatment visits, both chronic and flare-up visits. For these tables, the baseline value for the MOVE trial is the last available value prior to or on the date of first dose of chronic treatment. For the NHS study, the baseline value is the value taken at the Day 1 visit. Summary tables will also be provided for flare-up treatment visits only using date of flare-up dosing to derive change from baseline. For flare-up treatment, subjects may contribute more than one observation for summaries due to multiple cycles of flare-up treatment. Each cycle of flare-up treatment will be summarized separately. For these tables, the baseline value for the MOVE trial is the last available value prior to or on the date of first dose of flare-up treatment within the cycle. For the NHS study, the baseline value is taken at Day 1 of each individual flare-up.

The incidence of PCSs at any time (see Appendix 6.2 for the PCS list) during the treatment period will be summarized in the same manner as lab PCS tables.

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

Analyses of C-SSRS Variables

The subjects who report any type of suicidal ideations in the Columbia-Suicide Severity Rating Scale (C-SSRS) or any suicidal behavior during the study will be presented in listings.

A 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.

Analyses of Knee and Hand/Wrist Radiographs and WBCTs for Assessment of Bilateral Hand/Wrist and Knee Epiphyseal Growth Plate

The summary statistics of bone age and the difference between bone age and chronological age (ratio of change from baseline in bone age to change from baseline in chronological age) will be calculated for each visit.

Figures will include a by-subject spaghetti plot of the bone age by chronological age, and a summary figure by sex and age group (4-10, 11-18) with femur length and tibia length on the y-axis and visit on the x-axis. Similar figures will be produced using chronological age in place of bone age.

Summary tables of growth plate abnormalities identified in radiographs (MOVE Trial only) will include the following:

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

Summary tables of growth plate abnormalities identified in WBCT safety reads (MOVE Trial and NHS) will include the following:

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

Listings of all subjects with premature growth plate closure will be provided.

Analyses of Standardized Stadiometry and Knee Height for Assessments of Linear Growth

The summary statistics of linear height for age z-score and knee height (values and changes from baseline) will be calculated for each visit. Figures will include height z-scores over time and knee height over time. In addition, summary statistics of weight for age z-score and BMI (kg/m^2) for age z-score (values and changes from baseline) will be calculated for each visit and plotted against age.

The z-scores will be calculated by using US Centers for Disease Control and Prevention (CDC) growth charts, using the SAS macro from http://www.who.int/childgrowth/software/en/. (Note that the derivation of the z-scores depends on sex and age in months.)

Growth velocity will also be summarized in tables and figures. The calculation of growth velocity will be performed as follows: ((Current Height – Previous Height)/(duration (days)=(current date – previous date))) * 365.25 = Annualized Growth Velocity (cm/year). In addition to providing the calculated growth velocities per time period for each subject, the data will also be categorized according to worst growth velocity and last growth velocity by the following three categories: <4 cm/year; 4 to 5 cm/year; and >5 cm/year.

Analyses of WBCT Measurements of Femur and Tibia Lengths

The summary statistics of femur and tibia lengths (values and changes from baseline) will be calculated for each visit. Femur and tibia lengths will be summarized for the following subgroups:

- Male, Age 4-10 years
- Male, Age 11-18 years
- Female, Age 4-10 years
- Female, Age 11-18 years
- Female Age ≤8, Male Age ≤10 years
- Female Age >8 to <14, Male Age >10 to <14 years
- Age ≥14 years

Figures will include spaghetti plots of the femur and tibia change from baseline by chronological age.

Analyses of Evaluation of Bilateral Hips by WBCT Safety Read

Summary tables of hip abnormalities identified in WBCT safety reads (MOVE Trial and NHS) will include the following:

- Number and proportion of subjects with possible, probable, or definite AVN at baseline, and by specific abnormality (e.g., subchondral lucency, osteosclerosis, etc.)

- Number and proportion of subjects with any new post-baseline or prior time point AVN, and by specific abnormality

- Number and proportion of subjects with any new post-baseline or prior time point worsening in AVN, and by specific abnormality

- Number and proportion of subjects with improvement in baseline or prior time point AVN, and by specific abnormality

2.4.6 Analyses of Imaging Performance and Quality Metrics

Intra-Reader Agreement

For low-dose WBCT efficacy reviews, secondary reads will be summarized for the count and frequency of discrepancies for the post-baseline timepoints. Coefficients of variation, i.e., ratios of the standard deviation to the mean will also be computed by region and overall, for volumes of HO and for volumes of new HO, i.e., differences between timepoints.

For radiograph review and low-dose WBCT safety review, counts and frequencies for differences will be summarized.

Inter-Reader Agreement

Summary identical to those described above for intra-reader agreement will be performed to describe inter-reader agreement.

Summaries of differences for the primary reads, e.g., adjudication rates, will also be summarized.

2.4.7 Analyses of Pharmacokinetic Endpoints

Pharmacokinetic analyses will be conducted using the Principal PS and the Supplementary PS.

Pharmacokinetic Sample Concentrations

Blood samples for PK assessment of palovarotene will be collected at the first 3-month safety assessment during chronic-based treatment; if samples cannot be obtained during the first 3-month safety assessment, then they can be obtained during any subsequent 3-month safety visit. During flare-up dosing, blood samples for PK assessment of palovarotene will be collected once during the 20-mg regimen at any time between Study Days 4 to 28, and once during the 10-mg regimen at any time between Study Days 32 and 84, for the first treated flare-up only. If this is not possible for the first treated flare-up, PK blood samples can be obtained during any subsequent flare-up dosing. Pharmacokinetic blood samples will be collected at pre-dose and 3, 6, 10, and 24 hours post dose.

The determination of palovarotene plasma concentrations will be performed using a validated LC-MS/MS method. The time of sample collection as it relates to the time of dosing on the PK days will be recorded.

Plasma concentrations of palovarotene will be listed by subject, treatment group (chronic treatment (5 mg or weight equivalent dose) or flare-up-based treatment (10 mg and 20 mg or

weight equivalent dose)), and scheduled time. Concentrations will be summarized with descriptive statistics (N, number of participants with non-missing data (n), mean, SD, coefficient of variation [CV], median, minimum, and maximum) by treatment group and scheduled time using the Principal PS and Supplementary PS populations. Coefficient of variation will only be displayed if $n \geq 3$. Unscheduled timepoints will only be listed and not summarized. Plasma concentrations below the limit of quantification (BQL) will be set to 0. Data will be rounded for reporting purposes only, with data presented to three significant figures except for CV% which will be presented to 1 decimal place.

Figures of individual plasma concentration versus actual time profiles for palovarotene will be produced by treatment group and dose on both linear and semi-log scales for the Principal PS and Supplementary PS populations. Mean plasma concentration versus nominal time curves will be plotted by treatment group and dose on both linear and semi-log scales for the Principal PS and Supplementary PS populations.

Pharmacokinetic Parameters

Plasma concentrations versus time data will be analyzed using a standard non-compartmental model using Phoenix WinNonlin® (version 8.0 or later) or SAS applications. Actual collection time will be used in the calculation of plasma PK parameters.

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

- If PK concentration are missing at 0 and 24 hours, then the AUC and all PK parameters 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.
- Invalid pre-dose concentrations will be replaced with the 24-hour concentration. The pre-dose concentrations will be determined to be invalid if the concentrations are higher than any other concentrations following dosing as it will be assumed to have been obtained after dosing instead of before dosing.
- The following concentrations must be available to calculate the AUC: 3, 6, and 10-hour concentrations.

At least three data points not including T_{max} must be available to calculate λ_z (imputed 24-hour concentration can be used).

The following parameters will be determined where possible:

Parameter	Definition
$C_{max,ss}$	The maximum observed plasma concentration at steady state
$T_{max,ss}$	The time from dosing at which Cmax is observed
AUC_{0-24ss}	The area under the plasma concentration vs time curve, from time zero to 24 hours post-dose during the steady state dosing interval, calculated by the linear trapezoidal method
$C_{min,ss}$	The minimum observed plasma concentration at steady state
λ_z	Terminal elimination rate constant
$t_{½z}$	Terminal elimination half-life
CL/F	Apparent clearance following extravascular administration

For each treatment group and dose, descriptive statistics will be provided for all PK parameters using the Principal PS and the Supplementary PS populations; descriptive statistics for all parameters except T_{max} will include N, n, mean, SD, CV%, geometric mean, median, minimum and maximum. T_{max} will be summarized using median, minimum, and maximum.

2.4.8 Duration of Flare-up Treatment

For subjects receiving flare-up treatment, duration of flare-up treatment (days) is defined as total flare-up cycle dosing periods (days) across all flare-up cycles.

For a flare-up cycle, the flare-up cycle dosing period (days) is defined as days that subjects received flare-up-based treatment and is calculated as (last dose date of the flare-up cycle – first date of the flare-up cycle – days where flare-up medications were not taken/interrupted during the flare-up cycle + 1).

2.4.9 Duration of Non-Flare-up Treatment

Based on the study design, subjects are assumed to be on non-flare-up-based treatment from the date they received the first non-flare-up treatment unless they receive flare-up-based treatment. Thus, duration of non-flare-up treatment (days) will be calculated as (last non-flare-up treatment date - first non-flare-up treatment date – duration of flare-up treatment (days) - days where non-flare-up medications were not taken/interrupted + 1). Duration of flare-up treatment (days) is defined in Section 2.4.8.

3 INTERIM ANALYSIS

There were to be three early interim efficacy analyses and one final analysis in Part A, in addition to a descriptive analysis of safety and non-imaging efficacy endpoints, which was to be included in the interim CSR for the NDA for episodic palovarotene prevention of HO in patients with FOP. Hypothesis testing was to be performed using group sequential methods.

Approximately 80 subjects with the R206H mutation and no previous palovarotene experience were to be enrolled into the MOVE Trial and would be followed for 2 years. The first interim analysis was to occur when 35 subjects complete 1 year of follow-up, the second and third interim analyses were to occur when all subjects in the Principal EP had completed (i.e., have WBCT data) 12 months and then 18 months of follow-up, respectively.

3.1 Early Efficacy Testing

The primary efficacy analysis model and the one-sided significance threshold for each analysis are described in Section 2.4.1. Study success was to be declared if the posterior probability that palovarotene reduces the annualized change in new HO volume is greater than 1 minus the one-sided significance threshold. Under the assumptions in Section 1.4, the probability of declaring statistical significance at the interims is summarized in Table 2.

3.2 Futility Testing

A futility analysis was conducted for the second interim analysis based on the pre-specified criteria using square-root transformation and additional analyses without the square-root transformation. Because the futility boundary was crossed at the second interim analysis, all subsequent analyses of efficacy data will be supplied without applying thresholds for success or futility, as these thresholds are no longer relevant.

That said, at the request of the DMC, an analysis is planned to occur consistent with the planned timing of the third interim analysis, i.e., when all subjects in the Principal EP have completed 18 months of follow-up. The purpose is to re-evaluate the risk benefit ratio, both overall and in younger subjects with FOP. Because the dataset was unblinded to the sponsor after the second interim analysis, results at this third interim analysis will be shared with the sponsor – and the sponsor may participate in the conduct of the analyses.

3.3 Third Interim Analysis

An informal analysis will occur at the time of the planned third interim analysis. Table 8 lists the efficacy analyses that will be performed at this analysis.

Note that the third interim analysis dataset will also be summarized in the interim MOVE CSR for Study PVO-1A-301 (MOVE Trial) to be included in the palovarotene NDA and MAA for FOP. The efficacy analyses described in Table 8 will be included, in addition to the additional analyses, sensitivity analyses, analyses of secondary efficacy endpoints, and analyses of exploratory endpoints described in Section 2.4.4. Safety analyses in Section 2.1.4, analyses of imaging performance in Section 2.1.5 and analyses of pharmacokinetic endpoints in Section 2.1.6 will also be included. Further analyses may be added, as appropriate.

Table 8. Analyses to be Conducted at Interim Analysis #3

Analysis	Method	Cutoff	Trans.	Subset Subjects	Reason
Analysis of primary interest	Bayesian comp. P.	04Dec2019 for subjects <14 y.o., 24Jan2020 for subjects>=14.y.o.	None	All Subjects	Sponsor believes this is the more appropriate Bayesian analysis of efficacy
Alternative analysis of primary interest	wLME	04Dec2019 for subjects <14, y.o., 24Jan2020 for subjects>=14.y.o.	None	All Subjects	Explore if results of Bayesian analysis are consistent with results from simplified approach
Square-root transformation	Bayesian comp.P.	04Dec2019 for subjects <14, y.o., 24Jan2020 for subjects>=14.y.o.	Sq. Root	All Subjects	Show inconsistency of results with square-root transformation to results without square-root transformation
Alternative method w/ square-root transformation	wLME	04Dec2019 for subjects <14, y.o., 24Jan2020 for subjects>=14.y.o.	Sq. Root	All Subjects	Show that there is also an inconsistency of results with square-root transformation under simplified approach
Include scans after dosing stopped	Bayesian comp. P	All timepoints	None	All Subjects	Explore if results when with only scans done before dosing interrupted are consistent with results using all timepoints
Include scans after dosing stopped and square-root transformation	Bayesian comp.P.	All timepoints	Sq. Root	All Subjects	Protocol-specified primary analysis (ignoring dosing interruptions)
Older Population Cutoff 1	Bayesian comp. P	04Dec2019 for subjects <14, y.o., 24Jan2020 for subjects>=14.y.o.	None	Girls >=13 y.o., boys >= 15 y.o.	Explore if results differ by age
Younger Population Cutoff 1	Bayesian comp. P	04Dec2019 for subjects <14, y.o., 24Jan2020 for subjects>=14.y.o.	None	Girls < 13 y.o., boys < 15 y.o.	Explore if results differ by age
Older Population Cutoff 2	Bayesian comp. P	04Dec2019 for subjects <14, y.o., 24Jan2020 for subjects>=14.y.o.	None	Girls>=8 y.o., boys>=10 y.o.	Explore if results differ by age
Younger Population Cutoff 2	Bayesian comp. P	04Dec2019 for subjects <14, y.o., 24Jan2020 for subjects>=14.y.o.	None	Girls<8 y.o., boys<10 y.o.	Explore if results differ by age

4 SOFTWARE DOCUMENTATION

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

The Bayesian compound Poisson models will be fitted using the R statistical computing language and environment. The R programs will be run using v3.5.0 or later; packages used, including version number, will be identified in the CSR.

5 REFERENCES

1 Grantham JJ, Torres VE, Chapman AB, et al. Volume progression in polycystic kidney disease. N Engl J Med 2006;354:2122-2130.

2 Torres VE, Chapman AB, Deyvust O, et al. Tolvaptan in patients with autosomal dominant polycystic kidney disease. N Engl J Med 2012; 367:2407-2418

3 D'Agostino, RB., Jr., Tutorial in Biostatistics: Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group. *Statistics in Medicine*, 1998, 17, 2265-2281.

4 R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

5 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[5] allow a user to convert Global Physical Health, Global Mental Health, and Total 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.

Adult Global Physical Health		
Raw Score	**T-Score**	**Standard Error**
4	16.2	4.8
5	19.9	4.7
6	23.5	4.5
7	26.7	4.3
8	29.6	4.2
9	32.4	4.2
10	34.9	4.1
11	37.4	4.1
12	39.8	4.1
13	42.3	4.2
14	44.9	4.3
15	47.7	4.4
16	50.8	4.6
17	54.1	4.7
18	57.7	4.9
19	61.9	5.2
20	67.7	5.9

Adult Global Mental Health		
Raw Score	**T-Score**	**Standard Error**
4	21.2	4.6
5	25.1	4.1
6	28.4	3.9
7	31.3	3.7
8	33.8	3.7
9	36.3	3.7
10	38.8	3.6
11	41.1	3.6
12	43.5	3.6
13	45.8	3.6
14	48.3	3.7
15	50.8	3.7
16	53.3	3.7
17	56.0	3.8
18	59.0	3.9
19	62.5	4.2
20	67.6	5.3

Pediatric Self-Completed Total		
Raw Score	**T-Score**	**Standard Error**
7	16.0	3.4
8	17.1	3.6
9	18.3	3.7
10	19.7	3.8
11	21.2	3.8
12	22.8	3.7
13	24.4	3.6

Pediatric Proxy-Completed Total		
Raw Score	**T-Score**	**Standard Error**
7	14.7	2.9
8	15.3	3.1
9	16.0	3.2
10	16.9	3.4
11	18.1	3.6
12	19.4	3.7
13	21.0	3.8

Pediatric Self-Completed Total		
Raw Score	**T-Score**	**Standard Error**
14	26.1	3.6
15	27.6	3.5
16	29.2	3.5
17	30.8	3.5
18	32.4	3.6
19	34.0	3.6
20	35.6	3.6
21	37.2	3.6
22	38.8	3.6
23	40.4	3.6
24	42.1	3.7
25	43.9	3.7
26	45.7	3.6
27	47.5	3.6
28	49.2	3.6
29	51.1	3.7
30	53.3	3.9
31	55.7	4.2
32	58.3	4.5
33	61.1	4.9
34	64.2	5.4
35	67.5	6.1

Pediatric Proxy-Completed Total		
Raw Score	**T-Score**	**Standard Error**
14	22.7	3.8
15	24.4	3.7
16	26.1	3.7
17	27.7	3.7
18	29.4	3.8
19	31.2	3.8
20	32.9	3.8
21	34.6	3.8
22	36.2	3.8
23	37.9	3.9
24	39.7	4.0
25	41.7	4.0
26	43.6	3.9
27	45.4	3.8
28	47.3	3.9
29	49.3	4.1
30	51.8	4.4
31	54.5	4.7
32	57.3	5.0
33	60.2	5.4
34	63.2	6.0
35	66.1	6.5

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 x 10^3/mm^3	>16 x 10^3/mm^3

Laboratory Parameter	PCS Low	PCS High
Hemoglobin	<9.5 g/dL	>19.0 g/dL
Hematocrit	<34%	>54%
Platelet count	<75 x 10^3/mm^3	>700 x 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 msec only OR 2) increase from baseline ≥20 msec only OR 3) >200 msec and increase from baseline ≥20 msec
QRS Interval	None	1) >100 msec only OR 2) increase from baseline ≥10 msec only OR 3) >100 msec and increase from baseline ≥10 msec
QT Interval	None	1) >500 msec only OR 2) increase from baseline ≥60 msec only OR 3) >500 msec and increase from baseline ≥60 msec
QTcF and QTcB Interval	None	1) >500 msec only OR 2) increase from baseline ≥60 msec only OR 3) >500 msec and increase from baseline ≥60 msec

PCS Criteria for Vital Signs

Vital Sign	PCS Low	PCS High
Sitting SBP (mm Hg)		
Aged 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
Aged >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)		
Aged 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
Aged >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