FDA Briefing Document

NDA 215559

Drug name: Palovarotene

Applicant: Ipsen Biopharmaceuticals, Inc.

Endocrinologic and Metabolic Drugs Advisory Committee Meeting

6/28/23

Division of General Endocrinology
Office of Cardiology, Hematology, Endocrinology, and Nephrology

DISCLAIMER STATEMENT

The attached package contains background information prepared by the Food and Drug Administration (FDA) for the panel members of the Advisory Committee. The FDA background package often contains assessments and/or conclusions and recommendations written by individual FDA reviewers. Such conclusions and recommendations do not necessarily represent the final position of the individual reviewers, nor do they necessarily represent the final position of the Review Division or Office. We have brought the topic of palovarotene in patients with fibrodysplasia ossificans progressiva (FOP) to this Advisory Committee in order to gain the Committee's insights and opinions. The background package may not include all issues relevant to the final regulatory recommendation and instead is intended to focus on issues identified by the Agency for discussion by the Advisory Committee. The FDA will not issue a final determination on the issues at hand until input from the Advisory Committee process has been considered and all reviews have been finalized. The final determination may be affected by issues not discussed at the Advisory Committee meeting.

Table of Contents

Table of Tables

Table of Figures

Glossary

AC	Advisory Committee
CAJIS	Cumulative Analogue Joint Involvement Scale
CT	computed tomography
C-SSRS	Columbia Suicide Severity Rating Scale
DMC	Data Monitoring Committee
FDA	Food and Drug Administration
FOP	fibrodysplasia ossificans progressiva
FOP-PFQ	FOP-Physical Function Questionnaire
HO	heterotopic ossification
IA	integrated assessment
MRI	magnetic resonance imaging
NHS	natural history study
P-FAS	principal full analysis set
PROMIS	Patient-Reported Outcomes Measurement Information System
RAR	retinoic acid receptor
REMS	risk evaluation and mitigation strategy
WBCT	whole-body computed tomography
wLME	weighted linear mixed effects

1 Executive Summary/Draft Points for Consideration by the Advisory Committee

1.1 Purpose/Objective of the AC Meeting

The Food and Drug Administration (FDA) has convened this Advisory Committee (AC) meeting to discuss the proposed new drug palovarotene for the prevention of heterotopic ossification in patients with fibrodysplasia ossificans progressiva.

1.2 Context for Issues to Be Discussed at the AC

Fibrodysplasia ossificans progressiva (FOP) is a rare severely disabling disease caused by a sporadic or inherited gain-of-function mutation in the activin A type I receptor ACVR1 (ALK2) that also binds bone morphogenetic protein. The most common mutation (>90%) is arginine 206 to histidine (R206H), which renders ALK2 constitutively active to bone morphogenetic protein ligands, drives ectopic chondrogenesis and osteogenesis, and leads to heterotopic ossification (HO) in connective tissue, joints, and muscle.

Palovarotene is an orally bioavailable retinoic acid receptor (RAR) gamma (RARγ) selective agonist (retinoid) that appears to interfere with ALK2-mediated bone formation indirectly via dampening bone morphogenetic protein signaling by reducing SMAD1/5/8 phosphorylation. This is thought to lead to blockade of chondrogenic and osteogenic differentiation, reprogramming of progenitor cells into non-skeletal lineage, and prevention of abnormal endochondral bone formation.

The Applicant is seeking approval of palovarotene for prevention of HO in adults and children (aged 8 years and above for females and 10 years and above for males) with FOP.

The proposed dosing regimen is 5 mg daily, increasing at the time of flare-up symptoms to 20 mg daily for 4 weeks followed by 10 mg daily for 8 weeks.

1.3 Brief Description of Issues for Discussion at the AC

The Advisory Committee will discuss whether it is reasonable to conclude, based on available data, that palovarotene, used chronically, is a safe and effective drug in patients with fibrodysplasia ossificans progressiva. One phase 2 study with multiple open-label extension phases and one phase 3 study (Study 301) were conducted. Study 301 was a single-arm, open-label study with a natural history study (NHS) as the external control to support the efficacy of palovarotene 5 mg daily with flare-up dosing (20 mg daily for 4 weeks, followed by 10 mg daily for 8 weeks) as needed. The primary endpoint was annualized new HO volume. The study crossed the prespecified futility boundary at the second interim analysis. The Applicant believes that the failure of the primary analysis was due to the application of square root transformation on observed outcome, which moved the statistical conclusion from significant therapeutic benefit to showing futility of the treatment. The Applicant performed various post hoc analyses with and without square root transformation to support the efficacy of palovarotene.

With respect to the NHS external control, it is generally recognized that support for effectiveness can emerge using an externally controlled trial when certain conditions are met. However, discussion is necessary regarding whether these conditions are met in the context of this submission; specifically, whether the NHS subjects are sufficiently similar to Study 301 subjects, whether certain potential biases have been reasonably addressed, and whether the results provided compelling evidence of benefit remain questions to address.

An issue relevant to benefit-risk considerations is a numeric increase in reported flare-up events among palovarotene-treated patients. Retinoids have been associated with myositis and other musculoskeletal adverse effects, including back pain, arthralgia, and myalgia. There is uncertainty as to whether palovarotene may in some cases trigger flare-ups, or symptoms that could mimic flare-ups, and if so, whether the development of new HO over longer periods of treatment than were assessed in the palovarotene studies, e.g., greater than one year, could be worse than without palovarotene treatment.

1.4 Draft Points for Consideration

The Applicant is seeking approval of palovarotene for prevention of heterotopic ossification in adults and children (aged 8 years and above for females and 10 years and above for males) with fibrodysplasia ossificans progressiva.

- Discuss the evidence of effectiveness for palovarotene demonstrated in Study 301. In your discussion consider the following:
    - The use of post hoc analyses to support a demonstration of efficacy.
    - The interpretability of the results using an external NHS control.
- Discuss your view of the apparent increase in reported flare-up events in subjects treated with the proposed palovarotene dosing regimen and its relevance to benefit-risk considerations.

2 Introduction and Background

2.1 Background of the Condition/Standard of Clinical Care

Fibrodysplasia ossificans progressiva (FOP) is a rare disease with approximately 800 confirmed cases globally. Beginning in early childhood, patients with FOP develop extra-skeletal bone formation in muscles, tendons and ligaments known as heterotopic ossification (HO). HO is episodic, with some events starting with soft tissue inflammation (flare-up). The accumulation of extra-skeletal bone is cumulative and irreversible, causing restriction of movement, deformities, and severe disability. Complications may include chest wall deformity (thoracic insufficiency syndrome), ankyloses of the temporomandibular joints resulting in severe tooth decay and malnutrition, and localized skin breakdown. Most patients use mobility assistance by their 20s due to HO around the hips. Survival is shortened (median 56 years), most often by cardio-respiratory failure and pneumonia.

Currently, there are no approved targeted therapies for treatment of FOP. Current conventional therapy for FOP is aimed at symptom relief by decreasing inflammation and treatment of chronic pain (Kaplan FS, 2021). Attempts at surgical excision of lesions usually leads to reactivation of disease and new HO.

2.2 Pertinent Drug Development and Regulatory History

Clinical drug development of palovarotene for the prevention of HO in patients with FOP was initiated in 2014. Palovarotene was initially under development for treatment of chronic obstructive pulmonary disease but was found to be ineffective for this indication. The demonstration of inhibition of HO in animal models of FOP with palovarotene led to its clinical development program in patients with FOP (Appendix Section 5.1 (Clinical Studies Conducted for Palovarotene Treatment in Subjects With FOP) and Appendix Section 5.6 (Nonclinical Assessment of the Potential Effectiveness of Palovarotene)).

The Applicant began enrolling study PVO-1A-001 (study 001, NHS), the natural history study, in 2014, approximately the same time when clinical studies with palovarotene began. Study PVO-1A-201 (Study 201) was a randomized, 12-week, placebo-controlled study initiated as a proof of concept and dose-finding study in subjects with FOP with acute flare-up symptoms. Subjects who completed Study 201 could enroll in the open label extension study PVO-1A-202 (Study 202). Initially, Study 202 continued the flare-based dosing paradigm (10 mg daily for 2 weeks followed by 5 mg daily for 4 weeks) with evaluation of physical function indices (Part A).

As the NHS and studies 201 and 202 progressed, emerging data suggested that HO occurred not only with flare-ups, but also occurred without any flare-up symptoms. Additionally, results from Study 201 and Study 202 Part A suggested that the dosing regimen was not optimized. Therefore, the Applicant amended Study 202 (Part B) to include a 5 mg chronic daily dose and an increase in the dose and duration of flare-up dosing (20 mg daily for 4 weeks followed by 10 mg daily for 8 weeks). Assessment of HO by low dose whole-body CT (WBCT) was added to Study 202 Part B. Enrollment in 202 Part B consisted of subjects from Part A and 18 additional subjects who had not had flare-up symptoms for at least 4 weeks. In Study 202 Part C, palovarotene dosing was the same as in Part B, except that the criteria for initiating flare-up dosing were broadened. Assessment of HO by WBCT was the primary endpoint in Study 202 Part C. Enrollment into 202 Part C consisted of subjects continuing from 202 Part B; no new subjects were enrolled in Part C.

Discussion of the design of Study 301 occurred at an End of Phase 2 meeting. The patient-reported outcomes measures in the phase 2 trial did not show change, likely given the short duration of the trial. However, findings from the NHS demonstrated that higher amounts of HO seen on WBCT correlated with worsening functional status, supporting the use of whole-body HO as the primary endpoint in clinical studies investigating treatment of FOP. FDA agreed with the use of low-dose WBCT for assessment of HO for the primary endpoint for Study 301. The dose chosen for Study 301 was the chronic 5 mg daily plus flare up dosing of palovarotene 20/10 mg regimen that was evaluated in Study 202 Part C (modification of the regimen in 202

Part B). The WBCT scan results were not available to inform the dosing regimen, so the Applicant utilized the dose-response seen in animal models and pooled data from the flare-up dosing regimens from studies 201/202 to support the proposed phase 3 dosing regimen. The Applicant also proposed using the NHS as an external control arm. FDA recommended that randomized, placebo-controlled data would provide more convincing evidence of efficacy for the phase 3 trial. However, Study 301 was initiated in July 2017 as a single-arm study with the NHS control evaluating palovarotene for decreasing HO in adult and pediatric subjects with FOP as assessed by low-dose WBCT.

Because alteration in growth plate function is a known side effect of retinoid therapies, all palovarotene clinical studies contained a bone safety monitoring plan. In October 2019, the Data Monitoring Committee (DMC) increased the bone safety monitoring procedures after reports of premature epiphyseal fusion in pediatric subjects. Additional data were reviewed, and in December 2019, FDA placed a partial clinical hold for subjects under 14 years of age because of the risk of closure of the growth plates.

In January 2020, the Applicant informed FDA that based on Study 301's second interim analysis, futility was declared based on the protocol-defined rules. Dosing in studies 301 and 202C was stopped. The DMC was presented with additional post hoc analyses conducted after the sponsor had unblinded the efficacy dataset to confirm the results of the pre-specified interim analysis, and to better understand the results. It was confirmed that the futility analysis using the pre-specified Bayesian compound Poisson analysis was conducted appropriately. The Applicant explored additional post hoc analyses (discussed below), and based on these results, the DMC recommended that palovarotene therapy for subjects older than 14 years could be restarted in Study 301.

A pre-NDA meeting was held with the Applicant on July 28, 2020. FDA agreed that the data could be submitted for review. FDA outlined concerns regarding the change in statistical methods after data had been unblinded.

The palovarotene NDA was initially submitted March 21, 2021. During the review and preparation for a planned AC meeting, it was discovered there were issues with the imaging vendor that required image re-read and re-analysis of data for the primary endpoint. FDA met with the Applicant and discussed the situation. The Applicant withdrew the NDA in August 2021 to fully evaluate the issues with the imaging vendor and update datasets as needed.

The application was resubmitted April 29, 2022. The Applicant, however, informed FDA that additional WBCT data from Studies 301 and 202C was expected to become available, consisting of all scans conducted between February 2020 (the NDA efficacy cutoff date, at the time of treatment pause) and September 2022, when both studies concluded. During this interval, many subjects had re-started palovarotene treatment (after a pause in dosing ranging from 3 to 24 months), and it was anticipated that the additional data may help inform the efficacy and safety assessments and another review cycle would be needed. On February 16, 2023, the Applicant re-submitted the NDA.

3 Summary of Issues for the AC

3.1 Efficacy Issues

In support of the proposed indication, the Applicant intends to rely on a single arm, open-label, externally controlled, phase 3 study (Study 301). Of note, Study 301 failed its prespecified primary efficacy analysis, and the NDA submission relies on post hoc analyses from Study 301 to support the effectiveness of palovarotene. Therefore, the following are key efficacy issues during the FDA review:

- Key efficacy issue 1: appropriateness of reliance on post hoc analyses to support effectiveness
- Key efficacy issue 2: use of an external control group (NHS) for evaluation of the proposed dosing regimen

3.1.1 Sources of Data for Efficacy and Efficacy Results

As outlined in Section 5.1 (Clinical Studies Conducted for Palovarotene Treatments in Subjects With FOP), four studies contribute data to support the efficacy of palovarotene: study PVO-1A-001 (the NHS), study PVO-1A-201, study PVO-1A-202 (Parts A, B, and C), and study PVO-1A-301.

3.1.1.1 Natural History Study PVO-1A-001

Study PVO-1A-001 (study 001, NHS) was a multinational, prospective natural history study with an observation period of 36 months. The study had two parts. In Part A, 10 adult subjects were enrolled and both low-dose WBCT and DXA imaging were evaluated to determine which imaging modality provided the optimal method to assess whole body HO. The imaging committee chose WBCT as the optimal modality. Part B of the study was expanded to include all subjects less than 65 years old who were clinically diagnosed with FOP with a documented R206H mutation or who were believed to carry the mutation. Imaging for whole body HO was scheduled to occur at baseline, month 12, month 24, and month 36. Flare-up symptoms (e.g., pain and soft tissue swelling) were reported by subjects, and in a subset of confirmed flare-up events, the subjects underwent in-clinic evaluations and imaging of new HO over a 12-week period. In addition, functional assessments occurred yearly, including recording the use of assistive devices, the FOP-PFQ (FOP-Physical Function Questionnaire), PROMIS (Patient-Reported Outcomes Measurement Information System) Global Health Scale, and CAJIS (Cumulative Analogue Joint Involvement Scale for FOP).

A total of 114 eligible subjects were enrolled in the NHS, and 92 subjects completed the 12-month visit. Eight subjects left the NHS and joined Study 201 when they had flare-up symptoms before completing one year in the NHS. Thirteen subjects transitioned to Study 202 Part B. Additionally, all study sites for the NHS also participated in Study 301. Subjects could enroll in Study 301 if they met the inclusion and exclusion criteria. A total of 39 subjects chose to enroll in Study 301 from the NHS. All completed at least one year in the NHS prior to enrolling in Study 301.

3.1.1.2 Study PVO-1A-201

Study 201 was a randomized, 12-week, placebo-controlled study initiated as a proof of concept and dose-finding study in subjects with FOP and flare-up symptoms.

The study used an adaptive design. For the first 16 patients enrolling in this study (Cohort 1), the randomization ratio was 3:1 (palovarotene 10/5 mg: placebo). Subsequently, the 5/2.5 mg regimen was added, and the randomization ratio changed to 3:3:2 (10/5 mg: 5/2.5 mg: placebo) for the last 24 patients (Cohort 2).

Population – 201

Subjects with FOP and within 7 days of onset of a distinct active flare-up were enrolled. The index flare-up was defined as at least two out of six specified symptoms (pain, swelling, decreased range of motion, stiffness, redness, warmth), consistent with the subject's previous flare-ups, and confirmed by the investigator. Enrollment was initially limited to patients at least 15 years of age who were at least 90% skeletally mature (based on bone age, derived from hand/wrist x-ray). During enrollment of Cohort 2, the minimum age was lowered to 6 years.

Study Treatment – 201

Subjects received one of three regimens: placebo, palovarotene 10 mg daily for 2 weeks followed by 5 mg daily for 4 weeks (10/5), or palovarotene 5 mg daily for 2 weeks followed by 2.5 mg daily for 4 weeks (5/2.5). Dosing for subjects under 18 years of age was adjusted based on body weight. Treatments could not be extended beyond six weeks for any ongoing or new symptoms.

Imaging Methods – 201

In clinical practice, plain radiographs were sometimes used to assess formation of new HO at FOP flare-up sites, but the sensitivity and optimal timing of this and other potential HO imaging modalities were largely unknown prior to Study 201. In Study 201, imaging of the index flare-up site included assessments of HO by standard radiograph and CT scan, with measurement of HO volume and assessment of soft tissue edema at the site by MRI or ultrasound. This imaging was conducted at baseline, 6 weeks (end of treatment) and 12 weeks (end of post-treatment observation period). Qualitative scoring of the x-rays was on a six-point scale (Section 5.2: Scoring Scale to Measure New HO on Plain Radiographs in Studies 201 and 202A). Quantitative assessments were also done and reported. Several radiology reading processes were used. In the Primary Read process, two radiologists evaluated a single modality. In the Global Read process, radiologists evaluated all modalities (x-ray, CT, MRI/ultrasound) together to assess for change.

Endpoints – 201

The primary endpoint was the percentage of subject 'responders' as defined by no or minimal new HO (grade ≤3 on 0-6 scale) at the x-ray imaged flare-up site compared with baseline as

assessed by plain radiographs at Week 6. Secondary endpoints included assessment by other imaging modalities, including change in HO volume at the flare-up site as assessed by low dose CT scan, and presence of soft tissue swelling and/or cartilage formation at the flare-up site as assessed by MRI or ultrasound.

Results – 201

Disposition

All 40 subjects enrolled in Study 201 completed the study and are included in the full analysis population.

Demographics

The mean age of the study population was 21 ± 10.8 years, with a median of 21 years, and a range of 7 to 53 years. When grouped by age, 13 (32%) subjects were under 15 years of age and 27 (68%) were 15 years or older. Forty-five percent of the enrolled population was male and 55 percent female. The treatment groups were generally similar with respect to the subjects' FOP history and disease characteristics, including the number and location of regions with HO present by physical exam at baseline. The mean number of flare-ups per year reported at baseline was 2.3 in the placebo group, 2.0 in the palovarotene 5/2.5 mg group and 4.6 in the palovarotene 10/5 mg group. HO was present at the flare-up site at baseline by CT in 44%, 67% and 65% of subjects in these respective groups. Use of glucocorticoids to treat the flare-up was allowed during the study, and glucocorticoids were used in 78 to 95% of subjects.

Efficacy Analysis

The Primary Read of standard radiographs of the flare-up site at week 6 compared to baseline showed more than minimal new HO in only 2 subjects overall: 1/9 (11%) in the placebo group, 1/9 (11%) in the palovarotene 5/2.5 mg group and 0/21 (0%) in the palovarotene 10/5 mg group. Therefore, the study did not demonstrate efficacy based on this prespecified primary endpoint.

A greater number of new HO flare-up sites were detected at week 12 compared to week 6, and CT was found to be more sensitive than standard radiographs. At week 12, new HO was detected in 4 (40%) subjects in the placebo group, 2 (22%) in the palovarotene 5/2.5 mg group and 4 (20%) in the palovarotene 10/5 mg group (per-protocol population). The mean (SD) volume of new HO at the flare-up site was 16 (41.6) cm^3 in the placebo group, 1 (3.2) cm^3 in the palovarotene 5/2.5 mg group, and 4 (11.8) cm^3 in the palovarotene 10/5 mg group (averaged across all subjects, including those with no new HO). There were small declines from baseline in mean scores for flare-up pain in each of the 3 treatment groups.

3.1.1.3 Study PVO-1A-202

Study 202 was an uncontrolled, open-label extension of Study 201 in which dosing regimens, imaging techniques, and endpoints evolved over time. The data from 202 Parts A and B were used to inform the design of Study 301, but do not substantially contribute to the assessment

of palovarotene's efficacy of the proposed dosing regimen. For informational purposes, discussion of Study 202 Parts A/B is provided in Appendix Section 5.3 (Study PVO-1A-202)

Part C of Study 202 is considered relevant to the assessment of palovarotene efficacy in FOP because it followed the same treatment protocol as Study 301, in contrast to 202 Part B which used different criteria for treating flare-ups. Studies 202C and 301 began enrollment around the same time in 2017.

Population – 202C

Most of the subjects in 202 Part B continued from Studies 201 and 202A, while a smaller group of patients with FOP was newly enrolled and began palovarotene treatment in 202B. The 202B subjects continued into 202 Part C; no new subjects were enrolled in Study 202 Part C.

Study Treatment – 202C

Subjects received a palovarotene 5 mg daily dose, with flare-up dosing of palovarotene 20 mg for 4 weeks followed by 10 mg for 8 weeks, extended if needed for persistent symptoms (20/10 mg regimen). As in Study 301, all pediatric and adult subjects in 202C were to receive the chronic/flare-up regimen (with reduced doses for skeletally immature subjects); only one symptom was required to begin flare-up dosing; and intercurrent flare-ups at a different site could restart the flare-up dosing cycle. In 2019, protocol amendments for both studies 202C and 301 added a provision that flare-up dosing could also be triggered by a traumatic event considered likely to result in a flare-up. Beginning in Dec 2019-Jan 2020, all treatment was paused; the same regimen was later re-started in a subset of subjects in Study 202C and Study 301.

Imaging Methods – 202C

HO was assessed solely by whole body CT in 202 Part C; flare-up site-specific imaging was not conducted. For skeletally mature subjects, who began chronic treatment in 202B, Parts B and C were continuous, with WBCT performed annually starting at Part B baseline. For skeletally immature subjects, WBCT scans occurred at the initiation of chronic dosing in Part C and then annually. Scans were conducted at irregular intervals during and after the treatment pause, due in part to COVID restrictions. WBCT readings for 202C were conducted at the same central imaging facility used for Studies 301 and the NHS, using the same process involving HO measurement in each of 9 body regions.

Endpoints – 202C

The primary efficacy endpoint for 202C is the annualized new HO volume as assessed by low-dose WBCT scan. This is the same as the primary endpoint in Study 301. These were single-arm studies, with data from the NHS used for comparisons. However, unlike Study 301, Study 202 was an extension study that was not powered to demonstrate efficacy via comparisons to the NHS.

Analyses specific to 202 Part C were conducted on subjects who underwent at least two WBCT scans during Part C, provided that the first scan was not obtained within 1 month of a flare-up (in order to align with the enrollment criteria for Study 301 and the NHS). Analyses of data from the "Pre-Pause" treatment period were conducted on the subset of 202C subjects with at least two WBCT scans that were obtained prior to the extended treatment pause which started Dec 2019 – Jan 2020. Other analysis subgroups were subjects with data from the "Interruption" phase (mostly, but not entirely off-treatment for an interval of variable duration in 2020-2022); and subjects who subsequently restarted palovarotene in 2020-2021 ("Post-Pause" treatment population).

Results – 202C

Disposition

A total of 46 subjects enrolled in 202C and received treatment for a mean of 20 months in the Pre-pause period, including 36 subjects who required flare-up dosing in this phase. Among 202C subjects, there were 30 who underwent at least two WBCT scans in Part C, wherein the baseline scan was not within 1 month of a flare-up. Among these 30 subjects, there were 23 subjects with ≥2 scans prior to dose interruption (Pre-pause population); 19 subjects with data (i.e., ≥2 scans) during the Interruption period; and 15 subjects with data (≥2 scans) following restart (Post-Pause Treatment population). No information is available regarding the criteria used for selection of subjects to restart treatment.

Demographics

Study 202C subjects were 50% female, similar to Study 301 and the NHS. The mean age of the subjects in 202C was 21.1 years, which is older than the mean age for subjects in the NHS (17.5 years) or Study 301 (15.1 years). Whole body HO volume at baseline was also greater in 202C compared to the other studies, likely due to the age differential.

Efficacy

The following table below shows summary statistics for the primary endpoint, annualized volume of new HO, in 202C subjects with available data (≥2 WBCT scans) during each of the study phases. The rate of new HO in the pre-pause palovarotene treatment phase was 19.0 cm^3/year in 202C; for comparison, the rate in untreated NHS subjects was 23.7 cm^3/year. Among 15 subjects who restarted palovarotene following the pause in treatment, the rate of new HO in this phase was lower (6.4 cm^3/year).

Table 1. Study 202C: Annualized New HO Volume by Treatment Phase

Study Phase	Pre-Pause Treatment	Interruption	Post-Pause Treatment
n	23	19	15
Time interval (mean, months) [a]	11.7	19.6	11.8
Annual new HO volume (cm^3/year), mean	19.0	26.8	6.4

Source: Table 2 in Response to CRL, 02/16/23 resubmission M 1.11.4; IR response submitted 05/04/23
[a] Duration between first and last WBCT scan during the study/phase
Annualized new HO data represents change in volume between first and last WBCT scan within the phase, divided by time interval
Abbreviations: HO, heterotopic ossification

The following table below summarizes WBCT data in Study 202C over the pre-pause treatment period, the dosing interruption period, and the post-pause treatment period for subjects who contributed WBCT data in all three periods. As with the previous table above, rate of new HO tended to increase from the pre-pause treatment period to dosing interruption period, and then decrease during the post-pause treatment period.

Table 2. Study 202C: Annualized New HO Volume by Treatment Phase (Including Only Subjects With Data for All Three Phases)

Study Phase	Pre-Pause Treatment	Interruption	Post-Pause Treatment
n	9	9	9
Annual new HO volume (cm^3/year), mean	23.2	34.0	10.6

Source: FDA reviewer
Annualized new HO data represents change in volume between first and last WBCT scan within the phase, divided by time interval
Abbreviations: HO, heterotopic ossification

3.1.1.4 Study PVO-1A-301

Study 301 was a single-arm study with the primary objective of evaluating the efficacy of the chronic/flare-up regimen of palovarotene in decreasing new HO development in adult and pediatric patients with FOP. The Applicant chose to use untreated patients from the NHS as an external control arm. The single-arm design of Study 301, with use of the NHS as an external control group, raises issues of the comparability of these studies/groups, which are discussed in detail in Section 3.1.2 (Efficacy Issues in Detail), Issue 2. Sample size was calculated assuming 30% reduction in number of regions with any new HO and 50% reduction of new HO volume in subjects with new HO in one year.

Population - 301

Adult and pediatric patients ≥4 years of age with a clinical diagnosis of FOP and no flare-up symptoms within the previous 4 weeks were eligible to enroll in Study 301. The planned enrollment size was 110 subjects. To facilitate comparisons with the NHS, most of the subjects (i.e., the Principal Analysis Population) were required to have the classic R206H mutation and no prior exposure to palovarotene. There were exclusion criteria based on known retinoid safety issues, including pregnancy and increased suicide risk. Clinical trial study sites for the NHS were also study sites for Study 301. At the start of Study 301, subjects from the NHS could choose to enroll in Study 301 if they met the inclusion and exclusion criteria. A total of 39 subjects transitioned from the NHS into Study 301.

Study Treatment - 301

Study 301 subjects received chronic treatment of 5 mg palovarotene once daily with dose increases at the time of acute FOP flare-up episodes. At the onset of any symptom of a flare-up (e.g., pain, swelling) in any location, or a high-risk trauma event (blunt muscle trauma, surgery, intramuscular injection, mandibular block for dental procedure, muscle fatigue) as confirmed by the investigator, subjects increased to flare-up dosing of palovarotene 20 mg daily for 4 weeks, followed by palovarotene 10 mg daily for 8 weeks. If flare-up symptoms persisted, treatment could be extended with palovarotene 10 mg daily in 4-week increments. If a subject experienced an intercurrent flare-up (a flare-up at a new location, or marked worsening of the initial flare-up), the treatment cycle was restarted with palovarotene 20 mg daily. When acute symptoms had resolved and the flare-up treatment cycle was completed, chronic 5 mg daily treatment resumed. Subjects <18 years old with <90% skeletal maturity received reduced equivalents of both chronic daily and flare-up doses based on body weight.

Imaging Methods - 301

HO was assessed by WBCT scan every 6 months in Study 301 (during the first two years), and every 12 months in the NHS. At a central imaging facility, radiologists were blinded to protocol number, subject identifying information, exam date, visit name, total number of imaging timepoints, investigative sites and reason for exam. Radiology scans from these two studies were interspersed for readings that were blinded to the study/treatment group and other clinical data pertinent to the subject. In baseline scans from either study, each of nine body regions (excluding the head) was evaluated for the presence of HO by two radiologists. In each CT slice, regions of interest were delineated by reviewers around each area of HO; data from adjacent slices were integrated by image analysis software to yield volume of the regions of interest. Total volume of HO (cm^3) was calculated by consensus for each region, and summed to yield baseline total body HO.

For post-baseline scans, HO volume was recalculated for any body region that appeared to contain, upon initial qualitative review, new HO in comparison to the previous scan (either a new area of HO or enlargement of an existing lesion), with adjudication of any disagreements among radiologists. Any reduction in HO volume of other lesions in that region would also be captured, thus the change in HO volume from baseline could be positive or negative for any region and/or for total-body. HO volume was not remeasured post-baseline in regions where there was no apparent new HO compared to the previous scan. The data were summated by body region; no data pertaining to individual HO lesions were recorded.

For the subjects who transitioned from the NHS into Study 301, a subject's last WBCT scan from the NHS served as their baseline WBCT scan in Study 301. These scans were read twice (as NHS post-baseline, and as Study 301 baseline) because of the differing procedures involved.

Under the protocol, WBCT scans were to be conducted every 6 months through month 24 and then annually (months 36 and 48). During the latter phase, which corresponds to the period

that began with the extended treatment pause and the COVID pandemic, most of the scans obtained were in subjects who restarted treatment.

Endpoints - 301

The primary endpoint was the annualized change in new HO volume. The key secondary endpoint was the proportion of subjects with any new HO. The other secondary endpoints were the number of body regions with new HO, the proportion of subjects reporting flare-ups and rate of flare-ups per subject-month.

Exploratory functional endpoints (Section 5.4: Functional Outcome Scales) included change from baseline in range of motion assessed by CAJIS (Cumulative Analogue Involvement Scale for FOP), change from baseline in physical function using age-appropriate forms of the FOP-PFQ (FOP Physical Function Questionnaire), and change from baseline in physical and mental function for subjects ≥15 years old and mental function for subjects <15 years old using age appropriate versions of the PROMIS (Patient-Reported Outcomes Measurement Information System) Global Health Scale.

Statistical Analysis - 301

The prespecified primary analysis for the primary endpoint was based on a Bayesian compound Poisson model. Under this model, the change in HO since the previous CT scan was modeled as a compound distribution of the number of body regions with new HO and the new HO volume per region where new HO occurred (Section 5.5: Additional Statistical Analyses). The compound Bayesian model counted the occurrence of a positive new HO as an event. The treatment effect was expressed as the ratio of the annualized new HO between treated and untreated subjects, which was estimated by the ratio of annual event rates multiplied by the ratio of the growth rates per event. This method is equivalent to estimating the mean annualized new HO by multiplication of the mean annual number of events and the mean growth per event. The Bayesian compound model was estimated based on the number of regions with positive new HO and the average growth per region where positive new HO occurred since the previous WBCT scan. Negative new HO values were set to 0 in the analyses. The prespecified analysis included a square root transformation of each change in HO since the previous scan by region. The Applicant chose this statistical model "to more appropriately accommodate the high degree of variability in the volume of new HO assessed by WBCT." Additionally, square root transformation would likely decrease the variance of the endpoint by shrinking any extremely large values of new HO volume.

Three early interim analyses and one final analysis were pre-specified in the protocol for Study 301. The trial crossed the futility boundary at the second interim analysis and dosing was stopped by the Data Monitoring Committee (DMC). To confirm the pre-specified interim analysis and to better understand the results, the Applicant unblinded the efficacy dataset and the futility results were confirmed. The Applicant then conducted additional post hoc analyses, including the same Bayesian compound model but without the square root transformation and

a weighted linear mixed effects (wLME) model, which appeared to show evidence of benefit for palovarotene. Upon review of the additional post hoc analyses, the DMC noted that the impact of the pre-specified square root transformation was unexpectedly large and appeared to have moved the statistical conclusion from significant therapeutic benefit to showing futility of the treatment. The DMC concluded that "*the dilemma created by these highly disparate results precludes a confident conclusion about futility*." Dosing in Study 301 was thus allowed to resume.

In this NDA submission, the Applicant's evidence of efficacy relies on the post hoc analyses conducted after unblinding of the efficacy dataset.

Results - 301

Disposition

The principal full analysis set (P-FAS) included all Study 301 and NHS subjects who have the classic R206H mutation and had baseline and at least one post-baseline evaluable WBCT scan. The principal safety set consisted of 99 subjects who received at least one dose of palovarotene and 111 subjects in the NHS with no previous exposure to palovarotene.

Table 3. Studies 301 and NHS: Patient Disposition

Disposition Category	Study 301 Palovarotene	NHS Control
Subjects enrolled	107	117
Principal enrolled population	99	114
Supplementary enrolled population	8	N/A
Principal safety set (PSS)	99	111
Principal full analysis set (P-FAS)	97	101
Ongoing in study	47	2
Completed study	2	31
Discontinued study	58	81
Enrolled in another interventional study	0	66
Adverse event	12	N/A
Death	0	1
Worsening of clinical condition	0	1
Subject withdrew consent	31	9
Other	15	4

Source: Study 301 data are based on cutoff date 01/31/22, per Table 7 in Safety Update submitted on 08/10/22; NHS data per Study 301 CSR, Table 12 with cutoff date 02/28/20
Abbreviations: NHS, natural history study

Because of the safety related clinical hold in pediatric subjects (age <14 years) and the interim analysis finding of futility in this study (as described below), all dosing was stopped in Dec 2019-Jan 2020. Subsequently, 49 subjects in Study 301 (including 46 in the P-FAS) who were age ≥14 years restarted treatment after a pause that ranged from 3-24 months. The criteria that were used for restarting treatment remain unclear. The last subject visit of Study 301 occurred in Sept 2022; all WBCT data to this point have been submitted.

Among NHS subjects, more than half left this study before completion to enroll in an interventional study, including 8 subjects who enrolled in Study 201 at the onset of a flare-up, and others who enrolled in Study 202B (n=13), Study 301 (n=38), studies of other potential FOP therapies under different sponsors (n=5) or an unidentified study (n=1). One subject completed the NHS before enrolling in Study 301.

Demographics and Baseline Characteristics

As shown in Table 4, the two study populations were similar in demographics except that NHS subjects were somewhat older than Study 301 subjects (mean age 18 versus 15 years at baseline). NHS subjects also tended to have worse measures of FOP disease severity (whole-body HO volume, number of body regions with HO, CAJIS and FOP-PFQ scores measuring impaired range of motion and physical function, respectively). These differences are, at least in part, likely age-related, as cross-sectional analyses of baseline data showed that increasing age correlates with greater HO accumulation and with worsening FOP-related functional impairments.

Study 301 and NHS subjects were generally similar with respect to other FOP-related history, including mean age at FOP diagnosis (6.6, 7.5 years), history of great toe malformations (99%, 100%), cervical spine malformations (44%, 51%), hearing loss (46%, 35%), and family history of FOP (1%, 5%).

Table 4. Studies 301 and NHS: Baseline Demographic and Clinical Characteristics, Principal Safety Set[a]

Characteristic	Study 301 Palovarotene (N=99)	NHS Control (N=111)
Sex, n (%)		
Male	53 (54%)	60 (54%)
Female	46 (47%)	51 (46%)
Age, years		
Mean (SD)	15 (9.6)	18 (9.8)
Median (min, max)	13 (4,61)	15 (4,56)
Age groups (years), n (%)		
4 to <11	34 (34%)	29 (26%)
11 to <18	41 (41%)	37 (33%)
≥18	24 (24%)	45 (41%)
Race, n (%)		
White	70 (71%)	81 (73%)
Asian	9 (9%)	9 (8%)
Black/African American	1 (1%)	0
Other/Unknown[b]	19 (19%)	21 (19%)
Ethnicity, n (%)		
Hispanic	19 (19%)	23 (21%)
Non-Hispanic	69 (70%)	72 (65%)
Not reported[b]	11 (11%)	16 (14%)

Characteristic	Study 301 Palovarotene (N=99)	NHS Control (N=111)
Country of participation, n (%)		
United States	38 (39%)	38 (34%)
United Kingdom	9 (9%)	17 (15%)
France	10 (10%)	15 (14%)
Sweden	8 (8%)	0
Australia	6 (6%)	6 (5%)
Italy	5 (5%)	14 (13%)
Spain	6 (6%)	0
Canada	4 (4%)	1 (1%)
Brazil	5 (5%)	1 (1%)
Argentina	4 (4%)	19 (17%)
Japan	4 (4%)	0
Baseline characteristics		
Whole-body HO volume[a] (cm^3)		
Mean (SD)	231.2 (292.5)	312.5 (373.6)
Median (min, max)	127.2 (0, 1382.0)	195.4 (0, 1906.2)
Number of body regions with HO[a], mean	6.1	6.5
Number of flare-ups within 12 months of enrollment, mean (SD)	1.4 (1.86)	2.5 (5.98)
Number of flare-ups within 12 months of enrollment, median (min, max)	1.0 (0, 8)	1.0 (0, 40)
Mean CAJIS score	10.0	11.8
Mean FOP-PFQ score	44.3	47.0

Source: Study 301, ADSL and ADHOV, ISE ADNHSMOV and Tables S39A and S40.1
[a] HO data are based on Principal Full Analysis Set of Study 301 (N=97) and NHS (N=101)
[b] Large number of subjects with race/ethnicity data missing is partly due to French legal restrictions
Abbreviations: CAJIS, Cumulative Analogue Joint Involvement Scale; FOP, fibrodysplasia ossificans progressiva; HO, heterotopic ossification; N, number of subjects in treatment group; n, number of subjects with given characteristic; NHS, natural history study; PFQ, physical function questionnaire; SD, standard deviation

Primary Efficacy Endpoint Results

The major efficacy analyses, as detailed below, represent data collected in the "pre-pause" period, during which scans were scheduled at 6-month intervals, while scans were scheduled at 12-month intervals in the NHS. Table 5 summarizes the number of P-FAS subjects with evaluable CT data at each visit during this period.

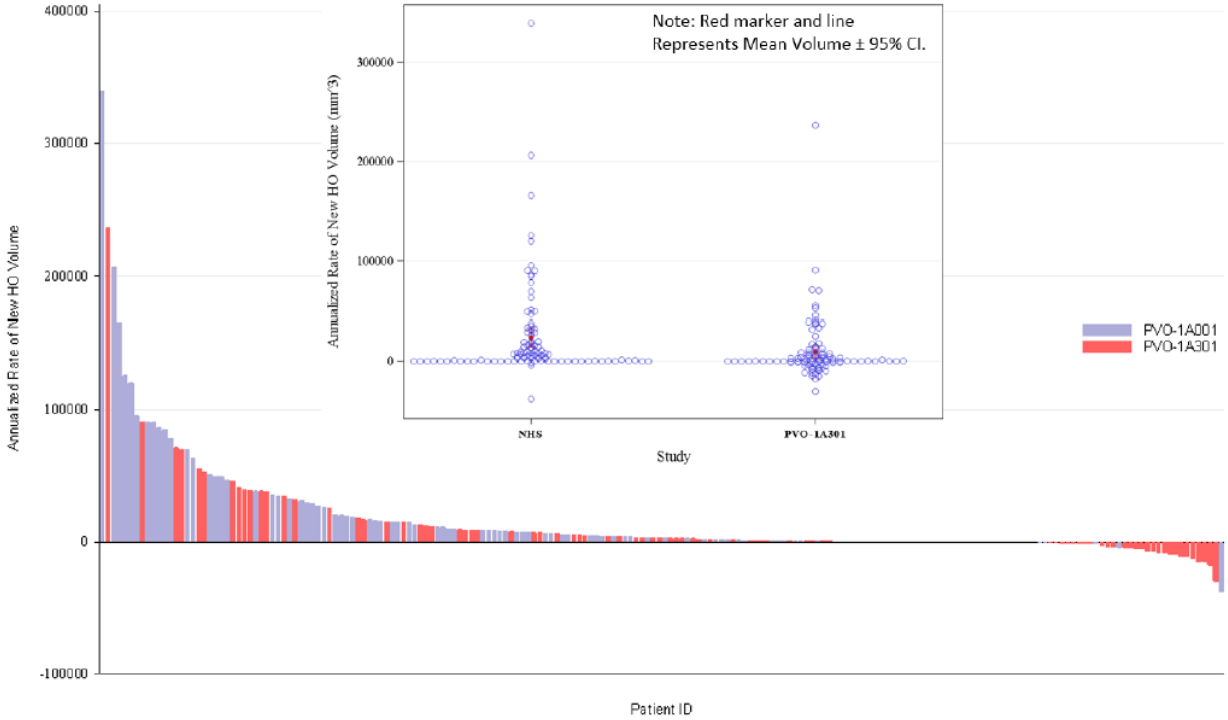
Table 5. Studies 301 and NHS: Number of Subjects With Whole Body HO Data in Pre-Pause Period by Visit (P-FAS)

Visit	Study 301 Palovarotene	NHS Control
Screening	97	101
Month 6	94	2
Month 12	93	90
Month 18	64	11
Month 24	1	63
Month 30	0	9
Month 36	0	33
Month 42	0	4
Month 48	0	0

Source: FDA reviewer; some subjects had some assessments outside the scheduled window
Data as of the following cutoff dates: in Study 301, 12/4/19 for <14 y/o and 1/24/20 for ≥14 y/o; in NHS, 2/28/20
Abbreviations: HO, heterotopic ossification; NHS, natural history study; P-FAS, principal full analysis set

Figure 1 shows the distribution of individual subject annualized new HO volume in the "Pre-pause" period of these studies, illustrating the variability of the data. On the left side of the plot, representing subjects with large HO increases, there is a preponderance of untreated subjects (in gray); 4 NHS subjects developed new HO at rates >100 cm^3/year, versus one palovarotene recipient in Study 301 (in red). On the right side, numerous subjects, mostly in Study 301, had net negative new HO.

Figure 1. Study 301 and NHS: Individual Subject Annualized New HO Volume Distribution for Palovarotene-Treated and Untreated Subjects (P-FAS)



Source: Study 301 Appendix 16.1.13.5 final IA3, Figure S14.1.1
Abbreviations: HO, heterotopic ossification; NHS, natural history study; P-FAS, principal full analysis set

Using the Applicant's pre-specified Bayesian analysis, the primary endpoint of annualized new HO failed to demonstrate efficacy of palovarotene with adequate posterior probability (Table 6). The posterior probability of a reduction in annualized new HO was 0.65. The 95 percent credible interval of the between study ratio of annualized new HO was (0.74, 1.22).

Table 6. Studies 301 and NHS: Ratio of Annualized New HO Volume

Study 301 Palovarotene	NHS Control	Median Ratio (95% CI)	Pr(Ratio <1)
N=97	N=101	0.95 (0.74, 1.22)	0.65

Source: CSR
Prespecified primary analysis method; change in HO values since previous scan were square root-transformed by body region and negative new HO volumes were set to 0. The median ratio is the median of the posterior distribution of the ratio (annualized rate of new HO for subjects in Study 301 divided by the annualized rate of New HO for subjects in NHS). Pr(ratio <1) is the probability that the ratio is less than 1, given the data (the posterior probability).
Abbreviations: CI, confidence interval; HO, heterotopic ossification; NHS, natural history study

Applicant's Post Hoc Bayesian Analyses

The Applicant's post hoc Bayesian analyses without square root transformation supported the efficacy with a high posterior probability of reduction (>0.99) in annualized new HO in Study 301 subjects compared with NHS subjects (Table 7). The Applicant's analyses with additional covariates yielded similar results.

Table 7. Studies 301 and NHS: Ratio of Annualized New HO Volume –Post Hoc Bayesian Analyses

Analysis Method	Median Ratio (95% CI)	Pr(Ratio <1)
No square root transformation	0.64 (0.45, 0.90)	>0.99
No square root transformation, added covariates[a]	0.58 (0.41, 0.82)	>0.999

Source: ISE
[a] Covariates: baseline HO divided by baseline age, baseline age, sex, baseline CAJIS, time since last flare-up.
Negative new HO set to zero.
The median ratio is the median of the posterior distr bution of the ratio (annualized rate of new HO for subjects in study PVO-1A-301 divided by the annualized rate of New HO for subjects in NHS). Pr(ratio <1) is the probability that the ratio is less than 1, given the data (the posterior probability).
Abbreviations: CI, confidence interval; HO, heterotopic ossification; NHS, natural history study

The Applicant also conducted a Bayesian analysis with square root transformation but adjusted for the visit schedule (that is, Study 301 visit schedule collapsed into a single Month 12 WBCT to match the timing of the first NHS post-baseline WBCT). The results were also in favor of Palovarotene, with a posterior probability of reduction of HO greater than 0.9.

Applicant's Post Hoc Weighted Linear Mixed Effect Model Analysis (wLME)

The Applicant's post hoc wLME analysis with baseline annualized HO (calculated as baseline total HO volume divided by baseline age) as the only covariate achieved nominal significance (Table 8). In this analysis, the annualized new HO for each subject was calculated using the change in total HO volume from baseline associated with the longest follow-up divided by the length of follow-up. The length of follow-up varied substantially among subjects and between studies. For example, annualized new HO was based on the Month 18 observation for 66% of subjects in Study 301 and the Month 36 observation for 32% of subjects in the NHS. The wLME model accounted for the various follow-up lengths using a weighted approach. We note that

the estimated treatment effect based on various lengths of follow-up could be difficult to interpret without the strong assumption of a constant rate of change in HO.

Table 8. Studies 301 and NHS: Applicant's wLME Analyses on Annualized New HO at Last Scan

Statistics	Study 301 Palovarotene N=97	NHS Control N=101
Annualized HO at baseline, mean (SD)	13.4 (13.1)	15.7 (14.0)
Mean annualized new HO, (SE)	9.4 (3.1)	23.7 (4.9)
Median (min, max)	0.4 (-30.2, 237)	6.1 (-37.9, 339)
Mean difference from NHS	-14.3	-
LS mean difference (95% CI)	-10.9 (-21.2, -0.6)	
Nominal p-value, comparison to NHS	0.039	-

Source: CSR

wLME analyses: subject level random effect included to account for the correlation among repeated measures on the same subject for those who transitioned from NHS to Study 301; No square root transformation, negative new HO included Covariates: annualized HO at baseline calculated by dividing baseline HO by baseline age; response variable- annualized change in HO at last visit.

Abbreviations: CI, confidence interval; HO, heterotopic ossification; LS, least square; NHS, natural history study; SD, standard deviation; SE, standard error; wLME, weighted linear mixed effects

Given that the distribution of annualized new HO is skewed, the Applicant performed sensitivity analyses to investigate the influence of large values on the analyses conclusion. These analyses indicate that the conclusions are not driven by the few extremely large values seen in subjects in the NHS (Table 20 and Table 21

Subgroup Summary by Sex and Age

Subgroup summaries indicate a smaller effect size among females and among subjects older than 18 years (Table 9). This finding may be explained by the smaller annualized growth of new HO in the control group (NHS) patients in these subgroups. However, it is not possible to rule out that the drug's effect is less in older patients for whom the disease stage is often more advanced. An explanation for the smaller effect in females is uncertain. Findings from these subgroup analyses should be interpreted with caution because with a smaller number of subjects in each subgroup, the influence of a few large values and imbalances in confounders in the absence of randomization could be greater on the mean difference, compared with analyses based on the overall population.

Table 9. Subgroup Summary on Annualized New HO (cm^3) Based on the Last Scan

Subgroup	Statistics	Study 301 (N=97)	NHS (N=101)	Difference
Sex				
Male	N	51	56	
	Mean (SD)	8.4 (22.7)	31.3 (55.3)	-22.9
Female	N	46	45	
	Mean (SD)	10.6 (37.3)	14.3 (37.7)	-3.7
Age				
>=18	N	22	40	
	Mean (SD)	8.6 (19.4)	10.7 (22.2)	-2.1
<18	N	75	61	
	Mean (SD)	9.7 (33.0)	32.3 (58.8)	-22.5

Source: FDA reviewer
Abbreviations: HO, heterotopic ossification; NHS, natural history study; SD, standard deviation

Secondary Efficacy Endpoint Results

The key secondary efficacy endpoint was the proportion of subjects with any new HO at month 12. There was no difference in the proportion of subjects with any new HO (64% in treated subjects, 62% in untreated subjects) at month 12. There was also no difference in the mean (SD) number of body regions with new HO at month 12 [1.3 (1.4) in treated subjects, 1.5 (1.6) in untreated subjects].

Exploratory endpoints included change from baseline in range of motion assessed by CAJIS at month 24, change from baseline in physical function assessed by FOP-FPQ at month 24, and change from baseline in physical and mental function in subjects 15 years and older as assessed by the PROMIS Global Health Scale at month 24. The results do not reveal any significant difference between treated and untreated subjects.

WBCT Data Following Treatment Pause – Study 301

The following table below summarizes WBCT data in Study 301, including available data following the treatment pause. The rate of new HO tended to increase from the pre-pause to dosing interruption period (mean 9.4 to 20.1 cm^3/year). In a subgroup of 17 subjects who subsequently restarted treatment, new HO volume was relatively low (7.7 cm^3/year), similar to an analogous subgroup in Study 202C. In another subgroup of 16 subjects who did not restart treatment and had available off-treatment data, mean new HO volume in this phase was 15.6 cm^3/year, representing an increase from a mean of 2.3 cm^3/year in these 16 subjects during the pre-pause Treatment period.

Table 10. Study 301: Annualized New HO Volume by Treatment Phase

Treatment Phase	Pre-Pause Treatment	Interruption	Post-Pause Treatment
N	97	42	17
Time interval (mean months)[a]	15.7	25.4	14.4
Annual new HO volume (cm^3/year), mean	9.4	20.1	7.7

Source: Table 2 in Response to CRL, 02/16/23 resubmission M 1.11.4
[a] Duration between first and last WBCT scan during the study/phase
Annualized new HO data represents change in volume between first and last WBCT scan in the phase, divided by time interval
Abbreviations: HO, heterotopic ossification

The following table below summarizes WBCT data in Study 301 over the pre-pause treatment period, the dosing interruption period, and the post-pause treatment period for subjects who contributed WBCT data in all three periods. As with the previous table above, rate of new HO tended to increase from the pre-pause treatment period to the dosing interruption period, and then decrease during the post-pause treatment period.

Table 11. Study 301: Annualized New HO Volume by Treatment Phase (Including Only Subjects With Data for All Three Phases)

	Study 301		
Treatment Phase	**Pre-Pause Treatment**	**Interruption**	**Post-Pause Treatment**
N	17	17	17
Annual new HO volume (cm^3/year), mean	5.0	29.8	7.7

Source: FDA reviewer
Annualized new HO data represents change in volume between first and last WBCT scan in the phase, divided by time interval
Abbreviations: HO, heterotopic ossification

3.1.2 Efficacy Issues in Detail

3.1.2.1 Issue 1: Appropriateness of Reliance on Post Hoc Analyses to Support Effectiveness

General Consideration

Post hoc analyses are generally considered hypotheses generating, as they could inflate type-I error and raise concerns of potential bias caused by selection of analyses toward the intended favorable outcome. However, in this specific context, FDA acknowledges that the prespecified primary analysis may not have been the appropriate method for analyzing the primary endpoint for the reasons outlined below. Acknowledging the limitations of the prespecified analyses, we think it is reasonable to consider alternative more appropriate analyses to assess evidence of efficacy.

Assessment of the Impact of Square Root Transformation

The primary endpoint was annualized new HO. However, the rate of growth in HO volume was estimated based on the square root transformation of each incremental change in HO by region between scans. Since WBCT scans were not performed at the same interval in the two studies (scheduled every 6 months in Study 301 and every 12 months in the NHS), the prespecified analysis involves summing the square root of each incremental change to compare the square root of the sum of each change, which is problematic as they are not equivalent. This could have resulted in a bias against palovarotene due to the differential schedule of new HO assessments between Study 301 and the NHS, which might have contributed to the failure of the primary analysis.

Table 12 illustrates how more frequent assessment could lead to a bias against palovarotene under the prespecified Bayesian analysis using the square root transformation. Assuming a new HO volume of 200 occurs in a body region before Month 6 and subsequently the region grows 200 more in HO volume during Month 6 to Month 12, the annual new HO will be 400. With the

prespecified statistical model, it will be counted as 2 events with growth rate of √200 per event under the square root transformation in Study 301 and 1 event with growth rate of √400 in the NHS. Therefore, the estimated annual new HO will be 2*√200=28.3 in Study 301 and √400=20 in the NHS, respectively. This example illustrates that the prespecified approach may lead to a nonsensible difference in the estimated annualized new HO and a bias against more frequent assessment.

Table 12. An Illustrative Example of Impact of Square Root Transformation on Annualized New HO

Study	Month 6 Change	Month 12 Change From Month 6	Total Change	Estimated Annual New HO[a]
301	200	200	400	$2*\sqrt{200}=28.3$
NHS	Not assessed	400	400	$\sqrt{400}=20$

Source: FDA reviewer
[a] Estimated annual new HO is annual number of events times the average amount of new HO per event. Each occurrence of a new HO is counted as an event.
Abbreviations: HO, heterotopic ossification; NHS, natural history study

FDA's Assessment of Post Hoc Bayesian Analysis Without Square Root Transformation

The Applicant's post hoc Bayesian analysis without square root transformation resulted in a high posterior probability in support of efficacy. We note that inference from the Bayesian model could be affected by different specification of prior distributions of the multiple parameters in the model and the results could differ if influential priors are specified. FDA's analyses will focus on frequentist methods that directly analyze the annualized new HO of each subject with minimal assumptions.

FDA's Assessment of the Applicant's Post Hoc wLME Analyses

As noted previously, the Applicant's post hoc wLME analysis (Table 8), based on the CT scan associated with the longest follow-up and without square root transformation, achieved nominal statistical significance. However, results from the Applicant's wLME could be difficult to interpret as the annualized new HO for each subject was based on various length of follow-up. As most subjects had WBCT assessments at Month 12 (Table 5), we compared the changes in HO volume between Study 301 and NHS subjects during a 12-month observation period as a landmark analysis, which does not rely on assumptions of a constant rate of change. Some subjects did not have changes in HO through a 12-month period (4 subjects in Study 301 and 11 subjects in NHS). Depending on how subjects with no 12-month HO data could be handled, there are different ways for performing the landmark analysis.

We performed the landmark analyses using different methods for subjects who did not have 12-month data to investigate the consistency of findings. For all these analyses, a robust sandwich variance estimator was used to account for heteroscedasticity in variance. Results from these landmark analyses with different ways of handling subjects without 12-month data and including different covariates were consistently in favor of palovarotene (Table 13). In addition, our analyses did not transform negative new HO values to zero.

Table 13. FDA's Landmark Analyses Without Square Root Transformation

Analysis Method[a]/Least Square Mean (SE)	Study 301 (N=97)	NHS (N=101)	Difference (95% CI)	Nominal p-Value
Include only subjects who provided Month 12 HO	93	90		
Covariates 1[b]	6.7 (2.4)	22.0 (6.4)	-15.3 (-28.5, -2.1)	0.0229
Covariates 2[b]	6.2 (2.4)	21.8 (6.1)	-15.5 (-28.8, -2.3)	0.0212
Covariates 3[b]	6.6 (2.5)	22.0 (6.2)	-15.4 (-28.8, -2.0)	0.0243
Include only subjects with HO data up to Month 12	97	92		
Covariates 1	8.1 (2.8)	22.5 (6.3)	-14.3 (-27.8, -0.9)	0.0368
Covariates 2	7.6 (2.8)	22.3 (6.1)	-14.7 (-28.2, -1.3)	0.0321
Covariates 3	8.1 (2.9)	22.7 (6.2)	-14.6 (-28.2, -1.0)	0.0354
Include all subjects using Month 12 data if available[c]	97	101		
Covariates 1	8.1 (2.8)	20.7 (5.3)	-12.5 (-24.0, -1.1)	0.0317
Covariates 2	7.5 (2.9)	21.0 (5.2)	-13.5 (-25.3, -1.7)	0.0249
Covariates 3	7.8 (2.9)	21.7 (5.5)	-13.9 (-26.3, -1.5)	0.0275
Include all subjects with Month 12 and multiple imputation for missing Month 12; with covariates 3	8.3 (3.0)	22.7 (6.3)	-14.4 (-28.1, -0.7)	0.0395

Source: FDA reviewer
[a] Generalized estimating equation, no square root transformation, negative new HO included
[b] Covariates 1: baseline rate of HO: baseline total HO divided by baseline age; covariates 2: baseline rate of HO, sex, baseline age; covariates 3: baseline rate of HO, sex, baseline age, baseline CAJIS, time since last flare up
[c] Used last HO changes at month 12 visit window if available; otherwise, used measurements closest to the 12-Month visit window. Conclusion from the multiple imputation-based approach will be overturned when a penalty greater than 45 cm^3 /year is added to the imputed values in treated subjects.
Abbreviations: CI, confidence interval; HO, heterotopic ossification; NHS, natural history study; SE, standard error

Compared with the treatment effect based on observations associated with the longest follow-up (Table 8), the estimated effects from these landmark analyses are generally larger but also have greater variability.

Conclusion

The prespecified Bayesian compound Poisson model incorporating square root transformation appears to have a bias against palovarotene due to the more frequent assessment in Study 301. Post-hoc analyses without the square root transformation conducted by either the Applicant or FDA support that the annualized new HO of Study 301 subjects was lower than that of NHS subjects. These analyses used regression adjustments to account for baseline differences between subjects in the two studies. Analyses methods based on matching and weighting to account for the use of an external control will be described below.

3.1.2.2 Issue 2: Use of an External Control Group (NHS) for Evaluation of Chronic/Flare-Up Regimen

General Consideration

Despite the limitations associated with using an external control rather than a randomized concurrent control, FDA recognizes that support for effectiveness can emerge using an externally controlled trial when the following are met (FDA, 2019):

(1) the natural history of the disease is well defined,
(2) the external control population is very similar to that of the treatment group,

(3) concomitant treatments that affect the primary endpoint are not substantially different between the external control and the trial population, and
(4) the results provide compelling evidence of a change in the established progression of disease.

The concomitant treatments in this population are not expected to affect the outcome of interest. The NHS study was performed to characterize the disease process and outcome. The focus here is whether the difference between the study populations in the NHS and Study 301 impacted the conclusion regarding efficacy and whether the results are compelling enough to overcome the general concerns of potential confounding factors and impacts from potential selection bias due to differential loss to follow-up between the NHS and Study 301.

FDA's Assessment of Comparability of Study Populations and Potential Selection Bias

The enrollment criteria for the NHS and Study 301 were generally similar. Because the NHS was restricted to subjects with the classic R206H mutation, which represents approximately 97 percent of FOP patients, the P-FAS population of Study 301 excluded a small number of participants with other mutations.

Recruitment of subjects into the NHS and Study 301 was sequential, as NHS enrollment was complete before Study 301 was initiated. All NHS study sites also participated in Study 301, and any NHS subject who met enrollment criteria (e.g., without retinoid-safety-related exclusions) was eligible for this transition. A total of 39 subjects transferred from the NHS into Study 301 and contributed efficacy data to both studies. In addition, eight subjects discontinued the NHS to enroll in Study 201 at the onset of a flare-up, 13 subjects discontinued the NHS to enroll in Study 202B, and 6 subjects discontinued the NHS to enroll in other interventional studies, including for other investigational products for FOP. Given that transfers out of the NHS were voluntary and driven in part by clinical factors which are largely unknown, unmeasured confounding is likely to have occurred. Further, selection bias due to differential loss to follow-up may be of concern if transferring from the NHS to interventional studies was related to disease progression or other outcome-related factors, whereas loss from Study 301 may not be similarly related to such factors, resulting in potential differential capture of outcome between the two studies.

Table 14 presents baseline data for three cohorts within the Principal FAS: the 39 transition subjects who contributed efficacy data to both the NHS and Study 301; 62 NHS subjects who did not enroll in Study 301; and 58 subjects in Study 301 who had not participated in the NHS. Transition subjects' data at both NHS baseline and Study 301 baseline (at an average of 2.3 years later) are presented.

As shown, NHS subjects who remained in the NHS were older (mean age of 20.5 years) than subjects who transitioned to Study 301 (mean age of 13.5 years) and other Study 301 subjects (mean age of 14.6 years). Consistent with older age, the NHS-only subjects tended to have more advanced disease, with generally higher whole-body HO volumes, greater number of

body regions with HO, and higher scores on CAJIS, indicating more severely impaired range of motion and physical function. Transition subjects reported a higher number of flare-ups prior to NHS enrollment (mean 5.8 within 12 months) than the other groups; however, this observation was driven by two subjects.

Table 14. NHS, Study 301, and Transition Subjects: Baseline Data (Principal FAS)

		Transition Subjects		Study 301
Statistics	NHS Only N=62	NHS Data N=39	Study 301 Data[a] N=39	Only N=58
Mean age, year	20.5	13.5	15.8	14.6
Sex, male/female (%)	52/48	62/38	62/38	47/53
Race, White/Asian/Other (%)	77/6/17	67/10/23	67/10/23	72/9/19
WBHO volume, cm^3, mean	378.2	207.9	259.2	212.4
Number of body regions with HO, mean	6.8	6.1	6.4	6.2
Number of flare-ups within 12 months of enrollment, mean	2.4	5.8	1.1	1.7
CAJIS score, mean	13.1	9.6	10.4	9.4

Source: reviewer's analysis.
[a] Transition subjects' baseline in NHS and baseline in Study 301 were summarized respectively.
Abbreviations: CAJIS, Cumulative Analogue Involvement Scale; FAS, full analysis set; FOP, fibrodysplasia ossificans progressiva; HO, heterotopic ossification; NHS, natural history study; WBHO. whole body heterotopic ossification.

Assessment of Impact of Subject Difference at Baseline

To assess the impact of subject difference between Study 301 and the NHS, the Applicant conducted sensitivity analyses by adjusting for additional baseline covariates, including the propensity score quartile, and performing matching and weighting based on propensity score among subjects who did not enroll in both the NHS and Study 301 (i.e., non-transition subjects, Table 15). Results from these analyses did not change the conclusion of the wLME analysis (Table 8).

Table 15. Propensity Score Matching and Weighting Analysis - Non-Transition Subjects

Analysis Method/Mean	Study 301 (N=58)	NHS (N=61)	Difference	Nominal p-Value
Propensity score matching[a] (39 matched)	5.6	24.1	-18.5	<0.05
Unstabilized propensity score weighting	9.5	28.9	-19.4	<0.05
Stabilized propensity weighting	9.5	28.9	-19.4	<0.05

Source: Applicant's matched analysis report.
[a] Nearest neighbor, caliper =0.2, covariates (specified by Agency): baseline rate of HO volume, baseline age, sex, months since last flare up, and baseline CAJIS.
Abbreviations: NHS, natural history study

The FDA review team performed landmark analyses on the P-FAS population using propensity score weighting, matching, and targeted maximum likelihood estimation methods that are used for causal inference for non-randomized studies. Results from these analyses are also consistent (Table 16). Baseline characteristics achieved reasonable balance after weighting or matching (Table 22 and Table 23). However, validity of conclusions on treatment effects from these analyses relies on the unverifiable assumption that there are no unknown or unmeasured confounding factors that could impact the difference between Study 301 and NHS subjects. The

majority of the subjects in these analyses completed the 12-month follow-up and thus loss to follow-up was less of a concern.

The propensity score-based weighting and matching analyses assume that crossover of the NHS subjects to Study 301 was not due to factors related to disease progression. The within-subject correlation was accounted for in the analysis after matching.

Table 16. FDA's Landmark Analyses Without Square Root Transformation Based on Methods for Causal Inference

Analysis Method	Study 301 (N=97)	NHS (N=101)	Difference (95% CI)	Nominal p-Value
Propensity score weighting/ mean (SE)				
Include all subjects using 12-Month data if available	97	101		
Covariates 1[a]	8.9 (3.1)	22.8 (5.9)	-14.0 (-27.0, -0.8)	0.0364
Covariates 2[a]	8.7 (3.0)	22.9 (5.9)	-14.2 (-27.2, -1.2)	0.0321
Covariates 3[a] (excluded 4 subjects missing covariates)	9.1 (3.2)	23.0 (6.1)	-13.9 (-27.4, -0.5)	0.0422
Propensity score matching[b]				
Exact match on sex and age group, nearest neighbor with caliper 0.2 for covariates 3	Matched n=61	Matched n=61	-16.2 (-25.7, -6.6)	0.0022
Targeted maximum likelihood estimation				
Covariates 1	n=97	n=101	-14.3 (-27.3, -1.3)	0.0309
Covariates 3 (excluded 4 subjects missing covariates)	n=97	n=97	-15.2 (-30.7, 0.03)[c]	0.0551
Covariates 3; including all 12-month changes	n=97	n=101	-11.6 (-21.8, -1.2)	0.0281

Source: FDA reviewer
[a] Covariates 1: baseline rate of HO, sex, baseline age; covariates 2: baseline rate of HO, sex, baseline age, baseline CAJIS; covariates 3: baseline rate of HO, sex, baseline age, baseline CAJIS, time since last flare up.
[b] Age group: (1) <12 years; (2) 12 to <18 years; (3) ≥18 years
[c] bootstrap 95% CI: (-31.2, -0.5)
Abbreviations: CI, confidence interval; NHS, natural history study; SE, standard error

Assessment of Analyses Based on Subjects Transitioned From NHS to Study 301

The Applicant performed analyses limited to subjects who transitioned from the NHS to Study 301 (i.e., baseline control). One of the Applicant's analyses compared the observations associated with the longest follow-up within each study. The other analysis compared the last 12-month change of HO in the NHS with the first 12-month change in Study 301. Results from the two analyses favored palovarotene (Table 17).

Table 17. Applicant's Analyses on Annualized New HO (cm^3) in Subjects Transitioned to Study 301

Analysis Method/Least Square Mean (SE)	Study 301 (N=39)	NHS (N=39)	Difference (95% CI)	Nominal p-Value
wLME* using the observation associated with the longest follow-up	8.1 (4.0)	16.7 (3.3)	-8.6 (-17.7, 0.5)	0.0634
wLME using last 12-month change in NHS and first 12-Month change in Study 301[#]	4.9 (4.9)	18.2 (5.2)	-13.2 (-25.7, -0.8)	0.0377

Source: CSR and response to information request
Analysis covariate: baseline rate of HO; random subject effect.
[#]For subjects who did not have 12-month observation, the last two measures in NHS or the first two measures in Study 301 were used
Abbreviations: NHS, natural history study; SE, standard error; *wLME, weighted linear mixed effect model

These within-subject comparisons complemented a between-subject comparison by using subjects' own observations in the NHS as the control. However, interpretation of results from this analysis should consider to what extent the disease would be expected to progress on average without treatment (similarly to what was observed in the NHS) after these subjects transitioned to Study 301. The impact of disease progression when enrolled into Study 301 could not be fully assessed without a concurrent control.

Nevertheless, the observed average within-subject increase in the rate of HO after the dosing interruption and the subsequent decrease in rate of HO after restarting palovarotene treatment (as shown in Table 1, Table 2, Table 10, and Table 11) in Study 202C and 301 subjects is also supportive to the conclusions from the wLME analysis.

Sensitivity Analysis to Address Potential Selection Bias

Propensity score weighting or matching considering baseline characteristics does not address the concern that loss to follow-up when transferring from the NHS to interventional studies may have been related to disease progression (or other outcome-related factors), whereas loss from Study 301 was minimal or was not similarly related to such factors. The implicit characteristics of NHS subjects that drove self-selection into interventional studies may have been unknown or uncaptured at study baseline. Even if captured, the confounding control methods are inadequate to address potential differential loss to follow-up (i.e., selection bias). To address this specific concern, the Applicant conducted sensitivity analyses considering palovarotene treatment a time-varying exposure and HO volume by WBCT scan a time-dependent confounder. The sensitivity analysis results were consistent with the wLME analysis in terms of the direction of association between the palovarotene treatment and the annualized new HO. Of note, specification of such a causal inference model required several strong and untestable statistical assumptions, and the post-hoc analysis was not powered to detect a clinically meaningful effect.

Conclusion

When an external control is used, there will be a lack of assurance of comparability between treated and untreated subjects given the lack of randomization. There were some differences in baseline characteristics between the NHS and Study 301, such as subjects in the NHS were

relatively older and of worse disease severity. To address the potential impact of measured confounding factors, various analyses using methods that are commonly used for causal inference were performed and these analyses provided consistent results. It appears that the difference in annualized new HO between treated and untreated subjects was not driven by any systemic difference in baseline covariates that are expected to be clinically important, whereas the impact of unknown confounding factors is uncertain. In addition, the Applicant's sensitivity analysis to assess the impact of selection bias provides limited support on the robustness of study conclusions, as these analyses require strong and untestable statistical assumptions.

3.2 Safety Issues

Palovarotene is a member of the retinoid class of medications and has the risks associated with the class. Boxed Warnings and Warnings and Precautions contained in the retinoid class product labels include teratogenicity; psychiatric disorders; pseudotumor cerebri; serious skin reactions; pancreatitis; lipid abnormalities (hypertriglyceridemia); hearing impairment; hepatotoxicity; inflammatory bowel disease; skeletal abnormalities, including decreased bone mineral density; hyperostosis and premature epiphyseal closure; and vision impairment, including corneal opacities and decreased night vision.

- Key Safety Issue: Apparent increased incidence of flare-ups

3.2.1 Sources of Data for Safety

Safety data are presented for the placebo-controlled Study 201 as well as analyses for the proposed dosing regimen of 5 mg palovarotene daily with a 20/10 mg dose increase for flare-up symptoms from Study 202 (B and C) and Study 301. The pooled safety data from studies 202 (B and C) and 301 were analyzed including all subjects who met the indicated age cut-off of 8 years for girls and 10 years for boys.

3.2.2 Safety Summary

The palovarotene studies in patients with FOP provide limited safety data, given the small number of FOP patients available for study. However, a large safety database, including prior chronic obstructive pulmonary disease studies of palovarotene (600 subjects with a maximum 5 mg daily dose), pediatric studies in multiple osteochondroma (n=160, 5 mg palovarotene), and extensive clinical experience with other retinoids, supplement the FOP clinical data and help to provide an adequate understanding of the safety of palovarotene treatment of this disease.

Study 201

In Study 201, 40 subjects with flare-up symptoms were enrolled and received placebo for 6 weeks, palovarotene 10 mg for 2 weeks followed by 5 mg for 4 weeks, or palovarotene 5 mg for two weeks followed by 2.5 mg for 4 weeks. All subjects completed the trial. The mean exposure to placebo was 42 days and 41.5 days (range 38-42 days) for palovarotene. No deaths occurred in the study. Serious adverse events occurred in 4 subjects. One subject in the placebo group had an asthmatic crisis that resolved with treatment. One subjects in the palovarotene 5/2.5

mg group suffered a hemorrhagic ovarian cyst and two subjects in the palovarotene 10/5 mg group had 'condition aggravated.' Condition aggravated included new FOP flare-ups that occurred during treatment. No subjects discontinued treatment or required dose reduction during the study. One subject in the palovarotene 10/5 mg group developed increased lipase that required dose interruption on study day 15. Study drug was restarted on study day 30 and lipase was again elevated to 330 U/L. Study drug dosing was stopped on study day 39. Lipase level returned to normal. Overall, 100% of subjects experienced at least one adverse event during the study. The most common adverse events reported at a higher rate in the palovarotene-treated groups when compared to the placebo group were dry skin (30% placebo, 56% palovarotene 5/2.5mg, 81% palovarotene 10/5 mg), condition aggravated (30% placebo, 22% palovarotene 5/2.5mg, 62% palovarotene 10/5 mg), and dry lip (10% placebo, 56% palovarotene 5/2.5mg, 33% palovarotene 10/5 mg).

Combined Study 202 (Parts B and C) and Study 301

In this combined safety database, 139 FOP subjects over the age of 8 years for girls and 10 years for boys are included. In this target population, the mean duration of exposure was 79 ± 50.4 weeks for the 5 mg daily chronic dose (N=131) and 35 ± 25.8 weeks for the 20/10 mg flare-up dosing (N=105).

Deaths

No deaths occurred in the FOP palovarotene treatment studies; however, one death did occur in a 13-year-old subject with a history of restrictive lung disease who died 2.5 months after discontinuing palovarotene treatment. Additionally, a 38-year-old subject in the NHS with substantial disease burden based on the CAJIS scale died of a cardiac arrest. This subject never received treatment with palovarotene.

Serious Adverse Events

Serious adverse events were reported in 50 (36%) subjects. Many events occurred in a single subject. The most common serious adverse events reported include premature epiphyseal closure in 9 (6%) subjects, coronavirus infection in 4 (3%) subjects, pneumonia in 4 (3%) subjects, condition aggravated in 3 (2%) subjects and cellulitis in 3 (2%) subjects.

Adverse Events Leading to Study Discontinuation

Four (3%) subjects discontinued palovarotene studies due to adverse events, all from Study 301. Two subjects were withdrawn for premature epiphyseal closure, one for intentional self-injury and one for malnutrition.

Adverse Events leading to Study Drug Dose Modification, Interruption or Discontinuation

Overall, 50 (36%) subjects had adverse events leading to palovarotene dose modification, 37 (27%) subjects had adverse events leading to palovarotene dose interruption and 11 (8%) subjects had adverse events leading to palovarotene discontinuation. In Study 301, 9 subjects

discontinued study drug. Four subjects stopped study drug due to premature epiphyseal closure and one each for pruritis, furuncle, mobility decreased, dry skin and intentional self-harm. In Study 301, adverse events leading to dose modification occurred in 34 (34%) subjects and more frequently when subjects were on treatment with the higher flare dose 20/10 mg regimen (28 (40%) subjects compared to 11 (11%) subjects on the 5 mg chronic daily dose regimen). The most common reason for dose modification was drug eruption and dry skin.

Adverse Events

Adverse events were reported in 100% of study subjects receiving palovarotene. The most common adverse events were dry skin (81%), dry lip (57%), arthralgia (49%), pain in extremity (42%), alopecia (42%) pruritis (42%), erythema (34%), rash (33%), headache (29%), and dry eye (27%). All of these events have been seen with other retinoids and are considered adverse reactions to the drug.

3.2.3 Safety Issues in Detail

3.2.3.1 Issue 1: Numeric increase in Incidence of Flare-Ups

In each of the NHS and palovarotene studies, subjects were instructed to promptly report potential FOP flare-up symptoms so that these events could be assessed by an investigator. In the phase 3 study (301), flare-up incidence and rate were designated as secondary efficacy endpoints to determine whether treatment may suppress the initiation of new flare-ups in addition to suppressing formation of new HO. However, an unexpected numerical increase in flareups was observed. Furthermore, retinoids (and vitamin A in toxic doses) have been associated with hyperostosis and calcification of ligaments and tendons; musculoskeletal adverse effects including back pain, arthralgia, myalgia; and rare reports of severe myositis, sometimes associated with strenuous physical activity. Because FOP flare-ups are frequently triggered by local inflammation or trauma and are characterized by ossification of muscle and other soft tissues, any palovarotene associated increase in flareups raises concern for worsening of HO.

Flare-Up Incidence During Treatment of an Index Flare in Studies 201/202

In early phase 2 studies, during the 12-week period of treatment/follow-up of an index flare-up, symptoms suggestive of a new flare-up at a different site were recorded as adverse events of ‘condition aggravated’ and were not specifically treated. In Study 201, such adverse events were reported by more subjects randomized to the palovarotene 10/5 mg regimen (13/21, 62%) in comparison to the 5/2.5 mg regimen (2/9, 22%) or placebo (3/10, 30%). There was no consistent pattern of adverse event onset that would suggest a “rebound” effect following discontinuation of palovarotene at 6 weeks.

These events were also evaluated by regimen across the NHS and phase 2 studies 201, 202A and 202B. Within the 12 weeks following onset of an index flare-up, new flare-ups (i.e., condition aggravated adverse events) were reported in 23% of untreated NHS flare-ups; 30% of flare-ups treated with placebo; 22% of flare-ups treated with palovarotene 5/2.5 mg; 37% of

flare-ups treated with palovarotene 10/5 mg; and 36% of flare-ups treated with palovarotene 20/10 mg.

Flare-Up Incidence and Rate During Chronic/Flare-Up Treatment in the NHS and Study 301

In Study 301, flare-ups were more broadly defined, to include index flare-ups and new events occurring during the flare-up dosing cycle (intercurrent flare-ups at a different site, or marked worsening at the original site), which re-started the 12-week flare-up treatment cycle. At least one flare-up was reported by 67% of subjects in Study 301 and 56% of subjects in the NHS through Month 12. Among subjects with ≥1 flare-up event, the number of reported events per subject was greater in Study 301 (mean 4.2, range 1-23) vs. the NHS (mean 2.5, range 1-9). Therefore, among all patients (with or without flare-ups), the rate of flare-ups per month of study participation was greater in Study 301 compared to the NHS (0.15 vs. 0.07 per subject-month, respectively). Data were similar for the subset of 39 subjects who crossed over from the NHS to Study 301.

The differences in flare-up rate between Study 301 and NHS may be explained, at least in part, by differences in study design and conduct. Although these two studies were conducted mostly at the same institutions, and the use of adjunctive care (e.g., corticosteroids) was similar, NHS subjects generally had less frequent contact with study staff compared to subjects in Study 301. Phone contacts were primarily every 6 months in the NHS, and every 3 months in Study 301. In addition, NHS subjects may have been less motivated to report symptoms that were not intended for any treatment. In contrast to Study 301, subjects in the NHS did not keep symptom diaries, and worsening of an existing flare-up was not captured as a new event. A published survey of untreated FOP patients has reported an average rate of 0.16 flare-ups per month, potentially supporting that these events may have been under-reported in the NHS. In Study 301, some subjects experienced multiple flare-ups in succession, requiring a prolonged cycle of flare-up dosing, due to intercurrent symptoms at a new location (58% of events) or worsening of an existing event (42%). Overall, in this study, the rate of reported new flare-ups occurring during chronic dosing (i.e., the first or index flare-up of a new cycle) was 0.12 events/month, and during flare-up dosing (any subsequent flare-ups within a cycle) was 0.33 events/month. This finding may be related to anecdotal reports that FOP flare-ups may occur in clusters or in rapid sequence.

The clinical significance of the number of flare-ups in these studies with regard to disease progression is unclear. Subjects reporting ≥1 flare-up, compared to subjects reporting no flare-ups, developed larger quantities of new whole-body HO in the NHS (mean 38.1 vs. 6.2 cm^3), and also in Study 301 (mean 11.7 vs. 3.2 cm^3). In both studies, there was a modest positive correlation of flare-up rate and annualized new HO volume, as expected. Among all subjects (with or without flare-ups), however, substantial proportions of the total volume of new HO developed away from flare-up sites.

Based on current evidence, it is uncertain whether palovarotene, particularly at higher doses, may trigger inflammatory processes leading to new flare-ups. Alternatively, musculoskeletal adverse events related to retinoid effects could be misinterpreted as flare-up symptoms, and rates of symptom reporting in the clinical studies may have been influenced by study design. Another potential factor is that flare-up symptoms may rebound following withdrawal of the high dose corticosteroids that are administered during most flare-ups, as has been documented in published surveys.

3.2.3.2 Issue 2: Teratogenicity

Embryo-fetal toxicity has been reported with multiple retinoids and may result in irreversible fetal malformations, adverse pregnancy outcomes and stillbirths. The labeling of all approved retinoids (isotretinoin, acitretin, tretinoin, bexarotene) includes boxed warnings for pregnancy and teratogenicity. Embryo-fetal toxicity consistent with retinoid effects was observed in fetal rat studies of palovarotene.

There are only rare reports of pregnancy in women with FOP. FOP poses life-threatening risks for both the mother and fetus, including cardio-respiratory complications, thromboembolism, delivery complications, fetal distress, and premature birth. Consensus guidelines for FOP advise of these life-threatening risks and urge serious consideration and family planning before attempting pregnancy. As a result of the high-risk nature of pregnancy to both mother and fetus, reports of miscarriage are common amongst the small number of FOP pregnancies documented in case reports. The risk of embryo-fetal toxicity from palovarotene would add to the incidence of these events in the population likely to use the drug. As a precaution, clinical studies of palovarotene have included frequent pregnancy testing and contraception requirements. To date, there are no reports of pregnancy or in utero exposure to palovarotene.

In men of reproductive potential, a phase 1 study conducted to evaluate the potential exposure of palovarotene through semen to a female partner showed that palovarotene treatment in male subjects resulted in very low palovarotene levels in semen.

3.2.3.3 Issue 3: Premature Epiphyseal Closure

Systemic retinoids, as well as vitamin A in toxic doses, have long been associated with a variety of adverse skeletal effects. Among these are premature epiphyseal closure (or premature physeal closure) and impaired growth in children with prolonged exposure to retinoids. Growth plate abnormalities and premature closure were also observed in juvenile rats exposed to palovarotene and were consistent with retinoid class effects. Therefore, a bone safety monitoring plan was developed for pediatric patients (<18 years) enrolled in the palovarotene studies that included hand/wrist and knee radiographs to determine bone age at the time of enrollment. Bone age was determined using the Greulich and Pyle atlas (Pyle SI, 1959).

For growth plate analyses, hand/wrist radiographs (PA view, left side preferably) and knee radiographs (AP view) were performed at baseline and every 6 months as well as every 3 months when subjects received flare-up dosing. All radiographs were read by two independent musculoskeletal radiologists at a central imaging laboratory using standardized procedures.

WBCT scans, performed for assessment of the volume of total body HO, were also used to assess for potential adverse growth plate effects.

Linear growth was measured in triplicate using calibrated stadiometers. Because linear height measurements can be challenging in subjects affected by FOP, measurement of knee height using a standard caliper was performed every 6 months. In addition, femur and tibia length were measured on WBCT scans to monitor growth.

In studies 202 Part B, 202 Part C and 301, premature epiphyseal closure was reported in 14 of 25 (56%) subjects under the age of 8 years (girls) and 10 years (boys); 13 of 39 (33%) subjects between the 8/10 cutoffs and 14 years, and in 0/38 subjects between 14 and 18 years. Five of the affected subjects had not had any flare-up dosing, and no safe level of palovarotene exposure could be identified. Small changes were seen in height Z-score that were similar to data from the NHS, but these data were not sufficient to assess long-term effects on growth.

In summary, treatment with palovarotene causes premature growth plate closure in subjects with open growth plates. This occurred despite the weight-based palovarotene dosing in younger patients. The Applicant proposes an indication that allows patients with at least approximately 80 percent skeletal maturity (bone age 8 years in girls and 10 years in boys) to receive palovarotene for the prevention of heterotopic ossification in patients with FOP, despite the risk of premature epiphyseal closure, given the severity of the disease.

3.2.3.4 Issue 4: Psychiatric Disorders

Treatment with retinoids have been associated with depression, psychosis, and suicidal ideation. Psychiatric adverse events were reported overall in 25 (18%) subjects with preferred terms depressed mood in 12 (9%) subjects, suicidal ideation in 6 (4%) subjects, depression in 6 (4%) subjects, mood altered, mood swings, and memory impairment in 2 (1%) subjects each, and drug dependence, intentional self-injury, and disturbance in attention in 1 subject each.

The palovarotene intervention studies and the NHS also included assessments using the Columbia-Suicide Severity Rating Scale (C-SSRS). In Study 202, four subjects had suicidal ideation. All were C-SSRS Type 1 - suicidal ideation is passive suicidal ideation ("wish to be dead"). No study or drug discontinuations occurred. In Study 301, no subjects had suicidal ideation at baseline and 3 subjects expressed suicidal ideation, one with C-SSRS Type 1 and two with C-SSRS Type 2 - Non-specific suicidal thought. In the NHS, four subjects expressed suicidal ideation at baseline and five subjects expressed suicidal ideation through the course of the study, four with C-SSRS Type 2 and one with C-SSRS Type 3 - Active suicidal ideation without plan or intent to act. Overall, suicidal ideation was seen in the NHS as well as in the interventional trials.

3.2.3.5 Issue 5: Effects on Bone Mineral Content, Bone Mineral Density and Fractures

There are literature reports to suggest that retinoids and vitamin A toxicity may cause bone demineralization, osteoporosis, or 'slender long bones' in children. Labeling for isotretinoin for the treatment of acne includes a warning for negative effects on bone mineral density (BMD)

seen in a clinical trial. Assessments of BMD and bone mineral content (BMC) were not included in the palovarotene FOP studies because of the likely confounding effects of widespread HO in overlying soft tissues. However, because of a signal of decreased BMC seen in a DXA study of palovarotene in patients with multiple osteochondroma, an assessment of vertebral structural integrity was done using novel vertebral biomechanical computed tomography (biomechanical CT) analysis and vertebral fracture assessment using the WBCT scans obtained in Study 301 and the NHS. It should be noted that these measures have not been validated by FDA for assessment of fractures or estimation of fracture risk in clinical trials.

In Study 301 and the NHS, biomechanical CT analysis was conducted at L1, L2 or adjacent vertebrae (i.e., T12 if either L1 or L2 was not evaluable) to derive the following parameters:

- Estimated bone compressive strength (Newtons) by finite element analysis
- Trabecular volumetric BMD (mg/cm^3)
- BMC (g) of the vertebral body

The Applicant analyzed the data for changes from baseline, with adjustments for covariates, and reported nominally significant associations of palovarotene treatment with lower values for each of these parameters.

Vertebral fracture analysis (VFA) of T4 through L4 vertebrae was calculated using automated placement of 6 points to outline vertebral morphometric shape; these were used to determine deformity ratios for grading using the established criteria of Genant (Genant HK, 1993):

- Normal (Grade 0; deformity <20%)
- Mild (Grade 1; deformity ≥20% to <25%)
- Moderate (Grade 2; deformity ≥25% to <40%)
- Severe (Grade 3; deformity ≥40%)

At 12 months, 28% of Study 301 subjects and 12% of NHS subjects were found to have at least one new-onset morphometric vertebral fracture that was not present at baseline. The corresponding proportions for at least one new onset moderate/severe fracture (Grade 2 or 3) were 9% and 5%, respectively.

The Applicant also analyzed fracture adverse events: bone fractures were reported in 13/164 subjects (8%), including 2/23 subjects (9%) in the placebo/untreated arm.

In summary, the results of the biomechanical CT and vertebral fracture analyses suggest that palovarotene treatment may be associated with decreases in vertebral bone strength, BMC and BMD, and a higher risk of new onset morphometric vertebral fractures. These effects are consistent with some data involving other retinoids and are consistent with the BMD findings by DXA in osteochondroma subjects. However, the data do not establish a definitive causal association due to lack of randomization, use of non-validated measures and differing amounts of follow-up data between the Study 301 and NHS populations. VFA has not been validated for assessment of vertebral fractures in clinical trials, and the biomechanical CT data on vertebral

density and estimated strength are exploratory measures that have not been demonstrated to predict fracture risk.

3.3 Risk Mitigation

The review Division has worked closely with FDA's Division of Risk Management to determine appropriate risk mitigation for the serious risks associated with palovarotene. The Applicant proposes a voluntary risk management plan with the goal of educating prescribers, pharmacists, and patients about the benefits and risks associated with the use of palovarotene. The Applicant also proposes enhanced labeling with teratogenicity and premature epiphyseal closure (PPC) both outlined in a boxed warning, contraindication of pregnancy, and precautions, including these risks and recommended monitoring. The educational materials package consists of a guide for prescribers and pharmacists, a guide for patients and their caregivers, a guide on the importance of avoiding pregnancy, and a caregiver guide for growing pediatric patients. The Applicant proposes these materials would be available through the exclusive specialty pharmacy that dispenses palovarotene, the Ipsen Call Center, the product website, and through a patient support program.

FDA has required a Risk Evaluation and Mitigation Strategy (REMS) with elements to assure safe use for isotretinoin, another systemic retinoid, to ensure the benefits of isotretinoin outweighs the risk of embryo-fetal toxicity that can lead to life-threatening birth defects. A REMS is a drug safety program that the FDA can require for certain medications with serious safety concerns to help ensure the benefits of the medication outweigh its risks. REMS are designed to reinforce medication use behaviors and actions that support the safe use of that medication.

There are many factors that FDA considers in determining whether a REMS is necessary, including the severity of the condition. Isotretinoin is approved for the treatment severe recalcitrant nodular acne, a condition, that while severe, is not life-threatening. FDA also considers the population likely to use the product. There is significant use of isotretinoin in individuals that can become pregnant. These and other factors have led the FDA to require a REMS for isotretinoin that is designed to ensure that safe use conditions, including pregnancy testing and appropriate contraception, be documented prior to dispensing the drug to the patient. Acitretin, tretinoin, and bexarotene are other systemic retinoids approved without a REMS due to their differences from isotretinoin in various factors considered when determining whether a REMS is necessary.

Both the review Division and Division of Risk Management agree that a REMS is unlikely to be necessary to mitigate the risk of teratogenicity for palovarotene for several reasons, including the extremely small FOP population, the reproductive fitness of patients with FOP, and the use of consensus guidelines emphasizing the high risk of pregnancy to a FOP mother and fetus, with recommendations for appropriate family planning. We also do not believe a REMS is necessary to mitigate the risk of PPC because growing children with FOP are followed closely by specialists who are expected to have comprehensive and individualized risk-benefit discussions with FOP patients and their caregivers upon diagnosis and throughout the course of the disease. This

determination of the need for a REMS is independent of the implementation status of the Applicant's proposed voluntary non-REMS risk management plan.

4 References

FDA. (2019). Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products, Draft Guidance for Industry.

Genant HK, e. (1993). Vertebral Fracture Assessment Using a Semiquantitative Technique. *Journal of Bone and Mineral Research 8(9)*, 1137-48.

Kaplan FS, e. (2021). The Medical Management of Fibrodysplasia Ossificans Progressiva: Current Treatment Considerations. *Proc Intl Clin Council FOP*, 1-127.

Pignolo RJ, e. (2016). The Natural history of flare-ups in firbodysplasia ossificans progressiva (FOP): A comprehensive review. *Journal of Bone and Mineral Research, 31*(3), 650-6.

Pyle SI, G. W. (1959). *Radiographic Atlas of Skeletal Developement of the Hand and Wrist.* Stanford University Press.

5 Appendix

5.1 Clinical Studies Conducted for Palovarotene Treatment in Subjects With FOP

Table 18. Studies Conducted for Palovarotene Treatment in Subjects With FOP

Study	Design	Population	Dose	Endpoints
PVO-1A-001	Natural history study	n=114	No treatment	HO by WBCT every 12 months
PVO-1A-201	R, DB, placebo controlled 12-week study 6 weeks flare-up treatment, followed by 6 weeks observation	Subjects ≥6 years with FOP and active flare-up n=40 enrolled	Flare-up only dosing Placebo (n=10) 10 mg for 2 weeks, then 5 mg for 4 weeks (n=21) 5 mg for 2 weeks then 2.5 gm for 4 weeks (n=9)	Single flare-up site evaluated Primary: Percentage of subject responders, defined by no or minimal new HO (≤3 on 0-6 scale) at the flare-up site vs baseline as assessed by plain radiograph at week 6.
PVO-1A-202, A	Single arm, open label extension of Study 201 Treatment of 2 new flare-ups Duration of treatment, 6 weeks	Subjects from 201 n=40 enrolled	Flare-up only dosing: 10/5	Up to two flare-up sites assessed Primary: Percentage of subject responders as assessed by plain radiographs at the follow-up visit. Percentage of subject responders at up to two, new, distinct flare-up sites compared to baseline as assessed by CT at flare-up Week 6.
PVO-1A-202, B	Open label, single arm	Subjects ≥6 years with FOP n=54 (subjects from Part A and 18 new subjects)	Adults: Chronic dosing of 5 mg daily with flare-up dosing of 20 mg for 4 weeks then 10 mg for 8 weeks Children: no chronic 5 mg, flare-up only dose	Primary: Proportion of flare-ups with no new HO (“responders”) at Week 12.

Study	Design	Population	Dose	Endpoints
PVO-1A-202, C	Open label, single arm	Subjects from Part B n=46	Adults: chronic dosing of 5 mg daily with flare-up dosing of 20 mg for 4 weeks then 10 mg for 8 weeks Children: chronic dose allowed with flare-up dosing	Primary: Annualized change in new HO volume as assessed by WBCT
PVO-1A-301	Single arm, open label with NHS as external control	Adult and pediatric subjects ≥4 years of age with FOP n=107	Chronic dosing of 5 mg daily with flare-up dosing of 20 mg for 4 weeks then 10 mg for 8 weeks	Primary: Annualized change in new HO volume as assessed by WBCT

Source: compiled by reviewer
The higher-dose flare-up regimen ((20/10 mg) could be extended beyond week-12 as needed for persistent symptoms
Abbreviations: DB, double blind; FOP, fibrodysplasia ossificans progressiva; HO, heterotopic ossification; R, randomized; WBCT, whole body CT

5.2 Scoring Scale to Measure New HO on Plain Radiographs is Studies 201 and 202A

Table 19. Analogue Scoring Scale to Measure New HO on Plain Radiograph

Score[a]	Description	Measurement Requirements (Longest Diameter of HO/Normotopic Bone)
0	No HO	No measurement
1	Single or multiple spicules (punctate) or islands (non-contiguous) of HO	No measurement
2	Coalescing islands or reticular complexes of bone	No measurement
3	Single contiguous HO with longest dimension ≤1/2 diameter of reference normotopic bone[b] in any projection	≤50%
4	Single contiguous HO with longest dimension >1/2 but ≤1 diameter of reference normotopic bone[b] in any projection	>50% but ≤100%
5	Single contiguous HO with longest dimension >1 but ≤2 diameter of reference normotopic bone[b] in any projection	>100% but ≤200%
6	Single contiguous HO with longest dimension >2 diameters of reference normotopic bone[b] in any projection	>200%
NE	HO lesion is not evaluable (comments required)	No measurement

Source: Clementia -PVO-1A-201 and PVO-1A-202 Independent Review Charter

[a] Score is based on highest grade feature

[b] Reference normotopic bone for the indicated location of HO is defined as follows:

Proximal / Upper Extremity – humeral shaft
Distal Upper Extremity = radial shaft
Hip/Proximal Lower Extremity – width of femoral neck
Knee – femoral shaft
Distal Lower Extremity – tibial shaft
Ankle – tibial shaft
Foot - metatarsal

Abbreviations: HO, heterotopic ossification; NE, not evaluable

5.3 Study PVO-1A-202

Subjects who completed Study 201 could enroll in the open label extension Study PVO-1A-202 (Study 202). Study 202 consists of Parts A, B, and C, as the enrollment criteria, treatments, and study endpoints evolved over time. Similar to Study 201, for skeletally immature subjects under age 18 years, dosing was adjusted based on subject weight for all parts of Study 202.

Study 202 Part A

In 202 Part A, the primary objectives were to evaluate the longer-term safety and efficacy of prior palovarotene treatment in FOP subjects who completed Study 201; to assess the original treated flare-up from Study 201 for up to 12 months; and to evaluate the safety and efficacy of palovarotene in FOP subjects who experienced up to two new, distinct flare-ups.

Population: Subjects who completed Study 201 were eligible for Study 202. All 40 subjects enrolled in Study 201 completed that study and enrolled in Study 202 Part A.

Treatment: Palovarotene dosing continued as flare-based dosing 10 mg daily for 2 weeks followed by 5 mg daily for 4 weeks. Flare-ups were treated if they met the following criteria: presence of at least 2 out of 6 predefined symptoms (pain, soft tissue swelling, decreased ROM, stiffness, redness, and warmth), and were within 7 days of onset, consistent with the subject's previous flare-ups, and confirmed by the investigator.

Imaging: Efficacy was based on the ability of palovarotene to prevent HO at the new, distinct flare-up site as assessed by low-dose CT scan (or plain radiographs for subjects unable to undergo CT scan). For subjects who experienced the first new, distinct flare-up, low-dose CT scans at the new site were obtained at Flare-up Screening/Baseline and on Flare-up Weeks 6 and 12 (Days 42 and 84). For subjects who experienced a second new, distinct flare-up, the imaging approach changed. Low-dose CT scans at the new flare site were obtained at flare-up baseline and on flare-up Week 6 (Day 42), and then a low-dose WBCT scan was obtained at Flare-up Week 12 (Day 84). The low-dose WBCT scan was used to assess the site-specific HO at Flare-up Week 12 (Day 84) of the second flare-up and served as the baseline assessment of HO for eligible subjects enrolling into Study 202 Part B.

Endpoints: The primary endpoints were:

(1) the percentage of responders as defined by no or minimal new HO at the original flare-up site compared with baseline (pre-dose data from Study PVO-1A-201) as assessed by plain radiographs at the follow-up visit (study month 12).
(2) the percentage of responders, defined by no or minimal new HO at up to two new, distinct flare-up sites compared with baseline (flare-up screening/Day 1) as assessed by low-dose CT scan (or plain radiographs for subjects unable to undergo CT scan) at Flare-up Week 6 (Day 42) and Week 12 (Day 84).

Statistical Analysis: The primary efficacy analysis was the proportion (i.e., incidence) of flare-ups with no new HO at Week 12 relative to baseline as assessed by CT scan (or plain radiograph for subjects unable to undergo CT scan). The presence of HO at baseline and the incidence of flare-

ups with no new HO and those with new HO at post-baseline visits were summarized with counts and percentages of flare-ups.

Results:

Disposition: Twenty (50%) subjects who enrolled in 202 Part A had an additional flare and were treated with the flare-based palovarotene regimen while 20 (50%) subjects did not have a flare-up and were untreated. No subject required dose reduction, dose interruption, or study discontinuation.

Demographics: The demographics of the enrolled population is the same as that for Study 201. When evaluated in terms of subjects who had no flare-up versus those who did have a flare-up, subjects who had a flare-up were older (24 ± 9 years for the flare group compared to 19 ± 12 years for the no flare group).

Primary Endpoint: In the subjects who had additional flare-ups, 13 (65%) had one additional flare-up, 6 (30%) had two additional flare-ups and one subject had three additional flare-ups.

Assessment of HO was only performed at the flare-up sites. At baseline, HO was present at 18 (64%) flare sites. At Week 6, 21 (75%) flare sites had no new HO while 7 (25%) did have new HO. At Week 12, 18 (64%) flare sites had no new HO while 10 (36%) did have new HO. The baseline mean (SD) volume of HO at the flare site was 25 (49.3) cm^3, and the median was 4 cm^3 (range of 0 to 165 $cm^{3)}$. The mean (SD) volume of new HO was 2 (4.0) cm^3 at Week 6, and 2 (4.7) cm^3 at Week 12.

Study 202 Part B

In 202 Part B, the primary objective was to evaluate the safety and efficacy of different palovarotene dosing regimens in subjects with FOP regardless of whether flare ups were experienced. Efficacy was based on the ability of palovarotene to prevent the formation of new HO, as assessed by low-dose CT scan (or plain radiographs for subjects unable to undergo CT scan).

Population: Subjects from Study 202 Part A could continue in the study. Enrollment was also opened to include additional subjects with FOP who had not had flare-up symptoms for at least 4 weeks before enrollment. A total of 54 subjects were enrolled in 202 Part B, 36 from 202 Part A and 18 additional subjects, 13 of whom came from the NHS.

Treatment: Dosing included a new palovarotene 5 mg daily dose with flare-up dosing increased in dose and duration to palovarotene 20 mg for 4 weeks followed by 10 mg for 8 weeks (20/10 mg regimen). The criteria for flare-up treatment were the presence of at least 2 out of 6 predefined symptoms (pain, soft tissue swelling, decreased ROM, stiffness, redness, and warmth), and were within 7 days of onset, consistent with the subject's previous flare-ups, and confirmed by the investigator. The daily dose regimen was restricted to ≥90% skeletally mature subjects. Enrolled subjects with less than 90 percent skeletal maturity were treated only with flare-based dosing as needed.

Imaging: Radiologic evaluation at the flare-up site by low dose CT scan (or radiographs if CT was not available) was performed at flare-up screening/baseline and flare-up Week 12 (Day 84). Interpretation of the flare-up site-specific CT scan was to have documented the absence or presence of new HO at the flare-up site compared with the baseline assessment, and the amount (volume) of new HO if present. WBCT scans were performed in subjects receiving the chronic 5 mg daily regimen at screening, Month 12, and Month 24. Interpretation of the low-dose WBCT scan was to have documented the absence or presence of HO across various body regions, volume of total body HO, and presence and volume of new HO at the follow-up visits.

Endpoints: The primary endpoint was proportion of flare-ups with no new HO ("responders") at Week 12. Secondary endpoints included assessment of change in HO volume at the flare site at end of treatment (Day 84). For subjects receiving the chronic daily palovarotene dose, change from baseline (Part B screening) in whole-body HO burden was also assessed by low dose CT at Months 12 and 24.

Statistical Analysis: The primary efficacy analysis was the proportion (i.e., incidence) of flare-ups with no new HO at Week 12 relative to baseline as assessed by CT scan (or plain radiograph for subjects unable to undergo CT scan). The presence of HO at baseline and the incidence of flare-ups with no new HO and those with new HO at post-baseline visits were summarized with counts and percentages of flare-ups. Similar analyses were performed for subjects who had extended treatment. The primary analysis is presented for the flare-up population (either palovarotene 20/10 mg alone or palovarotene 5 mg daily plus 20/10 mg flare-up regimen). WBCT data analyses were not part of the 202 Part B statistical analysis plan as most subjects continued in 202 Part C.

Results:

Disposition: A total of 54 subjects enrolled in 202 Part B, with 44 receiving the 5 mg daily palovarotene dose with flare-up dosing if needed and 10 receiving only the flare-up dosing regimen if needed. Two subjects reached skeletal maturity during 202 Part B and transitioned from the flare-up only regimen to the daily dose plus flare-up regimen. Overall, 33 subjects were treated for a flare-up. Eight subjects were still symptomatic and required continued flare-up dosing beyond 12 weeks. Nine subjects required dose reduction because of adverse events.

Demographics: The mean age of the study population was 21 ± 9.2 years with a range of 7 to 54 years. When grouped by age, 12 (22%) subjects were under 15 years of age and 42 (78%) were 15 years or older. Forty-three percent of the enrolled population was male, and 57 percent was female.

Endpoints: Of the 33 subjects who were treated for a flare-up, 18 (54%) had 1 treated flare-up, 12 (36%) had 2 treated flare-ups, 2 (6%) had 3 treated flare-ups, and 1 (3%) had 4 treated flare-ups. Of the 52 flare-up events, 34 occurred with the chronic/flare-up dosing group and 18 occurred in the flare-up only dosing group.

At baseline, the mean (SD) volume of HO at the flare site was 6 (7.7) cm^3, and the median was 3 cm^3 with a range of 0 to 30 cm^3. When evaluation includes the 202 Part A group, the mean (SD) volume of new HO at the flare site at Week 12 was numerically higher with the chronic palovarotene 5 mg daily plus flare-up palovarotene 20/10 mg group (6 (20.7) cm^3) than in the flare-up only palovarotene 10/5 mg group (2 (4.7) cm^3) or the flare-up only palovarotene 20/10 mg group (3 (5.5) cm^3).

Evaluation of total body HO by WBCT also occurred in 37 subjects. At baseline, mean whole body HO (SD) was 454 (357.6) cm^3. At month 12, 36 subjects were evaluated, and 15 subjects had new HO with the mean new HO of 28 (89.9) cm^3. At month 24, one subject was evaluated, and that subject had new HO, reported as 193 cm^3.

5.4 Functional Outcome Scales

Cumulative Analogue Joint Involvement Scale (CAJIS) for FOP

The examiner should evaluate the following:

Neck – flexion/extension; rotation; lateral bending [0–2]
Thoraco-lumbar spine – forward flexion; chest expansion with deep breathing [0–2]
Jaw – mouth opening [0–2]
Shoulders – abduction [0–2 for each shoulder]
Elbows – flexion/extension [0–2 for each elbow]
Wrists – dorsiflexion/volarflexion [0–2 for each wrist]
Hips – flexion (standing, sitting, or supine) [0–2 for each hip]
Knees – flexion/extension [0–2 for each knee]
Ankles – dorsiflexion/plantarflexion [0–2 for each ankle]

Scores for each assessed area:

0 = normal to <10% deficit
1 = 10% - 90% deficit
2 ≥90% deficit

Possible scores:

Axial = 0–6
Upper limbs = 0–12
Lower limbs = 0–12
Total = 0–30

The examiner MUST also note and record if the patient:

- Can walk [yes/no]
- Uses a wheelchair [yes/no]
- Needs SOME help with activities of daily living [yes/no]
- Needs COMPLETE help with activities of daily living [yes/no]

Forearm rotation (pronation/supination), subtalar motion, and finger/toe motion are not assessed. The CAJIS for each region is based on the greatest limitation in ANY aspect of assessed function. For example, if forward flexion of the thoracic and lumbar spine is not possible but chest expansion is only minimally affected (or vice versa) then the score for the thoracic and lumbar spine would be a 2 (maximal).
Source: Bone 101 (2017): 123-128.

FOP Physical Function Questionnaire

The FOP-PFQ consists of 28 questions on the adult form and 26 questions on the pediatric form ranging from 1 (not able to do) to 5 (with no trouble; without help or assistive device). Lower scores denote more difficulty, with items categorized into total, upper extremity, and mobility sections.

The total score was calculated as follows:

- The sum of the scores from each question (values range from 28 to 140 for the adult form)
- The sum of the scores from each question (values range from 26 to 130 for the pediatric form)

The upper extremities sub-score was calculated as follows:

- The sum of the scores from 15 questions (values range from 15 to 75 for the adult form)
- The sum of the scores from 18 questions (values range from 18 to 90 for the pediatric form)

The mobility sub-score was calculated as follows:

- The sum of the scores from 13 questions (values range from 13 to 65 for the adult form)
- The sum of the scores from 8 questions (values range from 8 to 40 for the pediatric form)

Source: Clinical Study Report for Protocol PVO-1A-001

The PROMIS Global Health Scale

Figure 2. The PROMIS Global Health Scale

The PROMIS Global Health short form consists of 10 questions on the adult form and nine questions each on the subject-completed and proxy-completed pediatric form, scored on varying scales.

For the adult form, Global Physical Health and Global Mental Health scores range from 4 (worse health) to 20 (better health).

In the calculation of the Global Physical and Mental Health scores, questions were rescaled as shown in Table 6.

Table 6. Rescaled Global Physical and Mental Health Scores

Parameter(s)	Raw Score	Rescaled Score
7	0	5
	1 to 3	4
	4 to 6	3
	7 to 9	2
	10	1
8 and 10	1	5
	2	4
	3	3
	4	2
	5	1

Source: Appendix 16.1.9 Statistical Analysis Plan Version 2 Table 7

For the pediatric form, only a total score was calculated, with values ranging from 7 (worse health) to 35 (better health). If a subject was missing some (but not more than three) of the contributing question scores, the total score was calculated as the average observed score multiplied by the number of expected question scores. Global Physical Health scores and Global Mental Health scores were also converted to T-scores. A T-score of 50 is normal. A T-score <50 indicates worse health while a T-score >50 indicates better health.

Source: Clinical Study Report for Protocol PVO-1A-001

5.5 Additional Statistical Analyses

Details of the Prespecified Bayesian Analysis

There are two components of the Bayesian analysis for the annualized rate of change of new HO volume. The first component, denoted as K, is the number of body regions with new HO. The second component, denoted as *Z*, is the new HO volume per region where new HO has occurred. Multiplication of these two components is used to estimate the annualized new HO for each study. The treatment effect is measured by the ratio of the annualized new HO for subjects in Study 301, divided by the annualized new HO for NHS subjects. The Bayesian model only counts the occurrence of positive new HO as an event.

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}~ $\text{Poisson}(\lambda_{i,j}*w_{ij}*\theta_{1,t(i,j)})$, where

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

The subject-level rate $\lambda_{i,j}$ is gamma distributed and accounts for 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 j_{th} WBCT scan and 0 if the subject was not on treatment. The marginal distribution of the number of body regions with new HO follows a negative binomial distribution. Setting $\theta_{1,0}$=1, the variable $\theta_{1,1}$ is the multiplicative effect of palovarotene treatment on the rate of body regions with new HO. Covariates adjust for 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 a factor variable equal to 0 for <18 years old and 1 for ≥18 years old. Covariate effects exp (β_1) and exp (β_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 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 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 restrict $\alpha_r = \alpha_{r'}$ is imposed, where r and r' are a left and right region pair (e.g., right arm versus left arm). The precision variables, τ_0 and τ_1, allow variability to differ between new HO volume in treated subjects verses untreated subjects. The treated effect is $\gamma = \theta_{1,1} * \theta_{2,1}$The priors for variables in Bayesian Compound Poisson Distribution are as follows:

The prior distribution for subject level rate $\lambda_{i,}$ is Gamma (a, b) where a and b are in turn both distributed as Gamma (1,1).

The prior distribution for the covariate coefficients β_1 and β_2 for sex and age, respectively, are N $(0,2^2)$.

α ~ N (0,100)

α_r ~Uniform (0,4)

τ_0, τ_1 ~Gamma (1,0.01)

$\theta_{1,1}$, $\theta_{2,1}$ ~Uniform (0,2)

Table 20. wLME for Annualized New HO Volume (No Square-Root Transformation and Negatives Included) for Palovarotene-Treated and Untreated Subjects, Replacing Annualized New HO >100 cm^3 With 100 cm^3 (Principal FAS)

Method	Study 301 N=97	NHS N=101	Difference (95% CI)	Nominal p-Value
Threshold analysis (HO volumes >100 cm^3 set to 100 cm^3)	8.7	17.7	-9.1(-15.8, -2.3)	0.01

Source: Table 25 in ISE. The annualized new HO wLME LSmean estimate and SEM are from a mixed model with dependent variable annualized new HO and independent variables including fixed effects of treatment and baseline total HO/baseline age and a random subject effect. Wilcoxon p-value 0.0003
Abbreviations: CI, confidence interval; FAS, full analysis set; HO, heterotopic ossification; NHS, natural history study; wLME, weighted linear mixed effect model

Similar analyses with different thresholds from 90 cm^3 (new HO >90 cm^3 set to 90 cm^3) to 5 cm^3 were performed. Results were even more significant. These results indicate that the difference was not likely driven by extreme large values of subjects in NHS.

Table 21. Sensitivity Analyses on Influence of Large Values

Threshold (cm^3)	Study 301	NHS	Difference	LCL	UCL	p-Value
100 (Table 25 in ISE)	8.686	17.742	-9.057	-15.848	-2.266	0.01
90	8.5	17.3	-8.8	-15.4	-2.2	0.01
70	8.0	15.6	-7.6	-13.5	-1.7	0.0125
50	7.0	13.6	-6.5	-11.4	-1.7	0.01
30	4.8	10.3	-5.5	-9.1	-1.8	0.004
10	1.2	4.85	-3.6	-5.5	-1.7	0.0004
5	-0.2	2.6	-2.8	-4.2	-1.4	0.0003

Source: FDA reviewers' analyses
Analyses of New HO with different thresholds from 90 cm^3 (new HO >90 cm^3 set to 90 cm^3) to 5 cm^3. New HO values exceeding threshold value were set to the threshold value. The annualized new HO wLME LSmean estimate and SEM are from a mixed model with dependent variable annualized new HO and independent variables including fixed effects of treatment and baseline total HO/baseline age and a random subject effect.
Abbreviations: ISE, integrated summary of effectiveness; LCL, lower confidence limit; NHS, natural history study; UCL, upper confidence limit

Table 22. Summary of Balance After Weighting (Study 301 vs. NHS)

Variable	Study 301 Means (N=97)	NHS Means (N=101)	Standardized Mean Difference
Propensity score, mean (min, max)	0.5 (0.2, 0.67)	0.5 (0.17, 0.67)	0
Age	16.83	16.60	0.025
Female (%)	0.45	0.45	0
Baseline rate of HO	15.17	15.11	0.0046
CAJIS	10.98	10.92	0.0091
Time since last flare up (month)	21.64	21.09	0.017

Source: FDA reviewer
Baseline rate of HO is the baseline total HO divided by age.
Abbreviations: CAJIS, Cumulative Analogue Joint Involvement Scale; HO, heterotopic ossification; NHS, natural history study

Table 23. Summary of Balance After Nearest Neighbor 1:1 Matching (Study 301 vs. NHS)

Variable	Study 301 Means (N=61)	NHS Means (N=61)	Standardized Mean Difference	Variance Ratio
Propensity score	0.51	0.51	0.04	1
Age	15.70	15.15	0.062	1.27
Female (%)	0.39	0.39	0	1.0
Baseline rate of HO	13.70	13.11	0.045	1.22
CAJIS	10.49	10.18	0.052	0.72
Time since last flare up (month)	21.18	18.40	0.21	0.93

Source: FDA reviewer
Exact match on sex and age group (<12;12 to <18: >=18) and caliper=0.2 for nearest neighbor propensity score matching; Baseline rate of HO is the baseline total HO divided by age.
Abbreviations: CAJIS, Cumulative Analogue Joint Involvement; HO, heterotopic ossification; NHS, natural history study

5.6 Nonclinical Assessment of the Potential Effectiveness of Palovarotene

A series of in vitro and in vivo studies were conducted in cell lines and murine models of FOP whose pathology is driven by excessive ALK2 activity, recapitulating the dysregulated BMP/ALK2 pathway described in human FOP.

In Vitro Studies Supporting the Mechanism of Action

Palovarotene demonstrated binding and functional selectivity towards retinoic acid receptor (RAR) gamma (RARγ) over RARα and RARβ, with IC_{50} and EC_{50} values of 450 nM and 8 nM, respectively. RARγ is expressed in chondrogenic cells and chondrocytes. In the initial studies, pretreatment with palovarotene inhibited BMP-induced chondrogenic differentiation of cultured ADTC5 cells (chondrogenic progenitor cells) and led to the loss of skeletogenic potential of bone marrow derived mesenchymal stem cells. Similar findings were recapitulated in a BMP-treated human FOP fibroblast cell line carrying human relevant gain-of-function R206H mutation of the ALK2 receptor. In these cell lines, palovarotene antagonized the BMP/ALK2 signaling pathway and reduced the levels of phosphorylated Smad proteins by approximately 43% to 50% (Smad1, Smad5 and Smad8). The effect of palovarotene on Smad proteins was absent in the presence of proteasome inhibitors, suggesting that palovarotene might promote proteasome-mediated degradation of Smad proteins.

In Vivo Nonclinical Efficacy Data Showing Proof of Concept

Proof of concept in vivo studies for palovarotene were conducted in three injury-induced animal models of FOP including BMP implant, transgenic mouse with constitutively active receptor ($ALK2^{Q207D}$), and a human relevant conditional knock-in mouse model ($ALK2^{R206H}$). Treatment with palovarotene demonstrated reductions in HO formation in a dose-related manner across these models as compared to vehicle-treated groups. Palovarotene also improved joint mobility, reduced mast cell infiltration and fibroproliferative responses at the site of injury. However, the effectiveness of palovarotene only covered a narrow window of treatment mainly during the pre/chondrogenic stage (Day 1 to Day 3 postinjury) but not the ossification phase. In addition to the injury-based models described above, palovarotene blocked non-injury (spontaneously) induced HO formation in the fore and hind limbs of pups when palovarotene was administered orally to nursing transgenic $Prrx1^{R206H}$ mice that carry the R206H mutation in $Prrx1^{+}$ skeletal progenitor cells.

Taken together, the completed nonclinical pharmacology studies provided an early proof-of-concept that treatment with palovarotene had a potential benefit in preventing injury-induced HO, supported the purported mechanism, aided dose selection for clinical trials, and justified the rationale for treatment at the time of flare-up before new cartilage and bone formation have taken place. However, findings from the preclinical models may not necessarily translate to clear clinical benefit due to species-specific differences and limitations of studies designed to show proof-of-concept but not powered or designed to replace clinical trials in FOP patients. Human FOP is a complex, heterogenous disease without a well understood pathology and flare-

ups are located in different tissues (muscles, tendons, or ligaments). However, the injury-based animal models used in this program lack heterogeneity in disease expression as the animals only had a single injury site at a specified location which was standardized across the treatment groups. Moreover, the disease mechanism between animal and human FOP is different as the BMP-implantation model is created with supraphysiologically high rhBMP-2 levels. The genetic model bearing the $ALK2^{Q207D}$ mutation is not found naturally in FOP patients. The $ALK2^{R206H}$ and $Prrx1^{R206H}$ mice do carry the most common mutation found in human FOP and do display many of the phenotypic features of FOP patients, including hind limb digit malformation, joint fusions, and extensive HO development. Treatment with palovarotene was initiated at an early age in all transgenic mouse studies (birth to 2 month), corresponding to the period of maximal skeletal growth and maturation and generally translating to pre-pubertal human childhood but not to sexually and skeletally mature adults. Furthermore, the timing for initiation of treatment varies between the nonclinical and clinical studies. In animal models, dosing with palovarotene was initiated prior to or immediately following injury (aligned with the catabolic and chondrogenic phase) whereas flare-up dosing in the clinical trials was targeted after the initiation of the catabolic and chondrogenesis process. Comparisons of nonclinical pharmacology models and study data are shown in Table 24 (FDA Summary) and Table 25 (Applicant table).

Table 24. Comparison of Nonclinical FOP Disease Models With Human FOP

	Mouse Model				
Parameters	**BMP-2 Implant**	**ALK2^{Q207D}**	**ALK2^{R206H}**	**Prrx1^{R206H}**	**Human FOP**
ACVR1 (R206H) mutation and location	− −	− −	+ *global*	+ *limbs*	+ *global*
Injury-induced HO	+ *single site*	+ *single site*	+ *single site*	− −	+ *multiple sites*
HO induction method	BMP2 implant (SC)	Transgene (ALK2^{Q207D}) activation and cardiotoxin (IM)	Transgene (ALK2^{R206H}) activation and cardiotoxin (IM)	ALK2^{R206H} mutation	ALK2^{R206H} mutation, trauma, infection, etc.
Spontaneous HO	−	−	−	+	+
Human equivalent dose	0.6 to 19 mg	1 to 15 mg	3 to 19 mg	7 mg	Chronic: 5 mg, Flare-up: 20 mg for 4 weeks followed by 10 mg for 8 weeks
Age at treatment initiation	2 months	16 days	1 months	Birth	Variable
Treatment initiation	Immediately after injury	3 days post injury	On the day of injury	Birth via breast milk	Variable
Treatment duration	12 days	15 days	14 days	30 days	Chronic and flareup regimen
Pharmacologic endpoint	HO volume in ectopic masses (mCT)	HO volume in injured muscle (mCT)	HO volume in injured muscle (mCT), mobility, histology	Whole body skeleton (mCT), mobility, histology	HO volume (WBCT)

Source: Compiled by the reviewer
Abbreviations: BMP bone morphogenic protein; BMP-2; bone morphogenetic protein-2, BV/TV; Bone volume/total volume; HO; heterotrophic ossification, IM, intramuscular; mCT, microcomputed tomography, SC; subcutaneous; WBCT; whole body computed tomography

Table 25. Applicant's Clinical FOP and Mouse Heterotopic Ossification Model Comparison Summary

Comparison of Clinical and Genetic Features of FOP with Mouse Models of Heterotopic Ossification

Feature	Human FOP (classic)	BMP implant models	Nse-BMP4 transgenic mice	caALK2 (Q207D) transgenic mice	*ALK2* (R206H) knock-in mice
Heterotopic endochondral ossification (HEO)	+	+	+	+	+
Progressive HEO in characteristic anatomic patterns	+	–	–	–	?
Malformed great toes	+	–	–	–	+
Orthotopic fusion of subaxial cervical vertebrae	+	–	–	–	+
Costovertebral malformations	+	–	–	–	+
Osteochondromas	+	–	–	–	+
Short broad femoral necks	+	–	–	–	+
Early degenerative joint disease	+	–	–	–	+
Spontaneous disease progression	+	–	+	–	+
Inflammation-induced lesion formation	+	–	+	+	+
ACVR1 (R206H) canonical mutation	+	–	–	–	+
Locus integrity of gene mutation	+	N/A	–	–	+
Stoichiometric fidelity of gene mutation	+	N/A	–	–	+
Germline transmission	+	N/A	+	Embryonic lethal	Perinatal lethal

?, unknown; caALK2, constitutively active ALK2.

Source: Applicant's NDA 215559 submission (Pharmacology Written Summary)
Abbreviations: FOP, fibrodysplasia ossificans progressiva; HEO, heterotopic endochondral ossification