Clinical and Regulatory Development

Biostatistics and Data Management

REGENERON
Biostatistics and Data Management

STATISTICAL ANALYSIS PLAN

VERSION: FINAL

Title: **A Randomized, Placebo-Controlled Study to Assess the Safety, Tolerability, Pharmacokinetics, and Effects on Heterotopic Bone Formation of REGN2477 in Patients with Fibrodysplasia Ossificans Progressiva**

Compound:	**REGN2477**
Protocol Number:	**R2477-FOP-1623**
Clinical Phase:	**Phase 2**
Sponsor:	**Regeneron Pharmaceuticals, Inc.**
Study Biostatistician:	[redacted]
Project Statistician	[redacted]
Clinical Trial Manager:	[redacted]
Study Medical Director:	[redacted]
Imaging:	[redacted]
Version/Date:	**FINAL/ October 2020**

The approval signatures below indicate that these individuals have reviewed the Statistical Analysis Plan (SAP) and agreed on the planned analysis defined in this document for reporting.

See appended electronic signature page

Study Biostatistician: [redacted]

See appended electronic signature page

Clinical Pharmacologist: [redacted]

See appended electronic signature page

Bioanalytical Scientist: [redacted]

See appended electronic signature page

Medical Director: [redacted]

See appended electronic signature page

Sr Director Imaging: [redacted]

See appended electronic signature page

Project Biostatistician: [redacted]

See appended electronic signature page

Head of BDM or designee: [redacted]

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS 7
1. OVERVIEW 9
1.1. Background/Rationale 9
1.2. Study Objectives 10
1.2.1. Primary Objective 10
1.2.2. Secondary Objective 10
1.2.3. Exploratory Objectives 11
1.2.4. Modifications from the Statistical Section in the Final Protocol 11
1.2.5. Modifications from the Approved Statistical Analysis Plan 11
2. INVESTIGATION PLAN 12
2.1. Study Design and Randomization 12
2.2. Sample Size and Power Considerations (Period 1) 12
2.3. Sample Size and Power Considerations (Period 2) 14
2.4. Study Plan 14
3. ANALYSIS POPULATIONS 15
3.1. Baseline-Active HO Analysis Set 15
3.2. Baseline-Active HO Classic ACVR1 [R206H] Mutation Analysis Set 15
3.3. The Full Analysis Set 15
3.4. The Safety Analysis Set 15
3.5. The Pharmacokinetic Analysis Set 16
3.6. The Total Target Analysis Set 16
3.7. The Anti-Drug Antibody Analysis Set 16
4. ANALYSIS VARIABLES 17
4.1. Demographic and Baseline Characteristics 17
4.2. Medical History 17
4.3. Pretreatment/Concomitant Medications 17
4.4. Efficacy Variables (Period 1) 18
4.4.1. Primary Efficacy Endpoints (Period 1) 18
4.4.2. Key Secondary Efficacy Endpoints (Period 1) 19
4.4.3. Other Secondary Efficacy Endpoints (Period 1) 20
4.4.4. Exploratory Endpoints (Period 1) 21

4.5. Efficacy Variables (Period 2) 23
4.5.1. Primary Efficacy Endpoint (Period 2) 23
4.5.2. Key Secondary Efficacy Endpoints (Period 2) 24
4.5.3. Other Secondary Efficacy Endpoints (Period 2) 24
4.5.4. Exploratory Endpoints (Period 2) 26
4.6. Efficacy Variables (Period 3) 27
4.7. Safety Variables 29
4.7.1. Adverse Events and Serious Adverse Events 30
4.7.2. Laboratory Safety Variables 31
4.7.3. Vital Signs 33
4.7.4. 12-Lead ECG 33
4.7.5. Physical Examination 33
4.8. Pharmacokinetics 33
4.9. Total Target 33
4.10. Anti-Drug Antibody Variables 33
5. STATISTICAL METHODS 34
5.1. Demographics and Baseline Characteristics 34
5.2. Medical History 34
5.3. Prior/Concomitant Medications 34
5.4. Subject Disposition 35
5.5. Measurement of Treatment Compliance 35
5.6. Treatment Exposure and Observation Period 36
5.7. Analysis of Efficacy Variables 37
5.7.1. Hypothesis test and Multiplicity Control (Period 1) 37
5.7.2. Hypothesis Test and Multiplicity Control (Period 2) 38
5.7.3. Analysis of Efficacy Variables (Period 1) 39
5.7.3.1. Primary Endpoints 41
5.7.3.2. Secondary Key Efficacy Endpoints 43
5.7.3.3. Other Secondary Efficacy Endpoints 43
5.7.4. Analysis of Efficacy Variables (Period 2) 44
5.7.4.1. Primary Efficacy Endpoint 47
5.7.4.2. Key Secondary Efficacy Endpoints 48

5.7.4.3. Other Secondary Efficacy Endpoints 52
5.7.5. Analyses of Exploratory Efficacy Variables (Period 1, Period 2 and Period 3) 56
5.8. Analysis of Safety Data....60
5.8.1. Analysis of Adverse Events 60
5.8.2. Analysis of Clinical Laboratory Measurements....63
5.8.3. Analysis of Vital Signs 63
5.8.4. Analysis of 12-Lead ECG 63
5.8.5. Analysis of Physical Examination 64
5.9. Analysis of Pharmacokinetics Drug Concentration Data 64
5.10. Analysis of Total Target Concentration Data 64
5.11. Analysis of Anti-Drug Antibody Data 64
6. DATA CONVENTIONS 66
6.1. Definition of Baseline for Efficacy/Safety Variables 66
6.2. General Data Handling Conventions 66
6.3. Data Handling Convention for Missing Data 66
6.4. Visit Windows 67
7. INTERIM ANALYSIS 71
8. SOFTWARE....72
9. REFERENCES 73
10. APPENDIX....74
10.1. Summary of Statistical Analyses (Period 1) 74
10.2. Schedule of Events 75
10.2.1. Footnotes for the Schedule of Events Table 2, Table 3, and Table 4 81
10.3. Criteria for Treatment-Emergent Potentially Clinically Significant Value 85
10.4. NRS Questionnaire 92
10.5. FOP Disease Activity Assessment (DAA) Questionnaire 92
10.6. SAS Code for EQ-5D Index Utility Scoring....96
10.7. Cumulative Analogue Joint Involvement Scale (CAJIS) for FOP 97

LIST OF TABLES

Table 1: Sample Size (Total Lesion Activity by ^{18}F-NaF PET, Total Volume of HO Lesions by CT, and SUVmax) at Week 28 for Comparisons between REGN2477 and Placebo Treatment Groups 13

Table 2: Schedule of Events: Screening/Baseline through Treatment Period 1 (Randomized Double-Blind Treatment Period) 75

Table 3: Schedule of Events: Period 2 (Open-Label Treatment Period) 78

Table 4: Schedule of Events: Period 3 (Follow-Up Treatment Period) 80

LIST OF FIGURES

Figure 1: Study Flow Diagram 14

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

ADA	Anti-drug antibody
AE	Adverse event
AHO	Baseline active HO analysis set
AHOC	Baseline active HO classic ACVR1[R206H] mutation analysis set
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
AUC	Area under the curve
CRF	Case report form
CAJIS	Cumulative Analog Joint Involvement Scale
CTCAE	Common terminology criteria for adverse events
CT	Computed tomography
ECG	Electrocardiogram
EOS	End of study
EQ-5D-3L	EuroQol 5 dimensions questionnaire with a 3-level scale
ET	Early termination
FDA	Food and Drug Administration
FEV1	Forced expiratory volume 1
^{18}F-NaF	Fluorine-18-labelled sodium fluoride
FOP	Fibrodysplasia ossificans progressive
FOP I-ADL	FOP Independent Activity of Daily Living
HDL	High-density lipoprotein
HIV	Human immunodeficiency virus
HO	Heterotopic ossification
ICF	Informed consent form
ICH	International Conference on Harmonisation
IV	Intravenously
LDH	Lactate dehydrogenase
LLN	Lower limit of normal
MedDRA	Medical dictionary for regulatory activities
MMRM	Mixed model repeated measures
NRS	Numeric rating scale
PCSV	Potentially clinically significant value
PET	positron emission tomography

PK	Pharmacokinetic
PT	Preferred term
RBC	Red blood cell
Regeneron	Regeneron Pharmaceuticals, Inc.
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SAS	Statistical Analysis System
SOC	System organ class
SUV	Standardized Uptake Value
SUVmax	maximum standardized uptake value (typically in reference to a specified ROI)
TEAE	Treatment emergent adverse event
ULN	Upper limit of normal
WBC	White blood cell
WHODD	World Health Organization Drug Dictionary

1. OVERVIEW

The purpose of the statistical analysis plan (SAP) is to ensure the credibility of the study results by pre-specifying the statistical approaches for analyses prior to database lock. The SAP is intended to be a comprehensive and detailed description of strategy and statistical technique planned to perform the analyses of data for study R2477-FOP-1623.

This SAP amendment specifies the details of the new Period 2 analyses. No changes have been made to the SAP with regard to the specified Period 1 analyses as those analyses have already been conducted.

1.1. Background/Rationale

REGN2477 is a fully human monoclonal antibody that specifically binds to and blocks signaling of activin A. REGN2477 is being developed for the treatment of fibrodysplasia ossificans progressiva (FOP), an ultra-rare genetic disorder affecting approximately 800 patients worldwide. Characteristics of FOP include progressive heterotopic ossification (HO) of skeletal muscle, ligaments, tendons, and fascia, which leads to joint immobility, significant disability, pain and premature death. There is no approved treatment for the prevention or reversal of HO associated with FOP. Fibrodysplasia ossificans progressiva appears to be caused by mutations in the region encoding for the intracellular domain of type I activin A receptor (ACVR1) with the most common mutation observed in patients being ACVR1[R206H] (Classic Mutation). Recent studies in a conditional-inducible mouse model of FOP have demonstrated that the ACVR1[R206H] variant drives HO in mice by conferring to the receptor the abnormal ability to recognize activin A as an agonistic ligand resulting in abnormal bone formation similar to the clinical condition. In this model, exogenous activin A drives exuberant HO activity and REGN2477 administration dramatically halts heterotopic bone formation. In vitro studies demonstrated that different mutations in the ACVR1 receptor transduced bone morphogenic protein (BMP) signaling when stimulated with activin A. These results indicate that REGN2477 may suppress HO for patients with FOP who have ACVR1 mutations other than ACVR1[R206H]. Recently, in support of the in vitro studies and the observations initially made with ACVR1[R206H] receptor tested in vivo, experiments in mice have been extended to another ACVR1variant, R258G, associated with the most severe FOP phenotype described thus far. This ACVR1[R258G] mouse model developed HO much like the ACVR1[R206H] mouse model, and HO was again inhibited by REGN2477 as evidenced by absence of HO by CT when dosed prophylactically. Taken together, these data suggest that REGN2477 may be an effective treatment for FOP irrespective of the underlying ACVR1 mutation.

Positron emission tomography (PET) with ^{18}F-NaF PET and volumetric computed tomography (CT) will be used to measure the change in heterotopic bone formation. ^{18}F-NaF PET has been shown in several other disease settings to provide a sensitive, specific and whole body quantitative measure of bone mineralization activity. Data published by Botman 2019 support the use of ^{18}F-NaF PET as predictor of HO growth in FOP and serve as the basis for evaluation of active HO lesions specifically associated with high-intensity ^{18}F-NaF PET signal. Furthermore,

volumetric assessment of HO lesions by CT over 6 months will also allow assessment of growth of HO lesions and transition of PET detectable lesions to CT detectable mature HO lesions.

The efficacy analyses for Period 1 of the study have indicated that REGN2477 prevents occurrence of new HO lesions. These data have prompted the redefinition of the study hypothesis and Period 2 analysis to test efficacy of REGN2477 in preventing the formation of new heterotopic bone lesions. The SAP amendment describes the Period 2 analyses to support the redefined hypothesis.

1.2. Study Objectives

1.2.1. Primary Objective

The primary safety objective of the study is to assess the safety and tolerability of REGN2477 in male and female patients with FOP.

The primary efficacy objective of the study is to assess the effect of REGN2477 versus placebo on the change from baseline in HO in patients with FOP, as determined by ^{18}F-NaF uptake in HO lesions by PET and in total volume of HO lesions by CT.

1.2.2. Secondary Objective

The secondary objectives of the study are:

- To compare the effect of REGN2477 versus placebo on pain due to FOP, as measured by the area under the curve (AUC) for pain based on daily Pain NRS scores
- To assess the effect of REGN2477 versus placebo on the change from baseline in HO, as determined by the number of new HO lesions identified by ^{18}F-NaF PET or by CT
- To assess the effect of REGN2477 versus placebo on the change from baseline in ^{18}F-NaF standardized uptake value maximum (SUVmax) of individual active HO site(s) by PET
- To assess the effect of REGN2477, between week 28 and week 56, on the number, activity, and volume of HO lesions identified by ^{18}F-NaF PET or by CT in patients who switch from placebo to REGN2477 at week 28 versus the same patients from baseline to week 28
- To assess the effect of REGN2477 versus placebo on the change from baseline in biochemical markers of bone formation
- To characterize the concentrations of total activin A at baseline and over time following the first dose of study drug
- To characterize the concentration of REGN2477 in patients with FOP

- To assess the immunogenicity of REGN2477

1.2.3. Exploratory Objectives

Exploratory objectives of the study are:

- To explore the effects of REGN2477 versus placebo on change from baseline on clinical endpoints such as the following: FOP disease activity, joint function, and patient reported activities of daily living, and glucocorticoids and pain medication use
- To explore the effect of REGN2477 versus placebo on normal bone metabolic activity as determined by change from baseline in ^{18}F-NaF PET uptake in skeletal bone
- To explore the effect of REGN2477 on the plasma concentration profile of ^{18}F-NaF
- To explore the effect of REGN2477 vs. placebo on the total dose of corticosteroids used over the treatment phase of the study
- To explore the correlation between change from baseline in ^{18}F-NaF PET signal of SUVmax at 8 weeks with change from baseline in volumetric CT evaluation of HO at 28 weeks
- To explore the effects of REGN2477 on lesion progression from PET detectable lesions to CT detectable lesions

1.2.4. Modifications from the Statistical Section in the Final Protocol

The MMRM analysis for time-weighted average (standardized area under the curve (AUC)) of the percent change from baseline in total lesion activity by 18F-NaF PET over 28 weeks was planned as a primary analysis in the final protocol. In response to FDA's recommendation (Advice/Information request, 18OCT2019), the primary analysis will now be based on an ANCOVA model. Ranked ANCOVA based on the endpoint will be used as sensitivity analysis. MMRM will be pre-specified in the SAP only as a supplementary analysis.

1.2.5. Modifications from the Approved Statistical Analysis Plan

The study R2477-FOP-1623 is currently ongoing and all patients have entered the open label treatment periods (Periods 2 and 3). At the time of the SAP amendment (Version 2), the treatment assignments in Period 1 are unblinded and results from the primary week 28 analyses are available. The Statistical Analysis Plan is revised to include Period 2 confirmatory efficacy analyses at week 56 to test the re-defined hypothesis that REGN2477 treatment leads to the reduction in new HO lesions. The modified SAP specifies additional week 56 endpoints and the statistical approach for the confirmatory efficacy analyses for study R2477-FOP-1623.

The purpose of the statistical analysis plan (SAP) amendment is to ensure conclusive evidence of efficacy by pre-specifying the statistical approaches for Period 2 analyses prior to the database lock for Period 2 at week 56. The SAP is intended to be a comprehensive and detailed

description of the strategy and the statistical techniques planned to perform the Period 2 confirmatory analyses of data for study R2477-FOP-1623. Imaging data from week 56 scans will not be available to the Sponsor until the week 56 database lock.

2. INVESTIGATION PLAN

2.1. Study Design and Randomization

This is a phase 2, randomized, placebo-controlled, study designed to evaluate the safety, tolerability, PK, and effects on heterotopic bone formation of repeated doses of 10 mg/kg IV REGN2477 Q4W in adult patients with FOP. Patients may be enrolled at multiple sites globally.

Forty-four FOP patients with or without current flare-up activities are randomized 1:1 to REGN2477 or placebo in the double-blind period. Randomization is stratified by presence/absence of baseline active HO lesions, gender, and classic ACVR1 [R206H] mutation/different ACVR1 mutations.

2.2. Sample Size and Power Considerations (Period 1)

The sample size estimation for at least 24 patients (12 patients per treatment group) with active HO at baseline and classic ACVR1[R206H] mutation is based on statistical considerations for the following efficacy endpoints: percent change from baseline in (1) total lesion activity by ^{18}F-NaF PET over 28 weeks, (2) total volume of HO lesion by CT at week 28, and (3) ^{18}F-NaF SUVmax at week 8. Accounting for a 20% dropout rate at week 28, the sample size would yield approximately 10 patients per treatment group for week-28 analyses. This sample size will provide 80% power at a 2-sided 0.05 significance level in allowing the detection of an observed treatment difference in the order of 57%, 65%, and 40% reduction in the total lesion activity by ^{18}F-NaF PET, the total volume of HO lesion by CT, and the ^{18}F-NaF SUVmax, respectively, if the measure of variability in FOP patients is similar to that observed in other bone diseases and the FOP mouse model.

No prior published data on treatment effects in patients with FOP using these endpoints are available; however, numerous case reports show positive ^{99m}Tc methylene diphosphonate (MDP) bone scan, and, an abnormally high ^{18}F-NaF PET signal in a small study of patients with FOP at sites of flare-up and "asymptomatic" locations.

There is no precedent in patients with FOP to allow estimation of variance or clinically meaningful effect sizes for these endpoints. Based on studies in other bone diseases in human and in the FOP mice, changes are estimated at a magnitude of 73% for the total lesion activity by ^{18}F-NaF PET, 65% for the total volume of HO lesion by CT, and 50% for ^{18}F-NaF SUVmax PET.

There are ^{18}F-NaF PET (SUVmax) data available from a bisphosphonate treatment study of 14 patients with Paget's disease. In this study, ^{18}F-NaF PET/CT was used to monitor treatment effects on bone. The SUVmax and bone influx constants (Ki) were found to be significantly

higher in Pagetic bone than normal bone (mean±SD are 35.07±23.86 vs. 7.15±4.76 and 0.114±0.115 vs. 0.014±0.008 mL/min/mL, respectively, where p<0.05 for both measures). Following treatment with bisphosphonates, a significant reduction of PET signal in Pagetic bone was observed compared to baseline (-27.84±24.59% at 1 month and -44.64±23.61% at 6 months, P<0.05 at both time points) (Installe 2005). These data yielded a log-scale SD~0.36 for 1-month change. Assuming a 50% reduction by REGN2477 versus placebo in total lesion activity by ^{18}F-NaF SUVmax PET and a log-scale SD~0.36, the sample size required for week-28 analysis assuming the effect size of 1.9 is 6 patients with active lesion per arm for 80% power at a 2-sided 0.05 significance level.

Prior data for total lesion activity by ^{18}F-NaF PET and total volume of HO lesion by CT in FOP are available only in mice.

REGN2477 treatment yielded an effect size of 2.2 and a log-scale SD of approximately 0.60 in reducing induced HO total lesion activity in mice as measured by ^{18}F-NaF PET AUC. Assuming an effect size of 2.2, the sample size required for week-28 analysis is 5 patients with active lesion per arm for 80% power at a 2-sided 0.05 significance level.

Effect sizes for total volume of HO by CT ranged 0.9 to 1.4 with a log-scale SD of 0.75 to 1.0. Assuming larger effect size of 1.4 and smaller log-scale SD of 0.75, the required sample size for week-28 analysis is 10 patients with active lesion per arm for 80% power at a 2-sided 0.05 significance level.

Considering that FOP is a rare disease and the robust preclinical efficacy of REGN2477, in the interest of maintaining smallest reasonable sample size, larger target effect sizes are used to compute power for potential sample size for this study. The calculations, which were carried out on log-scale (i.e., ln[post/pre]) and back-transformed to express differences in percent change scale, are summarized in Table 1.

Table 1: Sample Size (Total Lesion Activity by ^{18}F-NaF PET, Total Volume of HO Lesions by CT, and SUVmax) at Week 28 for Comparisons between REGN2477 and Placebo Treatment Groups

Endpoint	Sample Size[a]	TRUE Reduction[b]	Log-Scale SD	TRUE Effect Size
Time-weighted average (standardized AUC) percent change in total lesion activity by ^{18}F-NaF PET over 28 weeks	5	73% reduction	0.60	2.2
	10	57% reduction	0.60	1.4
Percent change in total volume of HO lesions by CT at week 28	10	65% reduction	0.75	1.4
	12	60% reduction	0.75	1.2
Percent change in ^{18}F-NaF SUVmax at week 8	6	50% reduction	0.36	1.9
	10	40% reduction	0.36	1.4

[a] N / treatment group for 80% power to yield statistically significant difference (alpha=0.05, 2-sided)
[b] percent reduction over placebo

2.3. Sample Size and Power Considerations (Period 2)

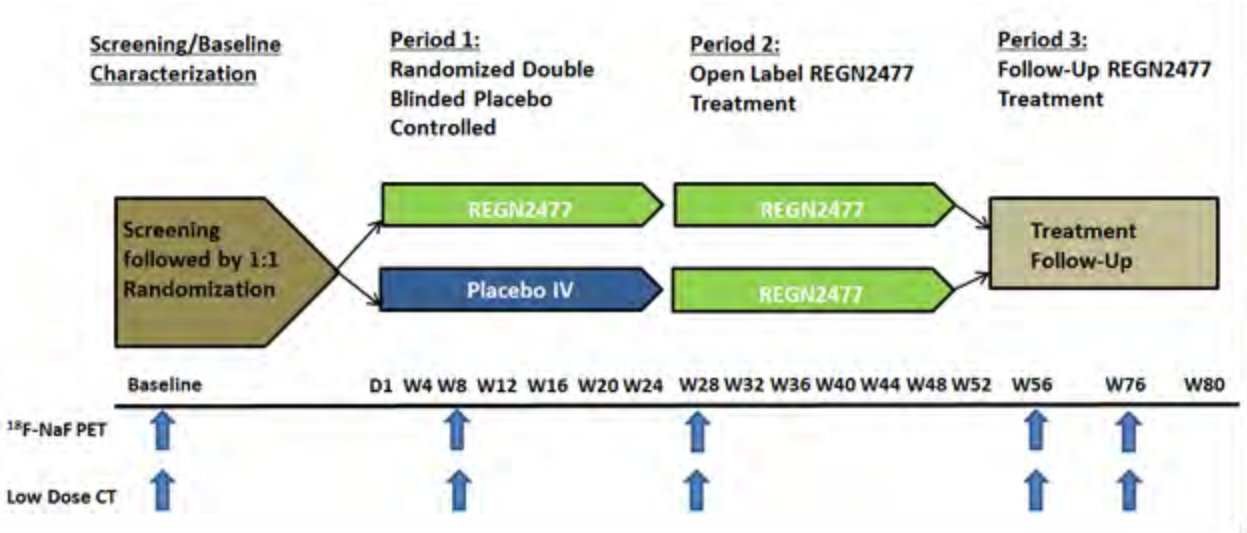
All 24 placebo patients have entered Period 2. Power calculations are based on simulations. Week 56 data of new HO lesions are generated using Poisson rate of 0.15. Wilcoxon test is used for the within group comparison on the simulated week 56 data and the observed week 28 data in patients who cross over to REGN2477 in Period 2. Twenty-four patients provide at least 90% power at two-sided 0.10 level of significance.

2.4. Study Plan

The study schedule of events is presented in Section 10.2 (Appendix). As depicted in the study design schematics (Figure 1), this study consists of a screening/baseline period (day -28 to day -1), two 6-month treatment periods, and a follow-up treatment period (Period 3). The 3 treatment periods are:

- Period 1: a 6-month randomized double-blind placebo-controlled treatment period
- Period 2: a 6-month open-label REGN2477 treatment period
- Period 3: a follow-up treatment period with REGN2477 continuing until patients have completed the week 76 visit, and all data have been collected and validated through the time when the last patient randomized into the study completes the week 28 visit (Period 1) and results of the primary analyses of safety and efficacy are available to the sponsor

Figure 1: Study Flow Diagram



3. ANALYSIS POPULATIONS

In accordance with guidance from the International Conference of Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guideline ICH E9 Statistical Principles for Clinical Trials (1998), the following population of analysis will be used for all statistical analysis:

3.1. Baseline-Active HO Analysis Set

The baseline-active HO analysis set (AHO) includes all randomized patients who had at least one active HO lesion at baseline; it is based on the treatment allocated (as randomized).

3.2. Baseline-Active HO Classic ACVR1 [R206H] Mutation Analysis Set

The baseline-active HO classic ACVR1 [R206H] mutation analysis set (AHOC) includes all randomized patients with the classic ACVR1 [R206H] mutation and who had at least one active HO lesion(s) at baseline, as defined by 18F-NaF PET positivity; it is based on the treatment allocated (as randomized)

3.3. The Full Analysis Set

The full analysis set (FAS) includes all randomized patients; it is based on the treatment allocated (as randomized).

3.4. The Safety Analysis Set

The safety analysis set (SAF) includes all randomized patients who receive any study drug and will be analyzed as treated.

The actual treatment group as treated is defined by the following rules based on treatment received in the double-blind placebo-controlled treatment period (Period 1):

- For a patient randomized to REGN2477,
  - if the patient received all placebo infusions, the actual treatment will be assigned as placebo
  - if the patient received at least one REGN2477 infusion, the actual treatment will be assigned as REGN2477
- For a patient randomized to placebo,
  - if the patient accidentally received one or more REGN2477 infusions, the actual treatment will be assigned as REGN2477; otherwise it will remain as placebo

The safety analyses will be based on the SAF. For safety summaries, three analysis periods are defined as follows:

- Double-blind period is defined as

  Day 1 from start of administration of the first dose of study drug through week 28 of the double-blind period.

- Open label and follow-up period is defined as duration starting from date of first dose of open label REGN2477 to the date of study completion (the end of the follow-up period).

- Entire study period is defined as from day 1 to the date of study completion, consisting of double-blind, open label and follow-up periods

Treatment compliance/administration and all clinical safety variables will be analyzed using the SAF.

3.5. The Pharmacokinetic Analysis Set

The PK analysis set includes all treated patients who received any study drug and who had at least 1 non-missing drug concentration following the first dose of study drug.

3.6. The Total Target Analysis Set

The total target analysis set includes all patients who received any study drug and who had at least 1 non-missing total target concentration following the first dose of study drug.

3.7. The Anti-Drug Antibody Analysis Set

The ADA analysis set includes all patients who received study drug and had at least 1 non-missing ADA result following the first study dose.

4. ANALYSIS VARIABLES

4.1. Demographic and Baseline Characteristics

The following demographic and baseline characteristics variables will be summarized:

Demographic variables:

- Age at screening with grouping (year; <30, ≥30) Sex, Ethnicity with grouping (Hispanic or Latino, Not-Hispanic or Latino), Race with grouping (White, Black, Asian, Other), baseline weight with grouping (kg, <60, ≥60), BMI (<25, ≥25 to <30, and ≥30), region (North America, Europe)

- Key Baseline disease characteristics: Age at FOP diagnosis, Duration of FOP, FOP genetic mutation type (Classic, other, number of active HO Lesions by ^{18}F-NaF PET and CT, total lesion activity, total volume of active HO lesions by 18F-NaF PET and CT, baseline average daily Pain NRS, baseline average daily Pain NRS in 7 days prior to first dose, joint function score (CAJIS), EQ-5D-3L total score, FOP independent activity of daily living (FOP I-ADL) score.

4.2. Medical History

Medical history will be coded to a Preferred Term (PT) and associated primary System Organ Class (SOC) according to the latest available version of the Medical Dictionary for Regulatory Activities (MedDRA).

4.3. Pretreatment/Concomitant Medications

Medications/Procedures will be recorded from the day of informed consent until the end-of-study visit. Medications will be coded to the anatomical therapeutic chemical (ATC) classification level 2 (therapeutic main group) and ATC level 4 (chemical/therapeutic subgroup), according to the latest available version of WHO Drug Dictionary (WHODD). Patients will be counted once in all ATC categories linked to the medication.

Pretreatment medications/procedures are defined as medications taken or procedures performed prior to administration of the study drug.

Concomitant medications/procedures are defined as medications taken or procedures performed from the time of first study drug administration to the final study visit. This includes medications that were started before the study and are ongoing during the study.

Prohibited concomitant medications/procedures

Treatment with the following concomitant medications/procedures is prohibited:

- Amiodarone
- Those that may have significant effects on skeletal bone turnover, such as Bisphosphonates (e.g., etidronate, pamidronate) Or denosumab
- Immunosuppressants (e.g., methotrexate)
- Chemotherapies
- Radiotherapy
- Bone marrow transplantation
- Antiplatelet therapy (e.g., clopidogrel)
- Anticoagulants (e.g., warfarin, heparin, factor Xa inhibitor, thrombin inhibitors)

4.4. Efficacy Variables (Period 1)

Efficacy variables for this study consist of assessments from imaging procedures, clinical endpoints and biomarkers of bone formation.

Imaging procedures for this study will include positron emission tomography (PET) with fluorine-18-labeled sodium fluoride (^{18}F-NaF) and volumetric computed tomography based on co-registered CT acquired using PET/CT scanner (^{18}F-NaF PET/CT). Positron emission tomography with ^{18}F-NaF and volumetric CT will be used to measure the change in heterotopic bone formation and to acquire dynamic and/or static emission and CT images from patients with FOP. The PET/CT scans will be used to assess changes to the uptake of ^{18}F-NaF in suspected active HO lesions as well as uptake into new lesions over time. All patients will undergo whole body static emission ^{18}F-NaF PET scan. The PET/CT scans will be performed at the baseline, week 8, 28, 56 and 76. The quantitative assessments of the ^{18}F-NaF PET and CT will be conducted by the two reviewers.

For ^{18}F-NaF uptake at HO site(s), all active HO lesions up to seven per patient will be identified and indexed at baseline, and their associated ^{18}F-NaF SUV_{mean}, SUVmax by PET, and their volume by CT will be derived. These index lesions will be assessed again on the post-dose treatment period visit scans. New lesions that arise during the post-dose treatment period will be assessed similarly.

The details of the imaging analysis procedures are given in the imaging charter.

4.4.1. Primary Efficacy Endpoints (Period 1)

Primary efficacy endpoints are:

- Time-weighted average (standardized AUC) of the percent change from baseline in total lesion activity by^{18}F-NaF PET over 28 weeks in AHO

- Percent change from baseline in the total volume of HO lesions as assessed by CT at week 28 in AHO

- Time-weighted average (standardized AUC) of the percent change from baseline in total lesion activity by ^{18}F-NaF PET over 28 weeks in AHOC

- Percent change from baseline in the total volume of HO lesions as assessed by CT at week 28 in AHOC

As specified in the imaging charter for the study, the quantitative assessments of the ^{18}F-NaF PET and CT are conducted by the two reviewers. If both reviewers are not in agreement regarding the number and location of new lesions during the post-baseline period, an adjudicator will select the assessment of either Reviewer 1 or 2. For the analysis of primary endpoints, the quantitative assessments done by the reviewer that adjudicator has selected at Week 28 will be used for the statistical analysis. If the adjudication at Week 28 is not done, the assessments done by Reviewer 1 will be used.

The following information based on the independent review will be provided:

- the number of replaced reviewers if any,

- the number of patients who had their imaging data adjudicated by timepoints,

- the number of times that either a partial or entire re-assessment of a patient's image was performed, which events occurred to cause a reassessment and their frequencies

The time weighted average (standardized AUC, i.e. AUC divided by time) for percent change of total lesion activity over 28 weeks for each patient will be calculated as AUC of percent change in total lesion activity divided by 28. Linear Trapezoidal rule will be used to calculate the AUC. The details are provided in Section 5.7.

Total HO volume by CT in each patient will be derived for each time point as the sum of all individual target and new HO lesion volume at that timepoint. Total lesion activity for each patient will be derived at each time point as the sum of lesion activity of individual target and new lesions.

If a lesion status is reported as non-evaluable for a particular lesion, then the HO volume, lesion activity and SUVmax for that lesion will be carried forward from the previous timepoint for that lesion.

4.4.2. Key Secondary Efficacy Endpoints (Period 1)

The key secondary efficacy endpoints are:

- Time-weighted average (standardized AUC) of the change from baseline in daily pain due to FOP, as measured using the daily NRS over 28 weeks in AHO

- Time-weighted average (standardized AUC) of the change from baseline in daily pain due to FOP, as measured using the daily NRS over 28 weeks in AHOC

The NRS (Numeric Rating Scale Pain Evaluation)

NRS is a categorical rating scale used by patients to rate their pain associated with FOP. Patients will be asked to rate their pain on a scale that ranges from "0" (no pain) to "10" (worst possible pain). The NRS assessments will be performed at the screening and baseline visits, and daily throughout the study and the results recorded in a pain diary (the e-diary device). Three pain scores are recorded each day in the e-diary, their current pain, worst pain, and least pain; the average of these three pain scores will be used for deriving these secondary endpoints.

Time weighted average of change from baseline in daily pain due to FOP will be calculated for each patient as AUC of change in daily pain divided by the number of days from first drug infusion to the last assessment in the double-blind period. Area under curve is calculated using "trapezoidal rule".

4.4.3. Other Secondary Efficacy Endpoints (Period 1)

The other secondary efficacy endpoints are:

- Percent change from baseline in ^{18}F-NaF SUVmax of individual active HO site(s) by PET at week 8 (AHOC)
- Percent change from baseline in ^{18}F-NaF SUVmax of individual active HO site(s) by PET at week 8 (AHO)
- Change from baseline in number of HO lesions as assessed by ^{18}F-NaF PET at week 28 (AHOC)
- Change from baseline in number of HO lesions as assessed by ^{18}F-NaF PET at week 28 (AHO)
- Change from baseline in number of HO lesions as assessed by ^{18}F-NaF PET at week 28 (FAS)
- Change from baseline in number of HO lesions detectable by CT at week 28 (AHOC)
- Change from baseline in number of HO lesions detectable by CT at week 28 (AHO)
- Change from baseline in number of HO lesions detectable by CT at week 28 (FAS)
- Time-weighted average (standardized AUC) of the change from baseline in daily pain due to FOP, as measured using the daily NRS over 28 weeks (FAS)
- Time weighted average (standardized AUC) of the percent change from baseline in biomarkers of bone formation levels in serum over 28 weeks, including Total

Procollagen Type 1 N-Terminal Propeptide (P1NP), bone specific alkaline phosphatase (BSAP), and total alkaline phosphatase (tAP) (FAS)

Biomarkers of bone formation

Biomarkers of bone formation include total procollagen type 1 N-terminal propeptide (P1NP), bone specific alkaline phosphatase (BSAP), and total alkaline phosphatase (tAP).

Time weighted average of percent change in biomarkers of bone formation will be calculated for each patient as area under curve (AUC) of percent change divided by the weeks from first infusion to the last assessment in the double-blind period. Area under curve is calculated using "trapezoidal rule".

4.4.4. Exploratory Endpoints (Period 1)

The exploratory endpoints are:

- Joint function assessment by physician at baseline and week 28, by the CAJIS
- Patient-reported quality of life by EQ-5D-3L at baseline and week 28
- Patient-reported assessment of activities of daily living by FOP I-ADL at baseline and week 28,
- FOP disease activity assessed by patient e-diary over time
  - Number and duration of flare-ups (defined in this study as experiencing at least 2 of the following: new onset of pain, swelling, joint stiffness, decrease in movement, or detection of HO)
  - Location and severity of FOP disease signs and symptoms including 1) pain, 2) swelling, 3) joint stiffness, 4) decreased movement, each scored using a 4-point scale (0: no symptom; 1: mild; 2: moderate; 3: severe)
- Total dosage of glucocorticoids use over time
- Number of patients with new HO lesions as assessed by ^{18}F-NaF PET through week 28
- Change in number of lesions that are only ^{18}F-NaF PET detectable at baseline to CT detectable lesions at week 28
- Change from baseline in the FEV1 of spirometry at week 28
- Time-weighted average (standardized AUC) change from baseline in the levels of hs-CRP over 28 weeks

Cumulative Analog Joint Involvement Scale (CAJIS)

The CAJIS is an assessment of 15 major joints; each major joint rated normal unaffected (0), affected (1), or completely functionally ankylosed (2). The total score ranges from 0 to 30.

Due to missing joints, the prorated average of the total CAJIS score will be calculated as 15*(total CAJIS score/number of non-missing joints). Also, the prorated average of axial score, upper limb, and lower limb score will be calculated.

Physicians at all sites will conduct CAJIS assessment at baseline, week 28, week 56, and week 76.

EQ-5D-3L

The EQ-5D-3L is a standardized measure of health status developed by the EuroQol Group in order to provide a simple, generic measure of health for clinical and economic appraisal. The EQ-5D-3L as a measure of health-related quality of life, defines health in terms of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, anxiety/depression. Each dimension has 3 ordinal levels of severity: absence of problems or issues with the specific dimension, e.g., "no problems" (1), "some problems" (2), "severe problems" (3). Overall health state is defined as a 5-digit number. Health states defined by the 5-dimensional classification can be converted into corresponding index scores that quantify health status, where 0 represents "death" and 1 represents "perfect health".

EQ-5D-3L health status will be converted into a single index value by using EQ-5D-3L value sets. Appendix 10.6 provides the SAS code to derive the index utility score using UK based population.

Patients will complete the EQ-5D-3L at baseline, week 28, week 56, and week 76.

FOP Disease Activity Assessment

FOP Disease Activity Assessment is an assessment of FOP disease activity, includes flare-up location, frequency, and duration, as well as the severity of selected FOP disease signs and symptoms, including pain, swelling, joint stiffness, and decrease in movement will be rated by patients using a 4-point scale for each (0: no symptom; 1: mild; 2: moderate; 3: severe). Patients and their caregiver (if applicable) will be trained on its use at the screening/baseline visit and are instructed to record their scores in the e-diary each day throughout the study.

FOP Independent Activity of Daily Living (FOP I-ADL)

FOP I-ADL is the questionnaire to assess patient's ability to independently perform activities of daily living. 28 questions in the FOP I-ADL are selected from the PROMIS bank of questions. For questions 1-24, ability to perform each activity is rated from (1) unable to do; (2) with much difficulty; (3) with some difficulty; (4) with a little difficulty; (5) without any difficulty and for questions 25-28, the extent to which health limits performance of an activity is rated from

(1) cannot do; (2) quite a lot; (3) somewhat; (4) very little; (5) not at all. Total score for I-ADL is the 140.

The FOP I-ADL will be collected at baseline, week 28, week 56, and week 76.

Total dosage of glucocorticoids

Disease flare-ups are typically treated with glucocorticoids. Total dosage of glucocorticoids will be calculated for each patient who take glucocorticoids.

Spirometry assessment of pulmonary function

FEV1 (forced expiratory volume in 1 second) will be determined by spirometry in patients who are able to undergo this procedure at baseline, week 28, week 56, and week 76.

Levels of hs-CRP

Inflammation marker hs-CRP will be collected at baseline, week 1, 4, 8, 16, 28, 32, 36, 44 and week 52.

Time weighted average of change of hs-CRP will be calculated as area under curve (AUC) of change divided by the weeks from first infusion to the last assessment day in the double-blind period. Area under curve is calculated using "trapezoidal rule".

4.5. Efficacy Variables (Period 2)

Efficacy variables for Period 2 analyses will be based on AHO population and will consist of assessments from imaging procedures, clinical endpoints and biomarkers of bone formation.

4.5.1. Primary Efficacy Endpoint (Period 2)

The primary efficacy variables in Period 2 is

- Number of new HO lesions as assessed by CT at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

The number of new CT and new PET lesions that develop after week 28 and appear on the week 56 scan will be identified and assessed for the primary and the key secondary endpoints that are defined in Section 4.5.2. As specified in the imaging charter for the study, the quantitative assessments of the ^{18}F-NaF PET and CT are conducted by the two reviewers. As done for week 28 analyses, on a per-patient basis, the quantitative assessments performed by the reviewer that the adjudicator selected at week 28 will be used for the week 56 statistical analysis. If the adjudication at week 28 was not required, the assessments performed by Reviewer 1 will be used in the analyses.

4.5.2. Key Secondary Efficacy Endpoints (Period 2)

The Key secondary efficacy endpoints for Period 2 analyses are:

- Total volume of new HO lesions as assessed by CT at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)
- Number of new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)
- Total lesion activity by ^{18}F-NaF PET in new lesions at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after double-blind period) (AHO)
- Percent of patients with new HO lesions as assessed by CT at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)
- Percent of patients with new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

4.5.3. Other Secondary Efficacy Endpoints (Period 2)

The other secondary efficacy variables for Period 2 analyses in AHO are as follows:

- Number of new HO lesions as assessed by CT only at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)
- Total volume of new HO lesions as assessed by CT only at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)
- Percent of patients with new HO lesions as assessed by CT only at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)
- Percent of patients with investigator-assessed flare-ups in Period 2 (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)
- Percent of patients with flare-ups assessed by patient e-diary in Period 2 (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

- Number of new HO lesions as assessed by CT at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)
- Number of new HO lesions as assessed by CT only at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)
- Number of new HO lesions as assessed by CT only at week 56 relative to week 28 scan (in patients who continue REGN2477 after the double-blind period) (AHO)
- Total volume in new HO lesions as assessed by CT at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)
- Total volume in new HO lesions as assessed by CT only at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)
- Percent of patients with new HO lesions as assessed by CT at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)
- Percent of patients with new HO lesions as assessed by CT only at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)
- Number of new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)
- Total lesion activity in new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)
- Percent of patients with new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)
- Percent change from week 28 in SUVmax to week 56 (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)
- Percent change from baseline in SUVmax to week 56 (in patients who continue REGN2477 after the double-blind period) (AHO)
- Percent change from week 28 in total lesion activity by ^{18}F-NaF PET to week 56 (in patients switching from placebo to REGN2477 after the double-blind period) versus the same patients between baseline and week 28 (AHO)

- Percent change from baseline in total lesion activity by ^{18}F-NaF PET to week 56 (in patients who continue REGN2477 after the double-blind period) (AHO)
- Percent change from week 28 in the total volume of HO lesions as assessed by CT to week 56 (in patients switching from placebo to REGN2477 after the double-blind period) versus the same patients between baseline and week 28 (AHO)
- Percent change from baseline in the total volume of HO lesions as assessed by CT to week 56 (in patients who continue REGN2477 after the double-blind period) (AHO)
- Change from week 28 in number of HO lesions by ^{18}F-NaF PET to week 56 (in patients switching from placebo to REGN2477 in double-blind period) versus the same patients between baseline and week 28 (AHO)
- Change from week 28 in number of HO lesions by CT to week 56 (in patients switching from placebo to REGN2477 in double-blind period) versus the same patients between baseline and week 28 (AHO)
- Daily pain due to FOP, as measured using the daily NRS (AHO)
- Total dosage of glucocorticoids use during Period 2 (AHO)

4.5.4. Exploratory Endpoints (Period 2)

The exploratory endpoints in AHO are:

- Change from week 28 in the total volume of new HO lesions to week 56 as assessed by CT (in patients switching from placebo to REGN2477 after the double-blind period)
- Joint function assessment by physician at week 56 by the CAJIS
- Patient-reported quality of life by EQ-5D-3L at week 56
- Patient-reported assessment of activities of daily living by FOP I-ADL at week 56
- FOP disease activity assessed by patient e-diary during Period 2
  - Number and duration of flare-ups (defined in this study as experiencing at least 2 of the following: new onset of pain, swelling, joint stiffness, decrease in movement, or detection of HO)
  - Location and severity of FOP disease signs and symptoms including 1) pain, 2) swelling, 3) joint stiffness, 4) decreased movement, each scored using a 4-point scale (0: no symptom; 1: mild; 2: moderate; 3: severe)
- Change from baseline in number of HO lesions (target and new lesions relative to baseline) as assessed by ^{18}F-NaF PET at week 56

- Change from baseline in number of HO lesions (target and new lesions relative to baseline) detectable by CT at week 56
- Number of patients with new HO lesions as assessed by ^{18}F-NaF PET through week 56 relative to baseline scan
- Percent change from baseline of mean SUV_{mean} of selected normal bones (e.g., lumbar spine and femoral heads) as assessed by PET/CT at week 56
- Percent change from baseline of ratio of mean SUV_{mean} of selected normal bones (e.g., lumbar spine and femoral heads) to venous plasma SUV during PET as assessed by PET/CT at week 56
- Percent change from baseline in venous plasma clearance ^{18}F-NaF SUV at week 56
- Percent change from baseline in ^{18}F-NaF incorporation rate (K_i) in individual active HO lesion(s) at week 56
- Percent change from baseline of ratio of SUVmax of individual active HO lesion(s) to venous plasma SUV during PET scan at week 56
- Change from baseline in the FEV1 of spirometry at week 56
- Change from week 28 in the FEV1 of spirometry at week 56

All the above primary, key secondary, other secondary and exploratory endpoints will also be assessed in the AHOC population. A separate analysis of the FAS will not be performed as the FAS is the same as the AHO set because all patients in the trial had active lesions at baseline.

4.6. Efficacy Variables (Period 3)

The exploratory endpoints for Period 3 analyses are:

- Number of new HO lesions as assessed by CT at week 76 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)
- Total volume of new HO lesions as assessed by CT at week 76 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)
- Number of new HO lesions as assessed by CT only at week 76 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

- Total volume of new HO lesions as assessed by CT only at week 76 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

- Number of new HO lesions as assessed by ^{18}F-NaF PET at week 76 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

- Total lesion activity by ^{18}F-NaF PET in new lesions at week 76 relative to week 28 scan (in patients switching from placebo to REGN2477 after double-blind period) (AHO)

- Percent of patients with new HO lesions as assessed by CT at week 76 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

- Percent of patients with new HO lesions as assessed by ^{18}F-NaF PET at week 76 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

- Percent of patients with investigator-assessed flare-ups in Period 2 and Period 3 (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

- Percent of patients with flare-ups assessed by patient e-diary in Period 2 and Period 3 (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

- Number of new HO lesions as assessed by CT at week 76 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)

- Total volume in new HO lesions as assessed by CT at week 76 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)

- Percent of patients with new HO lesions as assed by CT at week 76 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)

- Number of new HO lesions as assessed by ^{18}F-NaF PET at week 76 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)

- Total lesion activity in new HO lesions by ^{18}F-NaF PET at week 76 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)

- Percent of patients with new HO lesions as assessed by ^{18}F-NaF PET at week 76 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)

- Percent change from week 28 in SUVmax to week 76 (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

- Percent change from baseline in SUVmax to week 76 (in patients who continue REGN2477 after the double-blind period) (AHO)

- Percent change from baseline in total lesion activity by ^{18}F-NaF PET to week 76 (in patients who continue REGN2477 after the double-blind period) (AHO)

- Percent change from baseline in the total volume of HO lesions as assessed by CT to week 76 (in patients who continue REGN2477 after the double-blind period) (AHO)

- Daily pain due to FOP, as measured using the daily NRS during Period 3 (AHO)

- Joint function assessment by physician at week 76 by the CAJIS

- Patient-reported quality of life by EQ-5D-3L at week 76

- Patient-reported assessment of activities of daily living by FOP I-ADL at week 76

- FOP disease activity assessed by patient e-diary during Period 3

  – Number and duration of flare-ups (defined in this study as experiencing at least 2 of the following: new onset of pain, swelling, joint stiffness, decrease in movement, or detection of HO)

  – Location and severity of FOP disease signs and symptoms including 1) pain, 2) swelling, 3) joint stiffness, 4) decreased movement, each scored using a 4-point scale (0: no symptom; 1: mild; 2: moderate; 3: severe)

- Total dosage of glucocorticoids use during Period 3

- Change from baseline in the FEV1 of spirometry at week 76

- Change from week 28 in the FEV1 of spirometry at week 76

Additional exploratory endpoints will be analyzed if deemed appropriate.

4.7. Safety Variables

Subject safety will be assessed through the collection of AEs, laboratory data (hematology, chemistry, and urine), vital signs, 12-lead ECG, Physical exams.

4.7.1. Adverse Events and Serious Adverse Events

AEs and serious adverse events (SAEs) will be collected from the time of informed consent signature and then at each visit until the end of the study. All adverse events are to be coded to a "Preferred Term" and associated primary "System Organ Class (SOC)" according to the latest available version of MedDRA.

An AE is any untoward medical occurrence in a subject administered a study drug which may or may not have a causal relationship with the study drug. Therefore, an AE is any unfavorable and unintended sign (including abnormal laboratory finding), symptom, or disease which is temporally associated with the use of a study drug, whether or not considered related to the study drug. An AE also includes any worsening (i.e., any clinically significant change in frequency and/or intensity) of a preexisting condition that is temporally associated with the use of the study drug.

An SAE is any untoward medical occurrence at any dose that results in death, is life-threatening, requires in-patient hospitalization or prolongation of existing hospitalization, results in persistent or significant disability/incapacity, is a congenital anomaly/birth defect, or is an important medical event.

The severity of infusion reactions will be graded using the current version of the NCI-CTCAE (under "General Disorders and Administration Site Conditions").

The severity of AEs will be graded according to the following scale:

- **Mild**: Does not interfere in a significant manner with the subject's normal functioning level. It may be an annoyance. Prescription drugs are not ordinarily needed for relief of symptoms but may be given because of personality of the subject.

- **Moderate**: Produces some impairment of functioning but is not hazardous to health. It is uncomfortable or an embarrassment. Treatment for symptom may be needed.

- **Severe**: Produces significant impairment of functioning or incapacitation and is a definite hazard to the subject's health. Treatment for symptom may be given and/or subject hospitalized.

The criteria for determining whether an abnormal laboratory, vital sign or ECG finding should be reported as an AE are as follows:

- Test result is associated with accompanying symptoms, and/or

- Test result requires additional diagnostic testing or medical/surgical intervention, and/or

- Test result leads to a change in dosing (outside the protocol-stipulated dose adjustments), discontinuation from the study, significant additional concomitant drug treatment, or other therapy.

Adverse Events of Special Interest for this study are:

- Epididymitis
- Orchitis (inflammation of the testicles)
- Hydrocele (fluid buildup around 1 or both testicles)
- Scrotum pain
- Scrotum swelling
- Moderate to severe episodes of spontaneous non-traumatic bleeding
- Moderate epistaxis (defined as any episode lasting longer than 30 minutes or requiring professional medical intervention)
- Severe epistaxis (based on definition of a severe AE as per protocol)

Treatment-emergent adverse events are defined as those that are not present at baseline or those that represent an exacerbation of a preexisting condition during the double-blind period, open label and follow up treatment period.

4.7.2. Laboratory Safety Variables

Clinical laboratory tests will be collected at time points according to the schedule in Appendix 10.2. and include hematology, blood chemistry, and urinalysis.

Blood draws for laboratory testing at the screening visit will be collected in a fasted state (after an 8-hour fast). No other blood draw will require fasting.

Clinical laboratory values will be grouped by function in summary tables. Clinical laboratory tests are provided below, sub-grouped by function:

Blood Chemistry

Sodium	Total protein, serum	Total bilirubin
Potassium	Creatinine	Total cholesterol
Chloride	Blood urea nitrogen (BUN)	Triglycerides
Bicarbonate	Aspartate aminotransferase (AST)	Uric acid
Calcium	Alanine aminotransferase (ALT)	Creatine phosphokinase (CPK)
Glucose	Alkaline phosphatase	Magnesium
Albumin	Lactate dehydrogenase (LDH)	

Hematology and Coagulation

Hemoglobin	Differential:
Hematocrit	Neutrophils
Red blood cells (RBCs)	Lymphocytes
White blood cells (WBCs)	Monocytes
Red cell indices	Basophils
Platelet count	Eosinophils

Activated partial thromboplastin time	Prothrombin time
Thrombin Time	International Normalized Ratio (INR)
Platelet effector function	Fibrinogen

Urinalysis

Color	Glucose	RBC
Clarity	Blood	Hyaline and other casts
pH	Bilirubin	Bacteria
Specific gravity	Leukocyte esterase	Epithelial cells
Ketones	Nitrite	Crystals
Protein	WBC	Yeast

Other Laboratory Tests

Blood samples will be collected during screening to measure HbA1c and for DNA isolation for ACVR1 gene sequencing (the DNA samples will be stored until sequencing is performed).

Blood samples will be collected at the visits specified in Appendix 10.2 to measure adrenocorticotropic hormone (ACTH), growth hormone (GH), testosterone (total and free; male patients only), LH, FSH, estradiol, thyroid stimulating hormone (TSH) and free thyroxine (T4), amylase, and lipase.

Blood or urine samples will be collected from female patients at the visits specified in Appendix 10.2 for a pregnancy test before study drug administration.

4.7.3. Vital Signs

The following vital signs parameters will be collected at each clinic visit from screening to end of study (EOS) or at the time of early termination (ET):

- Pulse rate (beats/min)
- Systolic and diastolic blood pressures (sitting, mmHg)
- Respiratory rate (breaths/min)
- Body temperature (°F)

4.7.4. 12-Lead ECG

Standard 12-Lead ECG parameters include: Ventricular HR, PR interval, QRS interval, corrected QT interval (QTc Fridericia [QTcF] =QT/[RR0.33] and QTc Bazett [QTcB]=QT/[RR0.5]), ECG status: normal, abnormal not clinically significant or abnormal clinical significant. Electrocardiograms should be performed before blood is drawn during visits requiring blood draws. A standard 12-lead ECG will be performed at scheduled time points indicated in Appendix 10.2.

4.7.5. Physical Examination

The physical examination variable values are dichotomized to normal and abnormal.

A thorough and complete physical examination will be performed at time points according to the schedule in Appendix 10.2.

4.8. Pharmacokinetics

The PK variable is the concentration of functional drug in serum at scheduled time points.

4.9. Total Target

The total target variable is the concentration of total target in serum at nominal and actual time points.

4.10. Anti-Drug Antibody Variables

Anti-drug antibody (ADA) variables are ADA status, titer, and time-point/visit. Samples for ADA assessment will be collected at the clinical visits as specified in Appendix 10.2.

5. STATISTICAL METHODS

There are at least 3 database locks planned for this study: one for each of the study periods. The database lock for Period 1 occurred after the last patient completed the week 28 visit. The analyses of primary and secondary efficacy endpoints for Period 1 are considered final.

The database lock for Period 2 will occur after the last patient completes the week 56 visit. Following the Period 2 lock, the analyses of Period 2 endpoints will be conducted. In the context of FOP, an ultra-rare disease, an alpha of 0.10 (two-sided) will be used for the Period 2 analysis. An alpha of 0.10 will also mitigate any loss of power due to missing week 56 scans due to the COVID-19 pandemic.

The final database lock will occur when the last patient completes the last visit of the study (Period 3). Following the final database lock, final efficacy and safety analyses for Period 3 will be conducted.

For continuous variables, descriptive statistics will include the following: the number of subjects reflected in the calculation (n), mean, median, standard deviation, Q1, Q3, minimum, and maximum. For categorical or ordinal data, frequencies and percentages will be displayed for each category.

5.1. Demographics and Baseline Characteristics

Demographic and baseline characteristics will be summarized descriptively by treatment group and overall group based on AHOC, AHO and FAS. Listing of demographics and baseline characteristics will be presented.

5.2. Medical History

Medical history will be summarized by primary SOC and PT for each treatment group for FAS. The table will be sorted by decreasing frequency of SOC followed by PT based on the overall incidence across treatment groups.

Medical history will be listed based on the SAF.

5.3. Prior/Concomitant Medications

Number and proportion of subjects taking prior medication will be summarized by treatment group and overall patients in the double-blind safety population, sorted by decreasing frequency of ATC Level 2 and ATC level of overall group.

All concomitant medications during the double-blind period will be descriptively summarized by treatment group and overall in double-blind safety population, sorted by the decreasing frequency of ATC Level 2 and ATC level 4 of REGN2477 group.

For the entire study, concomitant medications will be summarized similarly by overall group as well as by the treatment group of double-blind period (actual treatment received).

5.4. Subject Disposition

The following summaries by each treatment group and overall group will be provided for double-blind period using FAS:

- The number of patients screened (signed the informed consent form) and the reasons for screen failure
- The number of patients randomized (received a randomization number)
- The total number of patients who completed the double-blind period
- The total number of patients who discontinued the study treatment during the double-blind period and the reasons for the discontinuation
- Number of patients who entered open label
- Number of patients who entered in the follow-up treatment period

For the open label period, the following summaries by overall group and each treatment group of double-blind period in open label safety analysis set

- The number of patients receiving open label treatment
- The number of patients completing the open label period
- The total number of patients ongoing in the open label period (applicable for 28 Week analysis)
- The total number of patients who prematurely discontinued study treatment during open label period and the reasons for the discontinuation

The total number of patients from open label period who complete the follow-up visit.

In addition, summary of number (and percentage) of patients in each analysis set will be provided by treatment group and overall group.

5.5. Measurement of Treatment Compliance

The compliance with protocol-defined study drug will be assessed for each patient for double-blind period and open-label period separately as follows:

Treatment Compliance = [(Number of study drug infusions received)/(Number of planned study drug infusions or until patient discontinuation in that period)] x 100%

The treatment compliance will be descriptively summarized. In addition, the number (%) of patients with treatment compliance <95% and ≥95% will be presented.

A Summary of study drug administration will be provided for the double-blind period by treatment group and will include the number of study drug doses administered.

A similar summary will be provided for overall group and by treatment group in the open label period.

A patient listing will be provided for those patients with interrupted or incomplete infusions. Listing of dose administration: including date/time, study day, number of infusions, locations of infusions, dosing information and whether the total dose is administered for each dose will be presented.

Listings of the subjects who received treatment different from the randomized treatment or stopped treatment due to an AE, or overdose will be provided.

5.6. Treatment Exposure and Observation Period

The duration of treatment during the double-blind period will be assessed and summarized by treatment group for patients in double-blind safety analysis set. The duration in weeks is calculated as:

((Date of last study drug infusion in double-blind period – date of first study drug infusion in double-blind period) + 28)/7

The duration of treatment during the open label period will similarly be assessed and summarized for all patients and by treatment group of the double-blind period for patients in open label safety analysis set.

In addition, the duration of REGN2477 treatment combining all periods will be assessed for all patients who received REGN2477. Note: exposure will be calculated based on the last study drug infusion date and first study drug infusion date in respective periods, regardless of temporary dosing interruption.

The duration of exposure during the study will be summarized using number of patients, number of infusions, duration of treatment using means, SD, minimums, Q1, medians, Q3, and maximums.

In addition, the duration of exposure will be summarized categorically by counts and percentages for each of the following categories as well:

≥1 week, ≥ 4 weeks, ≥ 8 weeks, ≥ 12 weeks, ≥ 16 weeks, ≥ 20 weeks, ≥ 24 weeks, ≥ 28 weeks, ≥ 32 weeks, ≥ 36 weeks, ≥ 40 weeks, ≥ 44 weeks, ≥ 48 weeks, ≥ 52 weeks, ≥ 56 weeks, ≥ 60 weeks, ≥ 64 weeks, ≥ 72 weeks, ≥ 76 weeks.

The duration of observation period will be similarly reported, is calculated in weeks as follows:

([Date of the last visit of study – date of the randomization] +1)/7.

The duration of observation period will be summarized descriptively as a quantitative data (n, mean, SD, median, Q1 and Q3, minimum and maximum). In addition, the number (%) of subjects with observation periods will be presented by specific time periods. The time periods of interest are specified as: ≥ 1 week, ≥ 4 weeks, ≥ 8 weeks, ≥ 12 weeks, ≥ 16 weeks, ≥ 20 weeks, ≥ 24 weeks, ≥ 28 weeks, ≥ 32 weeks, ≥ 36 weeks, ≥ 40 weeks, ≥ 44 weeks, ≥ 48 weeks, ≥ 52 weeks, ≥ 56 weeks, ≥ 60 weeks, ≥ 64 weeks, ≥ 72 weeks, ≥ 76 weeks, ≥ 80 weeks, ≥ 84 weeks, ≥ 88 weeks, ≥ 92 weeks.

5.7. Analysis of Efficacy Variables

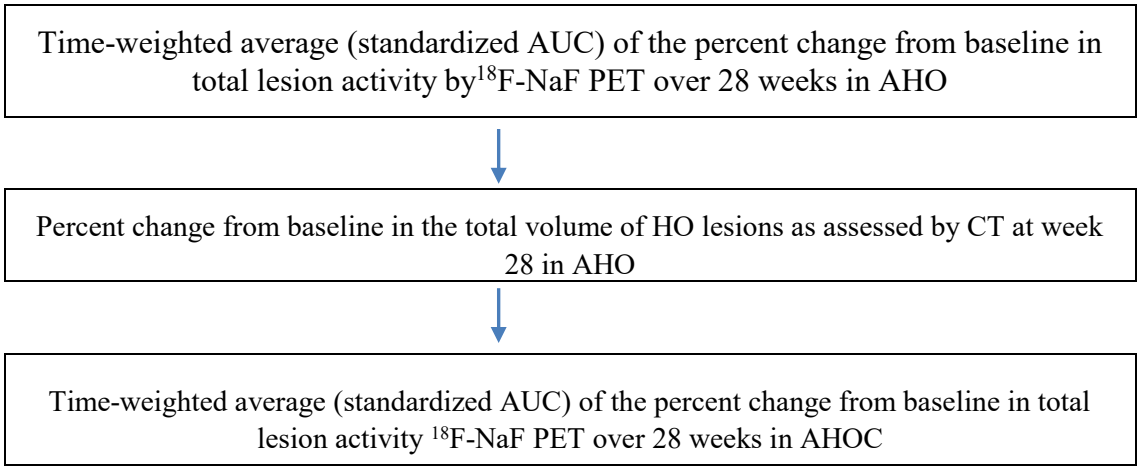
5.7.1. Hypothesis test and Multiplicity Control (Period 1)

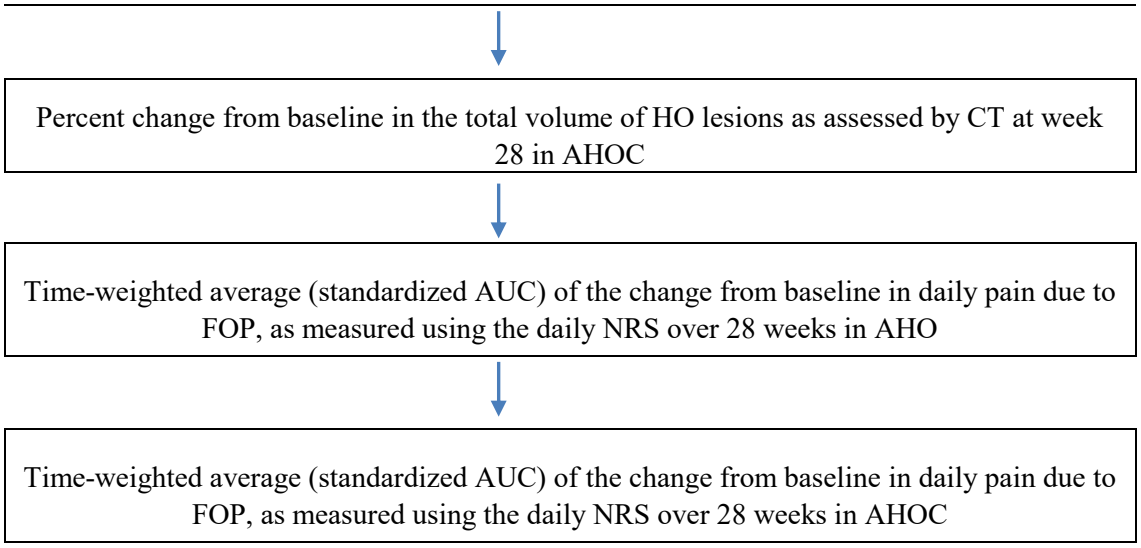
Let μ_0 and μ_1 be the population means of one of the primary and key secondary endpoints under placebo and REGN2477, respectively. The hypothesis that will be tested is "H0: $\mu 0 = \mu 1$" versus "H1: $\mu 0 \neq \mu 1$".

The analyses of efficacy variables are described in the subsections below and summarized in Appendix 10.1.

Multiplicity Considerations

To control the type-I error rate for the primary and key secondary endpoints, a hierarchical testing procedure will be applied at a 2-sided 5% significant level, key secondary efficacy endpoints will be tested only if statistical significance is established for all primary endpoints. The order of testing sequence for primary and key secondary efficacy endpoints is as follows:





No further adjustments will be made for other secondary and exploratory endpoints, for which nominal p-values will be provided for descriptive purpose only.

5.7.2. Hypothesis Test and Multiplicity Control (Period 2)

For the efficacy analyses in Period 2, the key statistical comparisons will be based on within group (Placebo/REGN2477) comparisons for the patients who were randomized to placebo in the double-blind period and switch to REGN2477 in the open label period.

For the Period 2 analyses, the following primary and key secondary null hypotheses will be tested:

- $H_{1,1}$: There is no difference in number of new (relative to week 28 scan) HO lesions as assessed by CT at week 56 compared to the number of new (relative to baseline scan) HO lesions as assessed by CT at week 28 in Placebo/REGN2477 group in AHO (primary)

- $H_{2,1}$: There is no difference in total volume of new (relative to week 28 scan) HO lesions as assessed by CT at week 56 compared to the total volume of new (relative to baseline scan) HO lesions as assessed by CT at week 28 in Placebo/REGN2477 group in AHO (key secondary)

- $H_{2,2}$: There is no difference in number of new (relative to week 28 scan) HO lesions by ^{18}F-NaF PET at week 56 compared to the number of new (relative to baseline scan) HO lesions by ^{18}F-NaF PET at week 28 in Placebo/REGN2477 group in AHO (key secondary)

- $H_{2,3}$: There is no difference in total lesion activity by ^{18}F-NaF PET in new (relative to week 28 scan) HO lesions at week 56 compared to the total lesion activity by ^{18}F-NaF PET in new (relative to baseline scan) HO lesions at week 28 in Placebo/REGN2477 group in AHO (key secondary)

- $H_{2,4}$: There is no difference in percent of patients with new HO lesions by CT at week 56 relative to week 28 scan in Placebo/REGN2477 group in AHO (key secondary)

- $H_{2,5}$: There is no difference in percent of patients with new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to week 28 scan in Placebo/REGN2477 group in AHO (key secondary)

Multiplicity Considerations

The type-I error rate will be controlled at 0.10 for the primary and key secondary null hypotheses in Period 2. To control the type-I error rate for the primary and key secondary endpoints in Period 2, a hierarchical testing procedure will be applied at a 2-sided 10% significance level to test the above hypotheses in the specified order. The key secondary efficacy endpoints will be tested only if statistical significance is established for the primary endpoint at 10% significance level.

No further adjustments will be made for other secondary and exploratory endpoints in Period 2, for which estimates, 95% CI, and/or nominal p-values will be provided for descriptive purpose.

The analyses of efficacy variables are described in the subsections (5.7.4).

5.7.3. Analysis of Efficacy Variables (Period 1)

Major methods for analyzing efficacy endpoints in the double-blind period are summarized below:

ANCOVA

Analysis of non-longitudinal efficacy endpoints will be conducted using ANCOVA. The model will include treatment, gender, baseline AHO status and ACVR1 mutation type as fixed effect, and the baseline value as continuous covariate.

Difference in LS mean change from baseline, the corresponding 95% CI and the p-value will be provided from ANCOVA model for comparison for REGN2477 group against placebo group.

Ranked ANCOVA

A non-parametric ranked analysis of covariance (ANCOVA) will be performed with ranked outcome variable. The model will include treatment, gender, baseline AHO status and ACVR1 mutation type and the ranked baseline value as continuous covariate.

Negative Binomial Model

The number of HO lesions assessed by ^{18}F-NaF PET and CT through 28 weeks will be analyzed using negative binomial model. The model will include treatment, gender, baseline AHO status, ACVR1 mutation type, and number of baseline active HO lesions. Log transformed number of scans over 28 weeks will be the offset variable.

The estimated mean number of lesions per scan during the 28 weeks for each treatment group and the two-sided 95% confidence intervals will be derived from the model. The ratio of mean number of lesions per scan for REGN2477 and placebo, two-sided 95% confidence interval of the ratio and p-value will be provided.

Mixed-effect model with repeated measures (MMRM)

Analysis of longitudinal efficacy endpoints will be conducted using MMRM. The MMRM model will include treatment, gender, baseline AHO status and ACVR1 mutation type as fixed effect, time point, and treatment-by-time point interaction, as well as, the continuous fixed covariates of baseline value. The MMRM model will be run using SAS Mixed procedure with an unstructured covariance to model the within-patient errors. Parameters will be estimated using restricted maximum likelihood method with the Newton-Raphson algorithm. Denominator degrees of freedom will be estimated using Kenward-Roger's approximation.

If the MMRM model fails to achieve convergence due to the complexity of model specifications, different estimation algorithms will be applied following the order: unstructured covariance → heterogeneous autoregressive (1) covariance → autoregressive (1) covariance →compound symmetry →using ANCOVA.

Difference in LS mean change from baseline, the corresponding 95% CI and the p-value will be provided from the MMRM model for comparison for REGN2477 group against placebo group.

The efficacy endpoints may be analyzed for one of the three analysis sets (AHOC, AHO, FAS). Table below lists the analysis set specific factors for each of the analysis set when different analysis method is utilized.

Analysis Set	**Analysis set specific factors**
AHOC	treatment, and gender
AHO	treatment, gender and ACVR1 mutation type
FAS	treatment, gender, baseline active AHO status and ACVR1 mutation type
Model type	**General terms in the model**
MMRM	time point, and treatment-by-time point interaction, as well as, the continuous fixed covariates of baseline value.
ANCOVA	baseline value as continuous covariate
Negative Binomial Model	Number of baseline active HO lesions, log transformed number of scans as an offset variable
Ranked ANCOVA	the rank of baseline value as continuous covariate

In case of substantial imbalance at baseline in the number of patients in the randomization strata: gender, AHO status, ACVR1 genotype, strata factor may be excluded from the model.

The stratification factors based on CRFs or imaging data will be used in the analyses.

5.7.3.1. Primary Endpoints

Time-weighted average (standardized AUC) of the percent change from baseline in total lesion activity by ^{18}F-NaF PET over 28 weeks in AHO

The AUC of the percent change from baseline in total lesion activity by ^{18}F-Naf PET over 28 weeks will be calculated for each patient. Nominal time (i.e. Wk 8 and Wk 28) and not the actual time will be used in the calculation of the AUC. If both imaging scans at week 8 and 28 are missing, then mean of percent changes of placebo group at week 8 and 28 will be used for imputation. If imaging scan for only week 8 is missing, linear interpolation of % change between the baseline and week 28 will be used to calculate AUC. If imaging scan for week 28 is missing, % change at week 8 will be carried forward to week 28 for calculating AUC. Time weighted average (standardized AUC, i.e. AUC divided by 28) of percent change from baseline will be calculated for each patient as the AUC of percent change from baseline in total lesion activity by ^{18}F-Naf PET over 28 weeks divided by 28. The analysis of covariance (ANCOVA) model will be used to analyze time weighted average of percent change over 28 weeks. The model will include treatment, gender, ACVR1 mutation type (classic, non-classic) and baseline total lesion activity as a covariate. In case of substantial imbalance at baseline in the number of patients in the randomization strata: gender, ACVR1 mutation type, strata factor(s) may be excluded from the model. Difference in LS mean change from baseline, the corresponding 95% CI and the p-value will be provided from ANCOVA model for comparison for REGN2477 group against placebo group.

Normality assumption will be checked and a nonparametric ranked analysis of covariance (ANCOVA) will be conducted as a supportive analysis, as needed. The ranked time weighted average of percent change from baseline will be used as the outcome variable. The model will include treatment, gender, ACVR1 mutation type (classic, non-classic) and the ranked baseline total lesion activity value as the covariate. In case of substantial imbalance at baseline in the number of patients in the randomization strata: gender, ACVR1 mutation type, strata factor(s) may be excluded from the model.

As a supplementary analysis, MMRM model based on the percent change from baseline in total lesion activity by ^{18}F-NaF PET as a dependent variable will be used. The fixed effects in the model will include independent variables of treatment, gender, ACVR1 mutation type (classic, non-classic), visit (Week 8, and 28), treatment-by-visit interaction and baseline total lesion activity. As specified above, in case of substantial imbalance at baseline in the number of patients in the randomization strata: gender, ACVR1 mutation type, strata factor(s) may be excluded from the model. An unstructured covariance will be used to account for within-patient correlation between time. The restricted Maximum Likelihood (REML) method will be used for estimates of variance components. Denominator degrees of freedom will be estimated using Kenward-Roger's approximation. Parameter estimates at Week 8 and Week 28 from the MMRM model, will be used to estimate the AUC for each treatment group (using Trapezoidal rule). Time weighted average over 28 weeks will be calculated by dividing the AUC estimates by 28. Difference in standardized AUC estimates, the corresponding 95% CI and the p-value will be provided.

Additional analyses will be conducted as deemed appropriate.

Percent change from baseline in the total volume of HO lesions as assessed by CT at week 28 in AHO

Percent change from baseline in the total volume of HO lesions as assessed by CT at week 8 and 28 weeks will be analyzed in AHO analysis set using a MMRM model. Model contains treatment, gender, ACVR1 mutation type (classic, non-classic), visit (Week 8, and 28), baseline total volume and treatment-by-visit interaction. In case of substantial imbalance at baseline in the number of patients in the randomization strata: gender, ACVR1 mutation type, strata factor(s) may be excluded from the model. An unstructured covariance will be used to account for within-patient correlation between time. Difference in LS mean change from baseline, the corresponding 95% CI and the p-value will be provided from MMRM model for comparison for REGN2477 group against placebo group.

A sensitivity analysis will be performed using ANCOVA model. Percent change from baseline to Week 28 in total volume of HO lesions as assessed by CT will be used as dependent variable. The post-baseline last observation carried forward (LOCF) method will be used to impute missing values at week 28. If both measurements at week 8 and 28 are missing, then mean of percent changes of placebo group at 28 will be used for imputation. Model contains treatment, gender, ACVR1 mutation type (classic, non-classic) and baseline total volume as covariate. In case of substantial imbalance at baseline in the number of patients in the randomization strata: gender, ACVR1 mutation type, strata factor(s) may be excluded from the model. Difference in LS mean change from baseline, the corresponding 95% CI and the p-value will be provided from ANCOVA model for comparison for REGN2477 group against placebo group. This parametric ANCOVA analysis will also serve as a primary analysis if use of MMRM as described above substantially violate the assumption of normal distribution of the residuals.

If normality assumption is not valid with the MMRM and parametric ANCOVA above, p-values will be based on a nonparametric ranked analysis of covariance (ANCOVA) with ranked percent change from baseline to Week 28 in total volume of HO lesions as assessed by CT as a dependent variable. The model will include treatment, gender, ACVR1 mutation type (classic, non-classic) and the ranked baseline total volume of HO lesions as the covariate. In case of substantial imbalance at baseline in the number of patients in the randomization strata: gender, ACVR1 mutation type, strata factor(s) may be excluded from the model.

Time-weighted average (standardized AUC) of the percent change from baseline in total lesion activity by ^{18}F-NaF PET over 28 weeks in AHOC

Time-weighted average (standardized AUC) of the percent change from baseline in total lesion activity by ^{18}F-NaF PET over 28 weeks in AHOC will be analyzed using the same method as for first primary endpoint in AHO.

Percent change from baseline in the total volume of HO lesions as assessed by CT at week 28 in AHOC

Percent change from baseline in the total volume of HO lesions as assessed by CT at week 28 in AHOC will be analyzed in the same method as for second primary endpoint (in AHO).

5.7.3.2. Secondary Key Efficacy Endpoints

Time-weighted average (standardized AUC) change from baseline in daily pain due to FOP, as measured using the daily NRS over 28 weeks in AHO

Time weighted average of change from baseline in daily pain scores (average of current pain, worst pain, and least pain) will be calculated for each patient. The baseline daily pain score is defined as the latest measurement before the first study drug infusion. If the pain score for intermediate days are missing, linear interpolation of change between the two adjacent measurements will be used to calculate time weighted average. If missing monotonically, the post-baseline last observation carried forward (LOCF) method will be used to impute missing values. Time-weighted average change in daily pain over week 28 will be analyzed in AHO using the ANCOVA model.

Time-weighted average (standardized AUC) change from baseline in daily pain due to FOP, as measured using the daily NRS over 28 weeks in AHOC

Time-weighted average (standardized AUC) change from baseline in daily pain due to FOP, as measured using the daily NRS over 28 weeks in AHOC will be analyzed in the same methods as previous key secondary endpoint (in AHO).

5.7.3.3. Other Secondary Efficacy Endpoints

Percent change from baseline in ^{18}F-NaF SUVmax of individual active HO site(s) by PET at week 8 in AHOC and AHO

Summary statistics for percent change of ^{18}F-NaF SUVmax will be provided on lesion level.

The patient level percent change in SUVmax will also be reported. Mean SUVmax at baseline and Week 8 will be calculated (sum of SUVmax for all lesions divided by the number of lesions). ANCOVA model will be used to analyze percent change from baseline to Week 8 in mean SUVmax. Model will include treatment, gender, ACVR1 mutation type (classic, non-classic) and baseline mean SUVmax as covariate. In case of substantial imbalance at baseline in the number of patients in the randomization strata: gender, ACVR1 mutation type, strata factor(s) may be excluded from the model. Difference in LS mean change from baseline, the corresponding 95% CI and the p-value will be provided from ANCOVA model for comparison for REGN2477 group against placebo group.

Change from baseline in number of HO lesions as assessed by 18F-NaF PET at week 28 in AHOC, AHO and FAS

Change from baseline in number of HO lesions as assessed by ^{18}F-NaF PET to week 28 will be analyzed using ranked ANCOVA. In addition, the number of HO lesions assessed by ^{18}F-NaF PET through week 28 will be analyzed using Negative Binomial model. This analysis will be done in AHOC, AHO and FAS.

Change from baseline in number of HO lesions detectable by CT at week 28 in AHOC, AHO and FAS

Change from baseline in umber of HO lesions as assessed by CT to week 28 will be analyzed using ranked ANCOVA. The number of HO lesions detectable by CT through week 28 will be analyzed using Negative Binomial model. This analysis will be done in AHOC, AHO and FAS.

Time-weighted average (standardized AUC) change from baseline in daily pain due to FOP, as measured using the daily NRS over 28 weeks in FAS

Time weighted average of change of daily pain scores (current pain, worst pain, and least pain, average pain) will be calculated for each patient. If the pain score for intermediate days are missing, linear interpolation of change between the two adjacent measurements will be used to calculate time weighted average. If the missing is monotone, the last observation carried forward (LOCF) method will be used to impute missing values for calculation of time weighted average over week 28.

Time-weighted average change in daily pain over week 28 will be analyzed in FAS using the ANCOVA model.

Time weighted average (standardized AUC) percent change from baseline in biomarkers of bone formation levels in serum over 28 weeks in FAS

Time weighted average of percent change from baseline in biomarkers of bone formation will be calculated for each patient. If the biomarkers for intermediate weeks are missing, linear interpolation of percent change between the two adjacent measurements will be used to calculate time weighted average. If the missing is monotone, the last observation carried forward (LOCF) method will be used to impute missing values for calculation of time weighted average over week 28.

Time-weighted average percent change in biomarkers of bone formation will be analyzed in the FAS using the ANCOVA model.

Percent change in biomarkers of bone formation will also be analyzed using ranked analysis of covariance (ANCOVA). This non-parametric approach is robust with concerns of departure from the normal distribution of biomarker data. The ranked ANCOVA model will be able to provide significance of the treatment effect as compared to placebo. Median treatment difference and 95% CI can be derived from the Hodges-Lehmann estimation and Moses distribution free CI respectively.

5.7.4. Analysis of Efficacy Variables (Period 2)

Overview

In the Placebo/REGN2477 group (placebo patients who switch to REGN2477 after the double-blind period), efficacy analyses will focus on the within group statistical comparisons of Period 2 vs Period 1.

Estimates of the treatment effects with 95% confidence interval on the endpoints related to new HO lesions from Period 1 (new relative to baseline scan) in the REGN2477/REGN2477 group patients will be descriptively compared with the estimates from Period 2 (new relative to week 28) in the Placebo/REGN2477 group.

Persistence of efficacy of REGN2477 in preventing the formation of new heterotopic will be assessed in the REGN2477/REGN2477 group (patients who continue receiving REGN2477 beyond week 28). The estimates of treatment effects at week 56 will be provided with 95% confidence interval on the endpoints related to new HO lesions (new relative to baseline scan). No formal statistical comparisons of Period 2 vs Period 1 will be performed. The estimation methods will be the same as described for the Placebo/REGN244 group in respective endpoints.

Analysis Sets (Period 2)

- Intent-to-Treat set (ITT): All AHO patients who receive a treatment in Period 2 and have a post-week 28 scan regardless of delayed/missed infusion.

- COVID-19 Modified Intent-to-Treat set (COVID-19 mITT): All AHO patients who receive a treatment in Period 2 for whom at least 1 post-Week 28 scan is collected and the period between any consecutive doses is less than 9 weeks (63 days) before the 1st post-week 28 scan.

The COVID-19 pandemic has affected the conduct of the study mainly by causing a delay in dose administration, and a delay in the collection of the PET/CT scans due to site closures and unavailability of ^{18}F-NaF due to interruption of tracer production. Some patients were not able to complete their week 56 scans within the protocol-specified visit window and some were only able to complete their scans using CT only modality. In order to mitigate the confounding effects of the pandemic on the study outcomes, the estimand framework will be applied with the COVID-19 mITT principle to handle intercurrent events.

COVID-19 mITT will be the primary efficacy set for all imaging endpoints. This will ensure that primary main analyses for all imaging endpoints will include only evaluable patients whose treatment is not greatly impacted due to the pandemic. Sensitivity analyses will be performed on the ITT set.

Delayed and Missing Scan Handling:

For patients whose week 56 scans are delayed or missed due to the pandemic, the first available scan after the week 28 scan will be used to impute the week 56 data in the primary analyses. In the absence of a PET/CT scan, data from the first available CT only scans after the week 28 scan will be used to analyze the primary endpoint and other secondary imaging endpoints as applicable. For missing Week 56 scans solely due to the COVID-19 pandemic in Placebo/REGN2477 patients, the assumption of 'missing completely at random mechanism' is reasonable.

Detailed listings of missed or delayed infusions/scans and the reasons will be provided. These will be based on the information provided and confirmed by the investigators.

Analysis Methods and Strategy

The primary and key secondary endpoints are related to the formation of new HO lesions. The frequency of new HO lesions may be small, however. The statistical methods and inference for analyzing the primary and the key secondary endpoints need to consider the effects of the sparseness and non-normality of the data. The inference for each endpoint, therefore, will be based on a comprehensive evaluation which will include patient level data, descriptive summary, graphical presentations, and suitable analysis methods. Graphical presentations such as waterfall plots of the number, volume or total lesion activity of new HO lesions for each patient in Period 2 and Period 1 will be provided. Non-parametric methods will be used for primary analyses as they do not require the assumption of Gaussian distribution. The parametric estimation methods, when applicable, will be based on non-Gaussian models (e.g. repeated negative binomial model). Some statistical methods that are model-based and/or use Gaussian distributional assumption may not work or may provide inadequate results. Estimates based on bootstrap methods, therefore, will be used to aid in interpretations. The key analyses methods and evaluation for the primary and key secondary efficacy endpoints are summarized in the following table.

Endpoint	Primary Analysis (COVID-19 mITT)	Sensitivity Analyses (ITT)
Primary: Number of new HO lesions as assessed by CT at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)	• Descriptive summary, aggregate and patient level graphical presentation for comparison of Period 2 and Period 1 • Hypothesis testing: Wilcoxon signed-rank test • Estimation: Descriptive and parametric using repeated measures negative binomial model (GEE)	• Descriptive summary, aggregate and patient level graphical presentation for comparison of Period 2 and Period 1 • Hypothesis testing: Wilcoxon Signed Rank test • Estimation**:** Descriptive and Parametric using repeated negative binomial model (GEE) • Multiple imputation using fully conditional specification (FCS) with predictive mean matching (PMM) method, Rubin's formulae with repeated negative binomial (GEE)
Total volume of new HO lesions as assessed by CT at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)	• Descriptive summary, aggregate and patient level graphical presentation comparing Period 2 and Period 1 • Hypothesis testing: Wilcoxon signed-rank test • Estimation: Descriptive, bootstrap methods and parametric using MMRM	• Descriptive summary, aggregate and patient level graphical presentation for comparison of Period 2 and Period 1 • Hypothesis testing: Wilcoxon Signed Rank test • Estimation**:** Descriptive, bootstrap and/or Parametric method using MMRM
Number of new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)	• Descriptive summary, aggregate and patient level graphical presentation for comparison of Period 2 and Period 1 • Hypothesis testing: Wilcoxon signed-rank test • Estimation: Descriptive and parametric using repeated measures negative binomial model (GEE)	• Descriptive summary, aggregate and patient level graphical presentation for comparison of Period 2 and Period 1 • Hypothesis testing: Wilcoxon Signed Rank test • Estimation**:** Descriptive and Parametric using repeated negative binomial model (GEE) • Multiple imputation using fully conditional specification (FCS) with predictive mean matching (PMM) method, Rubin's formulae

		with repeated negative binomial (GEE)
Total lesion activity by ^{18}F-NaF PET in new lesions at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)	• Descriptive summary, aggregate and patient level graphical presentation comparing Period 2 and Period 1 • Hypothesis testing: Wilcoxon signed-rank test • Estimation: Descriptive, bootstrap methods and parametric using MMRM	• Descriptive summary, aggregate and patient level graphical presentation for comparison of Period 2 and Period 1 • Hypothesis testing: Wilcoxon Signed Rank test • Estimation**:** Descriptive, bootstrap and/or Parametric method using MMRM
Percent of patients with new HO lesions as assessed by CT at week 56 relative to week 28 (in patients switching from placebo to REGN2477 after double-blind period) (AHO)	• Descriptive summary, aggregate and patient level graphical presentation for comparison of Period 2 and Period 1 • Hypothesis testing: McNemar's test • Estimation: Descriptive and parametric using repeated measures logistic regression model (GEE)	• Descriptive summary, aggregate and patient level graphical presentation for comparison of Period 2 and Period 1 • Hypothesis testing: McNemar's test • Estimation**:** Descriptive and Parametric using repeated measures logistic regression model (GEE) • Multiple imputation using fully conditional specification (FCS) with predictive mean matching (PMM) method, Rubin's formulae with repeated measures logistic regression model (GEE)
Percent of patients with new HO lesions assessed by^{18}F-NaF PET at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)	• Descriptive summary, aggregate and patient level graphical presentation for comparison of Period 2 and Period 1 • Hypothesis testing: McNemar's test • Estimation: Descriptive and parametric using repeated measures logistic regression model (GEE)	• Descriptive summary, aggregate and patient level graphical presentation for comparison of Period 2 and Period 1 • Hypothesis testing: McNemar's test • Estimation**:** Descriptive and Parametric using repeated measures logistic regression model (GEE) • Multiple imputation using fully conditional specification (FCS) with predictive mean matching (PMM) method, Rubin's formulae with repeated measures logistic regression model (GEE)

The details of the analysis methods and imputations are provided in the following Sections.

5.7.4.1. Primary Efficacy Endpoint

Number of new HO lesions as assessed by CT at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

Primary analysis:

Descriptive summary of the number of new HO lesions as assessed by CT at week 56 relative to week 28 scan will be summarized for the Placebo/REGN2477 group (COVID-19 mITT set). Wilcoxon signed rank test will be used for the within group comparison in the Placebo/REGN2477 group to compare the number of new HO lesions as assessed by CT at week 56 (relative to week 28 scan) with the number of new HO lesions as assessed by CT at week 28 (relative to baseline scan).

The estimate and the corresponding 95% confidence interval of the rate of new HO lesions as assessed by CT at Week 56 and that of the rate ratio (comparing Period 2 vs. Period 1) will be based on a negative binomial model with repeated measures. The Generalized Estimating Equation (GEE) method will be implemented. The response variable will be the number of new HO lesions as assessed by CT at week 28 relative to the baseline scan and the number of new HO lesions as assessed by CT at week 56 relative to the week 28 scan. The model will include visit (week 28 and week 56) and baseline total number of HO lesions by CT as a covariate. An unstructured covariance will be used to account for within-patient correlation.

Sensitivity analyses:

- Sensitivity analyses will be conducted based on all Placebo/REGN2477 patients in the ITT set using the analysis methods as described above for the primary analysis.

- A sensitivity analysis will be conducted using a multiple imputation approach in all Placebo/REGN2477 patients in the ITT set. For patients whose first available scan post week 28 scan is collected beyond the SAP specified analysis window (i.e. >study day 463), the number of new HO lesions at week 56 will be imputed using a fully conditional specification (FCS) with the predictive mean matching (PMM) method. The imputation model will include the total number of HO lesions detectable by CT at baseline, week 8 and week 28. A total of 100 imputed datasets will be generated and the number of closest observations to be used will be 3 in the PMM. The GEE method described previously will be applied to each of the imputed dataset. The log of the rate ratio and the standard error will be estimated using a contrast. These estimates will then be combined using Rubin's formulae to obtain the final estimate of log(rate ratio), 95% CI and the associated p-value. The rate ratio and 95% CI will be derived by taking the exponential of these estimates.

- Since the abnormal bone that is formed at an earlier timepoint is likely to be apparent on the later CT scans also, an additional sensitivity analysis will be conducted using multiple imputation approach as described above in the Placebo/REGN2477 group (ITT set) but will restrict the imputed datasets in which the imputed values do not exceed the number of new lesions on the later scan.

Supportive/Descriptive Analyses:

- Descriptive statistics for the number of new HO lesions as assessed by CT at week 56 relative to the week 28 scan will be provided separately for 3 groups of patients: a) those whose total number of HO lesions by CT decreased from baseline to week 28, b) those that did not change between baseline and week 28 and c) those patients that increased from baseline to week 28.

5.7.4.2. Key Secondary Efficacy Endpoints

Total volume of new HO lesions as assessed by CT at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

Primary analysis:

The total volume associated with new HO lesions as assessed by CT at week 56 relative to the week 28 scan will be descriptively summarized for the Placebo/REGN2477 group (COVID-19 mITT set). The Wilcoxon signed rank test will be used for the within group comparison in the Placebo/REGN2477 group to compare the total volume at week 56 associated with new HO lesions as assessed by CT (relative to week 28 scan) with the total volume associated with new HO lesions assessed by CT at week 28 (relative to baseline scan).

To estimate the difference between Period 2 and Period 1, a MMRM model will be implemented. The response variable will be the total volume of new lesions at week 56 (relative to the week 28 scan) and at week 28 (relative to the baseline scan). The model will include visit (week 28 and week 56) and the total volume by CT at baseline as a covariate. An unstructured covariance will be used to account for within-patient correlation between visits. LS mean of differences in total volume of new lesions and the corresponding 95% CI will be provided for comparison for week 56 against week 28.

Sensitivity analyses:

- Sensitivity analyses will be performed using all Placebo/REGN2477 patients in the ITT set. Analysis and estimation methods as described above for the primary analysis will be used.

- Additional sensitivity analysis will be performed using all Placebo/REGN2477 patients in the ITT set by considering the week 56 efficacy data as missing for patients whose first available scan post week 28 scan is beyond the SAP defined analysis window (i.e. >study day 463). The same MMRM model as described in the primary analysis will be used. LS means, 95% CI and p-value will be provided from this model.

Descriptive Analyses:

- Descriptive statistics for the total volume of new HO lesions as assessed by CT at week 56 relative to week 28 scan will be provided separately for two groups of patients: (a) those with increase and (b) those with decrease; in total volume in all lesions by CT from baseline to week 28.

Number of new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

Primary analysis:

Descriptive summary of the number of new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to week 28 scan will be provided for the Placebo/REGN2477 group (COVID-19 mITT set). The Wilcoxon signed rank test will be used for the within group comparison in the Placebo/REGN2477 group to compare the number of new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to week 28 scan with the number of new HO lesions as assessed by ^{18}F-NaF PET at week 28 relative to baseline scan.

The estimate and the corresponding 95% confidence interval of the rate of new HO lesions as assessed by ^{18}F-NaF PET at Week 56 and that of rate ratio (comparing Period 2 vs. Period 1) will be based on a negative binomial model with repeated measures. Generalized Estimating Equation (GEE) method will be implemented. The response variable will be number of new HO lesions as assessed by ^{18}F-NaF PET at week 28 relative to baseline scan and the number of HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to the week 28 scan. The model will include visit (week 28 and week 56) and the baseline total number of HO lesions by ^{18}F-NaF PET as a covariate. An unstructured covariance will be used to account for within-patient correlation.

In the absence of a PET/CT scan, the number of new HO lesions as assessed by CT only will be used for the imputation of new HO lesions by ^{18}F-NaF PET.

Sensitivity analyses:

- Sensitivity analyses will be conducted based on all Placebo/REGN2477 patients in the ITT set using the analyses methods as described above.

- Additional sensitivity analysis will be conducted based on all Placebo/REGN2477 patients in the ITT set using a similar multiple imputation approach as described in Section 5.7.4.1 for the primary endpoint. The imputation model will include the total number of lesions by PET at baseline, week 8 and week 28.

Descriptive Analyses:

- Descriptive statistics for number of new HO lesions as assessed by PET at week 56 relative to week 28 scan will be provided separately for 3 groups of patients: a) those whose total number of HO lesions by PET decreased from baseline to week 28, b) those whose total number of HO lesions did not change between baseline and week 28 and c) those whose number of HO lesions increased from baseline to week 28.

Total lesion activity by ^{18}F-NaF PET in new lesions at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

Primary analysis:

The total volume associated with new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to week 28 scan will be descriptively summarized for the Placebo/REGN2477 group (COVID-19 mITT set). The Wilcoxon signed rank test will be used for the within group comparison in the Placebo/REGN2477 group to compare the total lesion activity associated with new HO lesions assessed by^{18}F-NaF PET at week 56 relative to the week 28 scan with the total lesion activity associated with new HO lesions assessed by ^{18}F-NaF PET at week 28 relative to the baseline scan.

To estimate the difference between Period 2 and Period 1, a MMRM model described previously will be implemented. The model will include visit (week 28 and week 56) and the total lesion activity by PET at baseline as a covariate.

In the absence of a PET/CT scan, the number of new HO lesions as assessed by CT only will be considered as new HO lesions by ^{18}F-NaF PET and the new lesion activity in each of these new lesions will be imputed using the median new lesion activity in new lesions from available PET/CT scans in the Placebo/REGN2477 group.

Sensitivity and descriptive analyses:

Similar sensitivity and descriptive analyses as described for the endpoint of total volume of new HO lesions as assessed by CT will also be performed.

Percent of patients with new HO lesions as assessed by CT at week 56 relative to week 28 (in patients switching from placebo to REGN2477 after double-blind period) (AHO)

Primary analysis:

The number and percent of patients with new HO lesions as assessed by CT at week 56 relative to week 28 will be provided for the Placebo/REGN2477 group (COVID-19 mITT set). The within group comparison of the percent of patients with new HO lesions as assessed by CT at week 56 relative to week 28 scan versus the percent of patients with new HO lesions as assessed by CT at week 28 relative to baseline scan will be performed using McNemar's test in Placebo/REGN2477 group.

A logistic regression model with repeated measures will be used to estimate the odds ratio and 95% CI to compare Period 2 and Period 1 using the generalized estimating equation (GEE). The model will include visit (week 28 and week 56), and the baseline total number of lesions by CT as a covariate. An unstructured covariance will be used to account for within-patient correlation between time points.

Sensitivity analyses:

- Sensitivity analyses will be conducted based on all Placebo/REGN2477 patients in the ITT set who received REGN2477 in Period 2 using the analyses methods as described above.

- Additional sensitivity analyses will be conducted based on all Placebo/REGN2477 patients in the ITT set who received REGN2477 in Period 2 using a multiple imputation approach. The same multiple imputation method (for scans that fall outside the analysis window of study day 463) as described above for the number of new HO lesions as assessed by CT will be implemented to obtain the binary observation on each patient ('presence or absence' of new HO lesions by CT at week 56 relative to week 28). The logistic regression using GEE method as described in the primary analysis will be applied to each of the imputed dataset. The log of the odds ratio and the standard error will be estimated using contrast. These estimates will then be combined using Rubin's formulae to obtain the final estimate of log(odds ratio), and the corresponding 95% CI and the associated p-value. The odds ratio and 95% CI can be derived by taking the exponential of these estimates.

Descriptive Analyses:

- Descriptive statistics for the percent of patients with new HO lesions by CT at week 56 relative to the week 28 scan will be provided separately for patients whose total number of lesions by CT decreased from baseline to week 28, did not change between baseline and week 28, and increased from baseline to week 28.

Percent of patients with new HO lesions assessed by^{18}F-NaF PET at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

Primary analysis:

The number and percent of patients with new HO lesions assessed by ^{18}F-NaF PET at week 56 relative to week 28 scan will be provided for the Placebo/REGN2477 group (COVID-19 mITT set). The within group comparison to compare the percent of patients with new HO lesions assessed by^{18}F-NaF PET at week 56 relative to the week 28 scan with the percent of patients with new HO lesions assessed by ^{18}F-NaF PET at week 28 relative to baseline scan will be performed using McNemar's test in Placebo/REGN2477 group.

The logistic regression with GEE approach as described for the endpoint of percent of patients with new HO lesions by CT will also be implemented. The model will include visit (week 28 and week 56) and the baseline total number of HO lesions by ^{18}F-NaF PET as a covariate.

In the absence of a PET/CT scan, CT only scans will be used for the imputation of the endpoint (i.e. patients with new HO lesions by CT only scans will be considered having new HO lesions by ^{18}F-NaF PET).

Sensitivity and descriptive analyses:

Similar sensitivity and descriptive analyses as described for the endpoint of percent of patients with new HO lesions by CT will also be performed.

5.7.4.3. Other Secondary Efficacy Endpoints

Number of new HO lesions as assessed by CT only at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

The number of new HO lesions as assessed by CT only at week 56 relative to week 28 scan will be summarized for the Placebo/REGN2477 group. The Wilcoxon signed rank test will be used for the within group comparison in the Placebo/REGN2477 group to compare the number of new HO lesions as assessed by CT only at week 56 (relative to week 28 scan) with the number of new HO lesions as assessed by CT only at week 28 (relative to baseline scan).

Total volume of new HO lesions as assessed by CT only at week 56 relative to week 28 scan (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

The total volume associated with new HO lesions as assessed by CT only at week 56 relative to the week 28 scan will be summarized for the Placebo/REGN2477 group. The Wilcoxon signed rank test will be used for the within group comparison in the Placebo/REGN2477 group to compare the total volume at week 56 associated with new HO lesions as assessed by CT only (relative to the week 28 scan) with the total volume associated with new HO lesions assessed by CT only at week 28 (relative to the baseline scan).

Percent of patients with new HO lesions as assessed by CT only at week 56 relative to week 28 (in patients switching from placebo to REGN2477 after double-blind period) (AHO)

The number and percent of patients with new HO lesions as assessed by CT only at week 56 relative to week 28 will be provided for the Placebo/REGN2477 group. The within group comparison to compare the percent of patients with new HO lesions as assessed by CT only at week 56 relative to the week 28 scan with the percent of patients with new HO lesions as assessed by CT only at week 28 relative to the baseline scan will be performed using McNemar's test in Placebo/REGN2477 group.

Percent of patients with new investigator-assessed flare-ups in Period 2 (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

The number and percent of patients with new investigator-assessed flare-ups in the open label period (Period 2) will be provided for the Placebo/REGN2477 group. The within group comparison to compare the percent of patients with investigator-assessed flare-ups during the double-blind period (Period 1) with the percent of patients with new investigator-assessed flare-ups in the open label period (Period 2) will be performed using McNemar's test in Placebo/REGN2477 group.

Percent of patients with new flare-ups (using patient diary) in Period 2 (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

The number and percent of patients with new flare-ups (using patient diary) in the open label period (period 2) will be provided for the Placebo/REGN2477 group. The within group comparison to compare the percent of patients with flare-ups (using patient diary) during the double-blind period with the percent of patients with new flare-ups (using patient diary) in the open label period (period 2) will be performed using McNemar's test in Placebo/REGN2477 group.

Number of new HO lesions as assessed by CT at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)

The number of new HO lesions as assessed by CT at week 56 relative to baseline scan will be descriptively summarized for the REGN2477/REGN2477 group.

Number of new HO lesions as assessed by CT only at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)

The number of new HO lesions as assessed by CT only at week 56 relative to baseline scan will be descriptively summarized for the REGN2477/REGN2477 group.

Total volume in new HO lesions as assessed by CT at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)

The total volume in new HO lesions as assessed by CT at week 56 relative to baseline scan will be descriptively summarized for the REGN2477/REGN2477 group.

Total volume in new HO lesions as assessed by CT only at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)

The total volume in new HO lesions as assessed by CT only at week 56 relative to baseline scan will be descriptively summarized for the REGN2477/REGN2477 group.

Percent of patients with new HO lesions as assessed by CT at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period) (AHO)

Percent of patients with new HO lesions as assessed by CT at week 56 relative to baseline scan will be descriptively summarized for the REGN2477/REGN2477 group.

Percent of patients with new HO lesions as assessed by CT only at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period)

Percent of patients with new HO lesions as assessed by CT only at week 56 relative to baseline scan will be descriptively summarized for the REGN2477/REGN2477 group.

Number of new lesions as assessed by ^{18}F-NaF PET at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period)

The number of new HO lesions assessed by^{18}F-NaF PET at week 56 relative to baseline scan will be descriptively summarized for the REGN2477/REGN2477 group.

In the absence of a PET/CT scan, number of new HO lesions relative to baseline as assessed by CT only will be used for the imputation of new HO lesions by ^{18}F-NaF PET.

Total lesion activity in new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period)

The total lesion activity in new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to baseline scan will be descriptively summarized for the REGN2477/REGN2477 group. In the absence of a PET/CT scan, number of new HO lesions relative to baseline as assessed by CT only will be considered as new HO lesions by ^{18}F-NaF PET and the new lesion activity in each of these new lesions will be imputed using the median new lesion activity in new lesions relative to baseline in available PET/CT scans in REGN2477/REGN2477 group.

Percent of patients with new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to baseline scan (in patients who continue REGN2477 after the double-blind period)

Percent of patients with new HO lesions as assessed by ^{18}F-NaF PET at week 56 relative to baseline scan will be descriptively summarized for the REGN2477/REGN2477 group.

In the absence of a PET/CT scan, CT only scans will be used for the imputation of the endpoint (i.e. patients with new HO lesions relative to baseline by CT only scans will be considered having new HO lesions by ^{18}F-NaF PET).

Percent change from week 28 in SUVmax to week 56 (in patients switching from placebo to REGN2477 after the double-blind period)

Percent change in SUVmax from week 28 to week 56 will be summarized for the Placebo/REGN2477 group. Wilcoxon signed rank test will be used for the within group comparison in the Placebo/REGN2477 group.

Percent change from baseline in SUVmax to week 56 (in patients who continue REGN2477 after the double-blind period)

Percent change from baseline in SUVmax to week 56 will be descriptively summarized for of the REGN2477/REGN2477 group.

Percent change from week 28 in total lesion activity by 18F-NaF PET to week 56 (in patients switching from placebo to REGN2477 after the double-blind period) versus the same patients between baseline and week 28

Percent change from week 28 to week 56 and from baseline to week 28 in total lesion activity by ^{18}F-NaF PET will be summarized for Placebo/REGN2477 group. Wilcoxon signed rank test will be used for the within group comparison.

Percent change from baseline in total lesion activity by ^{18}F-NaF PET to week 56 (in patients who continue REGN2477 after the double-blind period)

Percent change from baseline in total lesion activity by ^{18}F-NaF PET to week 56 will be descriptively summarized for of the REGN2477/REGN2477 group.

Percent change from week 28 in the total volume of HO lesions as assessed by CT to week 56 (in patients switching from placebo to REGN2477 after the double-blind period) versus the same patients between baseline and week 28 AHO

Percent change from week 28 to week 56 and from baseline to week 28 in total volume of HO lesions assessed by CT will be summarized for Placebo/REGN2477 group. Wilcoxon signed rank test will be used for the within group comparison.

Percent change from baseline in the total volume of HO lesions as assessed by CT to week 56 (in patients who continue REGN2477 after the double-blind period)

Percent change from baseline in total volume of HO lesions as assessed by CT to week 56 will be descriptively summarized for of the REGN2477/REGN2477 group.

Change from week 28 in number of HO lesions by ^{18}F-NaF PET to week 56 (in patients switching from placebo to REGN2477 after the double-blind period) versus the same patients between baseline and week 28

Change from week 28 to week 56 and from baseline to week 28 in number of HO lesions by ^{18}F-NaF PET will be summarized for Placebo/REGN2477 group. Wilcoxon signed rank test will be used for the within group comparison.

Change from week 28 in number of HO lesions by CT to week 56 (in patients switching from placebo to REGN2477 after the double-blind period) versus the same patients between baseline and week 28

Change from week 28 to week 56 and from baseline to week 28 in number of HO lesions by CT will be summarized for Placebo/REGN2477 group. Wilcoxon signed rank test will be used for the within group comparison.

Daily pain due to FOP as measured using the daily NRS (AHO)

Change from Period 1 to Period 2 in average daily pain due to FOP as measured using the daily NRS will be summarized for the Placebo/REGN2477 group. Wilcoxon signed rank test will be used for the within group comparison.

Total dosage of glucocorticoids use over time (AHO)

Change from Period 1 to Period 2 in total dosage of glucocorticoids will be summarized for the Placebo/REGN2477 group. Wilcoxon signed rank test will be used for the within group comparison.

5.7.5. Analyses of Exploratory Efficacy Variables (Period 1, Period 2 and Period 3)

All imaging and flare-up related exploratory endpoints in Period 3 will be descriptively summarized. The COVID mITT and ITT principle described in Section 5.7.4 will be extended to the endpoints in Period 3.

Change from week 28 in the total volume of new HO lesions to week 56 as assessed by CT (in patients switching from placebo to REGN2477 after the double-blind period) (AHO)

Change from week 28 to week 56 in total volume of new HO lesions for new lesions developed at week 28 as assessed by CT will be descriptively summarized.

Change from baseline in Joint function score (CAJIS) by physician at week 28, week 56, and week 76

Change from baseline in prorated axial score, upper limb score, lower limb score, total score at week 28 will be analyzed in AHO using the ANCOVA model. Analyses in AHOC will be provided only if deemed necessary. The baseline observation carried forward (BOCF) method will be used to impute missing values at week 28 as week 28 is the first post-baseline visit for assessment of CAJIS.

Summary statistics for change from baseline in prorated axial score, upper limb score, lower limb score, total score at week 56 and 76 will be provided.

Patient-reported quality of life by EQ-5D-3L at baseline, week 28, week 56, and week 76

Change from baseline in EQ-5D-3L utility scores, total score and current health state score at week 28 will be analyzed in AHO using the ANCOVA model. The baseline observation carried forward (BOCF) method will be used to impute missing values at week 28 as week 28 is the first post-baseline visit for assessment of EQ-5D-3L.

Summary statistics for change from baseline in EQ-5D-3L utility scores, total score and current health state score at week 56 and 76 will be provided.

Change from baseline in patient-reported assessment of activities of daily living by FOP I-ADL at week 28, week 56, and week 76

Change from baseline in I-ADL total score at week 28 will be analyzed in AHO using the ANCOVA model. The baseline observation carried forward (BOCF) method will be used to impute missing values at week 28 as week 28 is the first post-baseline visit for assessment of I-ADL.

Summary statistics for change from baseline in I-ADL total score at week 56 and 76 will be provided.

FOP disease activity assessed by patient diary over time

Patient reported flare-ups based on e-diary will be descriptively summarized in AHO population as follows:

- number of new flare-ups in the double-blind period. Based on number of flare-ups, patients will be categorized into 5 flare-up groups as:
    1. 0 flare-up
    2. 1 flare-up
    3. 2 flare-ups
    4. 3 flare-ups
    5. ≥ 4 flare-ups

- severity of pain, swelling, joint stiffness, decreased movement associated with flare-up. If a patient has multiple flare-ups, highest severity will be selected for each sign and symptom.

- Presence of redness, and warmth. This will be captured if a patient presents at least one incidence of redness or warmth across flare-ups

Mean duration of the new flare-ups will be descriptively provided for each treatment group.

Frequency of location with flare-up will be provided.

Total dosage of glucocorticoids use over time

Summary statistics will be provided for total dosage of glucocorticoids at week 28 and 76 for patients who use glucocorticoids in AHO population.

Change from baseline in number of HO lesions as assessed by ^{18}F-NaF PET at week 56

Change form baseline in number of HO lesions by ^{18}F-NaF PET between baseline and week 56 will be descriptively summarized.

Change from baseline in number of HO lesions detectable by CT at week 56

Change from baseline to week 56 in number of HO lesions by CT will be descriptively summarized.

Number of patients with new HO lesions as assessed by ^{18}F-NaF PET through week 28 and week 56

Number and percent of patients with new HO lesions as assessed by ^{18}F-NaF PET will be provided.

Percent change from baseline of mean SUV_{mean} of selected normal bones (Pelvic Supra-Acetabular Region) as assessed by PET/CT at week 8, week 28, and week 56

Percent change in the mean of SUV_{mean} of selected normal bones at week 8 and week 28 will be analyzed in AHO using MMRM model.

Percent change from baseline of ratio of mean SUV_{mean} of selected normal bones (Pelvic Supra-Acetabular Region) to venous plasma SUV during PET as assessed by PET/CT at week 56

Percent change in the ratio of mean of SUV_{mean} to venous plasma SUV at week 56 will be descriptively summarized in AHO.

Percent change from baseline in venous plasma clearance ^{18}F-NaF SUV at week 56

Percent change in venous plasma clearance at week 56 will be analyzed in the AHO using the ANCOVA model.

Percent change from baseline in ^{18}F-NaF incorporation rate (Ki) in individual active HO lesion(s) at week 56

Summary statistics for percent change in ^{18}F-NaF incorporation rate (Ki) will be provided on lesion level and on patient level. The percent change for patient is the mean of percent change of all the lesions for the patient at the week 56.

Percent change from baseline of ratio of SUVmax of individual active HO lesion(s) to venous plasma SUV during PET scan at week 56

Summary statistics for percent change of ratio of SUVmax to venous plasma SUV at week 56 will be provided on lesion level and on patient level. The percent change for patient is the mean of percent change of all the lesions for the patient.

Change in number of lesions that are only ^{18}F-NaF PET detectable at baseline to CT detectable lesions at week 28

The number of lesions that are only ^{18}F-NaF PET detectable that become CT detectable at week 28 will be descriptively summarized.

Change from baseline in the FEV1 of spirometry at week 28, week 56 and week 76

Change from baseline in the FEV1 at week 28 will be analyzed in AHO using ANCOVA model. The baseline observation carried forward (BOCF) method will be used to impute missing values at week 28 as week 28 is the first post-baseline visit for assessment of FEV1.

Summary statistics for change from baseline in FEV1 at week 56 and 76 will be provided.

Change from week 28 in the FEV1 of spirometry at week 56 and week 76

Summary statistics for change from week 28 in FEV1 at week 56 and 76 will be provided.

Time weighted average (standardized AUC) percent change from baseline in the levels of hs-CRP over 28 weeks

Time weighted average of percent change of hs-CRP level will be calculated for each patient. If hs-CRP for intermediate weeks are missing, linear interpolation of percent change between the two adjacent measurements will be used to calculate time weighted average. If the missing is monotone, the last observation carried forward (LOCF) method will be used to impute missing values for calculation of time weighted average over week 28.

Time-weighted average percent change in hs-CRP will be analyzed in the AHOC using the ANCOVA model.

Percent change in hs-CRP will also be analyzed using ranked analysis of covariance (ANCOVA) model. This non-parametric approach is robust with concerns of departure from the normal distribution for biomarker data. The ranked ANCOVA model will be able to provide significance of the treatment effect as compared to placebo. Median treatment difference and 95% CI can be derived from the Hodges-Lehmann estimation and Moses distribution free CI respectively.

5.8. Analysis of Safety Data

Safety and tolerability will be descriptively summarized by treatment group, including AEs, clinical laboratory variables, vital signs, 12-lead ECG.

Safety evaluations will be performed on the SAF.

Thresholds for treatment-emergent potentially clinically significant values (PCSV) in laboratory variables, vital signs, and ECG are defined in Section 10 (Appendix).

5.8.1. Analysis of Adverse Events

The number and proportion of patients will be summarized separately for:

1. Double-blind period.
2. Open label and follow-up period.
3. Entire study period.

Adverse event (AE) incidence tables will be presented for each treatment group, including the number (n) and percentage (%) of patients experiencing an AE, where multiple instances of the same event occur in the same patient the event will be counted only once for that patient. The denominator for computation of percentages is the number of patients in each treatment group.

The number and proportion of patients reporting TEAEs will be summarized, sorted by decreasing frequency of SOC and PT for the REGN2477 group.

For each TEAE, an incidence rate will be calculated from the number of patients with the TEAE divided by the patient-years which is the cumulative time at risk within a treatment group. For a TEAE, the time to first TEAE will be used in the denominator for patients who report the TEAE, the time to last day in the study will be used in the denominator for patients who do not report this TEAE. In addition, for each TEAE, an event rate will be calculated from the number of total events divided by the cumulative exposure time in the study period.

Patient listings will be provided for all SAEs, death, and TEAEs leading to permanent treatment discontinuation.

The following variables will be included in the listing:

- Patient ID

- Treatment group
- Age/sex/race
- System Organ Class (SOC)
- High Level Term (HLT)
- Preferred Term (PT)
- Verbatim Term
- AE start date and end date/ongoing (using both calendar days and study days)
- AE Duration
- Relationship of AE to study drug: unrelated or related
- Action taken: Dose withdrawn temporarily, Dose reduced, Dose withdrawn permanently, Dose not changed, Not known or Not applicable
- Severity: using a 3–point scale (mild, moderate, or severe)
- Treatment: none, medication, surgery or others
- Outcome: recovered/resolved, recovered/resolved with sequelae, recovering/resolving, not recovered/not resolved, fatal or unknown
- AESI: yes/no

Overall TEAE summary

The overall summary of TEAEs will be provided with number and proportions of patients with any:

- TEAE
- Serious TEAE
- Severe TEAE
- TEAE related to study drug
- TEAE leading to permanent discontinuation from study
- TEAE of special interest (AESI)

- TEAE leading to death
- TEAE leading to permanent treatment discontinuation

AESI

- Epididymitis
- Orchitis (inflammation of the testicles)
- Hydrocele (fluid buildup around 1 or both testicles)
- Scrotum pain
- Scrotum swelling
- Moderate to severe episodes of spontaneous non-traumatic bleeding
- Moderate epistaxis (defined as any episode lasting longer than 30 minutes or requiring professional medical intervention)
- Severe epistaxis (based on definition of a severe AE as per protocol)

TEAE Incidence

Number and proportions of patients reporting TEAEs will be summarized for the following TEAEs:

- TEAEs
  - TEAEs by SOC/PT
  - TEAEs by SOC/HLT/PT
  - TEAEs by PT
  - TEAEs by severity by SOC/PT
  - Severe TEAEs by SOC/PT
  - TEAEs related to study medication as assessed by the investigator by SOC/PT
  - Severe TEAEs related to study medication as assessed by the investigator by SOC/PT
  - TEAE of special interest/TEAE by category

- Serious TEAEs
  - Serious TEAEs by SOC/PT
  - Serious TEAEs related to study medication as assessed by the investigator by SOC/PT
- TEAEs leading to permanent discontinuation of study treatment by SOC/PT
- Death by SOC/PT
- TEAE by PT associated with infusion reaction

5.8.2. Analysis of Clinical Laboratory Measurements

Laboratory measurements include clinical chemistry, hematology and urinalysis results, and will be converted to standard international units and US conventional units. Summaries of laboratory variables will include:

- Descriptive statistics of laboratory result and change from baseline and from week 28 by visit
- The number (n) and percentage (%) of subjects with treatment-emergent PCSVs during study, depending on data
- Shift tables based on baseline normal/abnormal may be used to present the results for parameters of interest

Listing of all laboratory parameters value, normal range, abnormal flag and treatment-emergent PCSV by subject and visit will be provided.

5.8.3. Analysis of Vital Signs

Summaries of vital sign variables will include:

- Descriptive statistics of vital sign variable and change from baseline and from week 28 by visit
- The number (n) and percentage (%) of subjects with treatment-emergent PCSV, depending on data

Listings will be provided with flags indicating the treatment-emergent PCSVs, depending on data.

5.8.4. Analysis of 12-Lead ECG

Summaries of 12-lead ECG parameters by treatment group will include:

- Each ECG parameter and change from baseline and from week 28
- The number (n) and percentage (%) of subjects with PCSV, depending on data
- ECG status (i.e. normal, abnormal) summarized by a shift table

Listings will be provided with flags indicating PCSVs, depending on data.

5.8.5. Analysis of Physical Examination

The number (n) and percentage (%) of subjects with abnormal physical examination will be summarized by baseline and each scheduled measurement time with descriptive statistics

5.9. Analysis of Pharmacokinetics Drug Concentration Data

Summary of functional REGN2477 and will be presented by nominal time point (i.e., the time points specified in the protocol) using the pharmacokinetic analysis set. Plots of drug concentration will be presented over time (linear and log scales). When the scale is linear, concentrations below the lower limit of quantification (LLOQ) will be set to zero. In the log-scaled figures, concentrations below the LLOQ will be imputed as LLOQ/2.

5.10. Analysis of Total Target Concentration Data

Summary of total target concentrations will be presented by nominal time point (i.e., the time points specified in the protocol) using the total target analysis set. Plots of total target concentration will be presented over time (linear and log scales). When the scale is linear, concentrations below the lower limit of quantification (LLOQ) will be set to zero. In the log-scaled figures, concentrations below the LLOQ will be imputed as LLOQ/2.

5.11. Analysis of Anti-Drug Antibody Data

Immunogenicity will be characterized by the ADA response observed using the anti-drug antibody analysis set:

Negative - If all samples analyzed are negative in the ADA assay, or a positive ADA response at baseline with all post-dose ADA results negative, or the baseline sample is positive and all post baseline ADA titers are reported as less than 9-fold over the baseline titer values.

Treatment-boosted - defined as any post-dose positive ADA assay response that is 9-fold over baseline titer levels when baseline is positive in the ADA assay.

Treatment-emergent - defined as any post-dose positive ADA assay response when the baseline results are negative.

- Persistent - A positive result in the ADA assay detected in at least 2 consecutive post baseline samples separated by at least a 16-week post baseline period, with no ADA-negative results in-between, regardless of any missing samples

- Indeterminate - A positive result in the ADA assay at the last collection time point analyzed only, regardless of any missing samples
- Transient - Not persistent or indeterminate regardless of any missing samples
- Maximum ADA Titer values
  - Low (titer <1,000)
  - Moderate (1,000≤ titer ≤10,000)
  - High (titer >10,000)

Listings of ADA positivity and ADA titer presented by patient/subject, time point, and dose cohort/group will be provided. Incidence of treatment-emergent ADA will be assessed as absolute occurrence (N) and percent of patients (%), grouped by study cohorts and ADA titer level.

Plots of drug concentrations will be examined for the influence of ADAs on individual PK profiles evaluated. Assessment of impact of ADA on safety and efficacy may be provided.

6. DATA CONVENTIONS

The following analysis conventions will be used in the statistical analysis.

6.1. Definition of Baseline for Efficacy/Safety Variables

Unless otherwise specified, the baseline measurement for all measurements will be the latest available valid measurement taken prior to the administration of study drug. If any randomized patients are not treated, the baseline will be the last value on or prior to the randomization. The following rules specify the determination by both date/time information:

1. For the AE, lab (including biomarker), drug concentration and ADA data, both date and time of the measurement will be used to determine baseline by comparing with the first infusion date and time.

2. For other data except AE, lab (including biomarker), drug concentration or ADA, only date of the measurement will be used to determine baseline by comparing with the first infusion date.

For the rescreened patients, all data from the same patient will be used to derive baseline regardless if the data is from the screen- failure subject ID or enrolled subject ID.

6.2. General Data Handling Conventions

For the laboratory safety variables and biomarker data, if the data below the lower limit of quantification (LLOQ) / limit of linearity, half of the lower limit value (i.e., LLOQ/2) will be used for quantitative analyses. For data above the upper limit of quantification (ULOQ) / limit of linearity, the upper limit value (i.e., ULOQ) will be used for quantitative analyses.

6.3. Data Handling Convention for Missing Data

Missing data will not be imputed in listings. This section includes the methods for missing data imputation for some summary analyses, if necessary.

Missing/incomplete dates

Every effort will be made to collect the start dates of all AEs and concomitant medications. However, in the case the start date of an AE or concomitant medication is incomplete or missing, it will be assumed to have occurred on or after the first dose of study medication, except if an incomplete date (e.g., month and year) clearly indicates that the event started prior to treatment. If the partial date indicates the same month or year of the first dose of study medication date, then the start date of the first dose will be imputed, otherwise, the missing day or month by the first day or the first month will be imputed.

Adverse event

If the intensity of a TEAE is missing, it will be classified as "severe" in the frequency tables by intensity of TEAE. If the measurement of relationship of a TEAE to the investigational product is missing, it will be classified as "related" in the frequency tables by relation to the investigational product.

Medication coding

Medications whose ATC level 4 cannot be coded will be summarized by setting ATC4=ATC2 in the table programs. However, these un-coded ATC level 4 records still need to be confirmed with study DM and study MD.

PCSV

Patients who had post-baseline PCSV but missing baseline value will be regarded as having treatment emergent PCSV.

6.4. Visit Windows

Data analyzed by-visit-analysis (including efficacy, laboratory data, visit sign, ECG) will be summarized by the study scheduled visits described Appendix 10.2, "Schedule of Event". The analysis visit windows will be exhaustive so that all available values obtained from unscheduled visits, early termination visit (ETV) and end of treatment (EOT)/end of study (EOS) have the potential to be summarized. No analysis visit windows will be applied for the study scheduled visits. The visit windows are constructed using ranges applied to the number of days in study (study days) when the measure is collected. Day 1 is defined as the first date of study treatment administration and is labeled as baseline.

The analysis visit windows will generally be not applied when calculating time weighted average of % change/change of total lesion activity, daily pain score, biomarker of bone formation and hs-CRP. However, in the absence of data at scheduled visits, unscheduled visits with the visit windows can be applied in calculating these outcome variables.

Time Window for efficacy variables

Visit No.	Visit	Targeted Study Days	Measurement from PET and CT	Biomarker of bone formation, hs-CRP	FOP I-ADL, EQ-5D-3L, CAJIS, FEV1
1	Screening	<1			<1
2	Baseline	1	1	1	1
4	Week 1	7		[2,18]	
5	Week 4	29		[19,43]	
6	Week 8	57	[2,127]	[44,85]	
7	Week 12	85			
8	Week 16	113		[86, 155]	
9	Week 20	141			
10	Week 24	169			
11	Week 28 (End of double-blind period)	197	[128,218]	[156, 211]	[2,295]
12	Week 32	225		[212, 239]	
13	Week 36	253		[240, 281]	
14	Week 40	281			
15	Week 44	309		[282, 337]	
16	Week 48	337			
17	Week 52	365		[338, 386]	
18	Week 56	393	[219,463]		[296,463]
19	Week 60	421			
20	Week 64	449			
21	Week 68	477			
22	Week 72	505			
23	Week 76	533	≥464		>=≥464
24	Week 80	560			

Time Window for safety variables

Visit No.	visit	Targeted Study Days	Lab	Vital Sign	ECG
1	Screening	<1			<1
2	Baseline	1	1	1	1
4	Week 1	7		[2,18]	
5	Week 4	29	[2,43]	[19,43]	[2,43]
6	Week 8	57	[44,71]	[44,71]	[44,113]
7	Week 12	85	[72, 99]	[72, 99]	
8	Week 16	113	[100, 127]	[100, 127]	
9	Week 20	141	[128, 155]	[128, 155]	
10	Week 24	169	[156, 183]	[156, 183]	[114,183]
11	Week 28 (End of double-blind period)	197	[184, 211]	[184, 211]	[184,239]
12	Week 32	225	[212, 239]	[212, 239]	
13	Week 36	253	[240, 267]	[240, 267]	
14	Week 40	281	[268, 295]	[268, 295]	[240,337]
15	Week 44	309	[296, 323]	[296, 323]	
16	Week 48	337	[324, 351]	[324, 351]	
17	Week 52	365	[352, 379]	[352, 379]	
18	Week 56	393	[380, 407]	[380, 407]	[338, 407]
19	Week 60	421	[408, 435]	[408, 435]	[408, 435]
20	Week 64	449	[436, 463]	[436, 463]	[436, 463]
21	Week 68	477	[464, 491]	[464, 491]	[464, 491]
22	Week 72	505	[492, 519]	[492, 519]	[492, 519]
23	Week 76	533	[520, 547]	[520, 547]	[520, 547]
24	Week 80	560	≥548	≥548	≥548

In general, the following order will be used to select the record for analysis at given visit:

1. Scheduled visit
2. Early termination (ET) or end of study (EOS), whichever comes first if scheduled visit not available

3. Unscheduled visit if both scheduled visit and ETV/EOT/EOS are not available

For the multiple measurements of the same test in the same window, the following rules will be used to pick up the analysis value:

- If multiple valid values of a variable within an analysis visit window, the closest from the target study day will be selected.
- If the difference is a tie, the value after the targeted study day will be used.
- If multiple available values of a variable exist within a same day, then the first value of the day will be selected.

7. INTERIM ANALYSIS

No formal interim analysis is planned.

8. SOFTWARE

All analyses will be done using SAS Version 9.4.

9. REFERENCES

ICH. (1998, February 5). ICH Harmonized tripartite guideline: Statistical principles for clinical trials (E9). International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use.

E Botman, PGHM Raijmakers, M Yaqub, B Teunissen, C Netelenbos, W Lubbers, et al. (2019) Evolution of heterotopic bone in fibrodysplasia ossificans progressiva: An [^{18}F]NaF PET/CT study. Bone 124, 1-6. Installe J, Nzeusseu A, Bol A, Depresseux G, Devogelaer JP, and Lonneux M. (2005) (18)F-fluoride PET for monitoring therapeutic response in Paget's disease of bone. Journal of nuclear medicine : official publication, Society of Nuclear Medicine 46(10):1650-8.

10. APPENDIX

10.1. Summary of Statistical Analyses (Period 1)

Endpoint	Analysis Populations	Statistical Method	Supportive Analysis	Subgroup Analysis	Other Analyses
Efficacy analysis	AHOC, AHO, FAS	ANCOVA, MMRM,Ranked ANCOVA, Negative Binomial regression, Fisher exact test	Yes	Yes	Yes
Adverse Events	SAF	Descriptive Statistics	No	No	No
Laboratory Measures	SAF	Descriptive Statistics	No	No	No
Vital sign	SAF	Descriptive Statistics	No	No	No
ECG	SAF	Descriptive Statistics	No	No	No

10.2. Schedule of Events

Table 2: Schedule of Events: Screening/Baseline through Treatment Period 1 (Randomized Double-Blind Treatment Period)

		Period 1 (Randomized Double-Blind Treatment)								
Visit[1s]	**Screening/ Baseline Period**	**Visit 2**	**Visit 3[2]**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**	**Visit 8**	**Visit 9**	**Visit 10**
Day ± visit window (d)	**-28 to -1**	**1**	**2**	**7 ±1d**	**29 ±3d**	**57 ±7d**	**85 ±7d**	**113 ±7d**	**141 ±7d**	**169 ±7d**
Week		**1**	**1**	**1**	**4**	**8**	**12**	**16**	**20**	**24**
Screening/Baseline[3, 20]:										
Informed consent	X									
Informed consent (optional genomic sub-study)	X									
Inclusion/Exclusion	X									
Medical/Prior Medication/Surgical history[4]	X									
Urogenital exam (male patients only)	X									
Scrotum ultrasound with Doppler (male patients only)	X									
Blood sample for HbA1c measurement	X									
Confirm clinical FOP diagnosis, including documentation of ACVR1 mutation[5]	X									
Whole blood DNA sample for ACVR1 gene sequencing[21]	X									
Demographics	X									
Treatment:										
Study enrollment randomization		X								
Administer Study Drug (REGN2477 or placebo)[6, 8, 9]		X			X	X	X	X	X	X
Concomitant medications	X	X	X	X	X	X	X	X	X	X
Safety:										
Adverse events	X	X	X	X	X	X	X	X	X	X
Height (if feasible)[22]	X									
Body Weight	X	X			X	X	X	X	X	X
Vital signs[7, 8]	X	X[8]	X	X	X	X	X	X	X	X
Electrocardiogram[7, 9]	X	X[9]	X		X	X				X
Physical examination	X									
Menstrual history, pregnancy status reporting and confirmation of contraception use[10]	X	X			X	X	X	X	X	X

Visit[1s]	**Screening/ Baseline Period**	**Period 1 (Randomized Double-Blind Treatment)**								
		Visit 2	**Visit 3[2]**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**	**Visit 8**	**Visit 9**	**Visit 10**
Day **± visit window (d)**	**-28 to -1**	**1**	**2**	**7** **±1d**	**29** **±3d**	**57** **±7d**	**85** **±7d**	**113** **±7d**	**141** **±7d**	**169** **±7d**
Week		**1**	**1**	**1**	**4**	**8**	**12**	**16**	**20**	**24**
Laboratory Testing[3]:										
Hematology[7, 11]	X	X	X		X	X	X	X	X	X
Blood chemistry[7, 11]	X	X	X		X	X	X	X	X	X
Urinalysis[7, 23]	X	X	X		X	X	X	X	X	X
Blood samples for FSH[7, 11]	X	X			X	X				
Blood samples for ACTH, GH, cortisol, TSH, free T4, testosterone (total and free; male patients only), estradiol, LH, amylase, and lipase[7, 11]		X			X	X				
Blood sample for coagulation parameters (activated partial thromboplastin time, prothrombin time, INR, thrombin time), fibrinogen, and platelet effector function	X[25]				X		X			
Pregnancy test (WoCBP)[7, 10, 11]	serum	urine			urine	urine	urine	urine	urine	urine
Imaging Assessments:										
^{18}F-NaF PET[7, 12, 13]	X					X				
Blood sample for ^{18}F-NaF tracer concentration (optional)[7, 11, 12, 13]	X									
Low dose CT[7, 12]	X					X				
Clinical Endpoint Measures[24]										
Daily Pain NRS[14, 18]	X	X	X	X	X	X	X	X	X	X
Daily FOP disease activity e-diary[14, 18]	X	X	X	X	X	X	X	X	X	X
FOP I-ADL	X									
EQ-5D-3L	X									
Joint function (CAJIS)	X									
Pulmonary function by spirometry (FVC, FEV1)	X									
Biochemical Bone Formation and Mechanism based Biomarkers										
Blood samples for BSAP and P1NP[7, 11]	X	X		X	X	X		X		
Blood samples for mechanism of action or safety	X									
PK/Drug Concentration and ADA Samples:										
PK/Drug conc. and activin A sample[7, 11]		X		X	X	X	X	X	X	
ADA sample[7, 11]		X						X		
Biomarker Procedures:										
Whole blood for RNA gene expression[7, 11]		X	X	X	X					

		Period 1 (Randomized Double-Blind Treatment)								
Visit[1s]	**Screening/ Baseline Period**	**Visit 2**	**Visit 3[2]**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**	**Visit 8**	**Visit 9**	**Visit 10**
Day ± visit window (d)	**-28 to -1**	**1**	**2**	**7 ±1d**	**29 ±3d**	**57 ±7d**	**85 ±7d**	**113 ±7d**	**141 ±7d**	**169 ±7d**
Week		**1**	**1**	**1**	**4**	**8**	**12**	**16**	**20**	**24**
Future Biomedical Research Samples										
Research biomarker serum and biomarker plasma[7, 11]	X	X		X	X	X	X	X		
Optional Genomic Study: DNA Analysis										
Whole blood DNA sample (optional)[7, 15]		X								

Table 3: Schedule of Events: Period 2 (Open-Label Treatment Period)

	Period 2 (Open-Label REGN2477 Treatment)						
Visit	**Visit 11[16]**	**Visit 12**	**Visit 13**	**Visit 14**	**Visit 15**	**Visit 16**	**Visit 17**
Day ± visit window (d)	**197 -7d/+15d**	**225 ±7d**	**253 ±7d**	**281 ±7d**	**309 ±7d**	**337 ±7d**	**365 ±7d**
Week	**28**	**32**	**36**	**40**	**44**	**48**	**52**
Treatment:							
Administer REGN2477[6, 8, 9]	X	X	X	X	X	X	X
Concomitant medications	X	X	X	X	X	X	X
Safety:							
Adverse events	X	X	X	X	X	X	X
Body Weight	X	X	X	X	X	X	X
Vital signs[7, 8]	X	X	X	X	X	X	X
Electrocardiogram[7, 9]	X[9]			X			
Physical examination	X						
Menstrual history, pregnancy status reporting and confirmation of contraception use[10]	X	X	X	X	X	X	X
Laboratory Testing:							
Hematology[7, 11]	X	X	X	X	X	X	X
Blood chemistry[7, 11]	X	X	X	X	X	X	X
Urinalysis[7, 23]	X	X	X	X	X	X	X
Blood samples for ACTH, GH, cortisol TSH, free T4, testosterone (total and free; male patients only), FSH, estradiol, LH, amylase, and lipase[7, 11]	X	X		X			
Blood sample for coagulation parameters (activated partial thromboplastin time, prothrombin time, INR, thrombin time), fibrinogen, and platelet effector function	X	X		X			
Pregnancy test (WoCBP)[7, 10]	urine	urine	urine	urine	urine	urine	urine
Imaging and Other Assessments:							
^{18}F-NaF PET[7, 13]	X						
Low dose CT[7, 13]	X						
Clinical Endpoint Measures[24]							
Daily Pain NRS[14]	X	X	X	X	X	X	X
Daily FOP disease activity e-diary[14]	X	X	X	X	X	X	X
FOP I-ADL	X						
EQ-5D-3L	X						
Joint function (CAJIS)	X						

	Period 2 (Open-Label REGN2477 Treatment)						
Visit	**Visit 11[16]**	**Visit 12**	**Visit 13**	**Visit 14**	**Visit 15**	**Visit 16**	**Visit 17**
Day **± visit window (d)**	**197** **-7d/+15d**	**225** **±7d**	**253** **±7d**	**281** **±7d**	**309** **±7d**	**337** **±7d**	**365** **±7d**
Week	**28**	**32**	**36**	**40**	**44**	**48**	**52**
Pulmonary function by spirometry (FVC, FEV1)	X						
Biochemical Bone Formation and Mechanism based Biomarkers							
Blood samples for BSAP and P1NP[7, 11]	X	X	X		X		X
Blood samples for mechanism of action or safety	X						
PK/Drug Concentration and ADA Samples:							
PK/Drug conc. and activin A sample[7, 11]	X	X	X	X	X	X	X
ADA sample[7, 11]	X			X			X
Future Biomedical Research Samples							
Research biomarker serum and biomarker plasma[7, 11]	X	X			X		X

Table 4: Schedule of Events: Period 3 (Follow-Up Treatment Period)

	Period 3 (Follow-Up Treatment Period)					End of Study[17]	
Visit	**Visit 18[19]/ Early Termination Visit**	**Visit 19**	**Visit 20**	**Visit 21**	**Visit 22**	**Visit 23**	**≥Visit 24**
Day	**393**	**421**	**449**	**477**	**505**	**533**	561
± visit window (d)	**±14d**	**±7d**	**±7d**	**±7d**	**±7d**	**±7d**	**±7d**
Week	**56**	**60**	**64**	**68**	**72**	**76**	**80**[17]
Treatment:							
Administer REGN2477[6, 8, 9]	X	X	X	X	X	X	X
Concomitant medications	X	X	X	X	X	X	X
Safety:							
Adverse events	X	X	X	X	X	X	X
Vital signs[7, 8]	X	X	X	X	X	X	X
Electrocardiogram[7, 9]	X	X	X	X	X	X	X
Physical examination[9]	X	X	X	X	X	X	X
Menstrual history, pregnancy status reporting and confirmation of contraception use[10]	X	X	X	X	X	X	X
Laboratory Testing:							
Hematology[7, 11,]	X	X	X	X	X	X	X
Blood chemistry[7, 11]	X	X	X	X	X	X	X
Urinalysis[7, 23]	X	X	X	X	X	X	X
Blood samples for ACTH, GH, cortisol TSH, free T4, testosterone (total and free; male patients only), FSH, estradiol, LH, amylase, and lipase[7, 11]	X					X	
Pregnancy test (WoCBP)[7, 10, 11]	Urine	urine	urine	urine	urine	serum	urine/serum
Imaging and Other Assessments:							
^{18}F-NaF PET[7, 13]	X					X	
Blood sample for ^{18}F-NaF tracer concentration (optional)[7, 11, 13]	X						
Low dose CT[7, 13]	X					X	

	Period 3 (Follow-Up Treatment Period)					End of Study[17]	
Visit	**Visit 18[19]/ Early Termination Visit**	**Visit 19**	**Visit 20**	**Visit 21**	**Visit 22**	**Visit 23**	**≥Visit 24**
Day	**393**	**421**	**449**	**477**	**505**	**533**	561
± visit window (d)	**±14d**	**±7d**	**±7d**	**±7d**	**±7d**	**±7d**	**±7d**
Week	**56**	**60**	**64**	**68**	**72**	**76**	**80**[17]
Clinical Endpoint Measures[24]							
Daily Pain NRS[14]	X	X	X	X	X	X	X
Daily FOP disease activity e-diary[14]	X	X	X	X	X	X	X
FOP I-ADL	X					X	
EQ-5D-3L	X					X	
Joint function (CAJIS)	X					X	
Pulmonary function by spirometry (FVC, FEV1)	X					X	
Biochemical Mechanism based Biomarkers							
Blood samples for mechanism of action or safety	X						
Future Biomedical Research Samples							
Research biomarker serum and biomarker plasma[7, 11]	X					X	
PK/Drug Concentration and ADA Samples:							
PK/Drug conc. and activin A sample[7, 11]	X	X	X	X	X	X	X
ADA sample[7, 11]		X	X	X	X	X	X

10.2.1. Footnotes for the Schedule of Events Table 2, Table 3, and Table 4

1. Procedures other than those associated with drug administration and imaging procedures may be conducted by trained study staff at or near patients' residences (i.e., at home visits).

2. Study visit 3 (day 2) will be conducted in the clinic following the first dose of study drug, although no study drug administration or imaging procedure is conducted at this visit.

3. Procedures may be conducted on different days during the screening/baseline period, if needed. Blood draws for laboratory testing at the screening visit will be collected in a fasted state (after an 8-hour fast). No other blood draw will require fasting.

4. Medical history should include detailed FOP history, and prior medication and surgery for treatment of FOP.

5. Documentation of ACVR1 genotype, including classic [R206H] or different ACVR1 mutations should be obtained for study file. If at the time of screening, the patient doesn't have documentation of what ACVR1 mutation he/she carries or it cannot be retrieved, then the study site staff should collect a blood sample for the testing by the investigator to determine the ACVR1 mutation prior to randomization. This testing, which will be performed on an as needed basis, is separate from the mandatory whole blood collection described in footnote #21.

6. All patients who receive study drug will be closely monitored after each study drug administration for at least 2 hours.

7. Study procedures within each visit may be conducted on different days, within the stated visit window. If conducted on the same day, vital signs and ECG will be conducted before blood sampling, and blood samples (except ^{18}F-NaF tracer concentrations) will be collected before imaging procedures. Study drug administration will be performed last at that study visit.

8. On study day 1 (Period 1) and on study day 197 (Period 2), vital signs will be collected pre-dose and post-dose at 1 hour and 4 hours after the end of infusion. On subsequent visits, vital signs will be collected prior to dosing.

9. On study day 1 (Period 1) and on study day 197 (Period 2), ECG will be conducted pre-dose and post-dose 1 hour and 4 hours after the end of infusion. On subsequent visits, ECG will be conducted prior to dosing. ECG and symptom-directed physical examination may be performed during Period 3 at the investigator's discretion.

10. Menstrual history and pregnancy status of WoCBP will be determined at the screening visit. Menstrual events and pregnancy status of WoCBP will be monitored at each visit, except visit 3 and visit 4. For all patients, contraception use will be confirmed. A negative result for the pregnancy test must be obtained prior to each study drug administration. After week 56, urine pregnancy testing will be performed prior to each REGN2477 dose administration. A serum pregnancy test will be performed at the patient's end of study visit.

11. In some patients, it may not be possible to obtain venous blood samples for all tests. Serum pregnancy tests and other blood tests for determining eligibility must be obtained. Blood samples for safety labs will be prioritized. After week 56, PK/ADA sampling will continue to be performed at all visits.

12. The baseline imaging assessments by ^{18}F-NaF-PET and low dose CT must be conducted between day-7 to day -1.

13. The ^{18}F-NaF-PET may be performed at a separate location than the study clinic, within the visit window, before study drug administration. Blood samples for the determination of the tracer concentration in plasma will be collected for baseline and week 56 imaging sessions from a subset of patients who will undergo an optional dynamic PET scan. Any incidental findings identified through PET/CT scans will be reported to the investigator and medical monitor. The investigator will communicate any findings to the patient.

14. Daily pain NRS score and FOP disease activity will be collected as entries in a diary completed by. If the patient is physically unable to or requires assistance to complete the e-diary on their own caregivers can complete or provide assistance in completion of the e-diary. Caregivers can only physically assist with completion of the e-diary and there is to be no interpretation of patient disease activity on their part.

15. Separate consent is required for participation in the optional genomic DNA sub-study and collection of blood sample (DNA). The blood sample for genomic DNA should be collected on day 1 or may be collected at any visit.

16. Patients will undergo end of double-blind treatment (Period 1) assessments at the week 28 visit (Period 1), which is the first visit in the open-label treatment (Period 2) with administration of open-label REGN2477.

17. The week 80 column illustrates additional visits and procedures to be performed every 4 weeks, provided that the patient has reached the week 76 visit, and data for the last patient randomized into the study through the week 28 visit (ie, end of Period 1) have not yet been collected and validated, and results of the primary analyses of safety and efficacy are not yet available to the sponsor.

18. The baseline entry for both the daily pain NRS and the daily FOP disease activity diary will be on day -1.

19. Patients will undergo end of open-label treatment (Period 2) assessments at the week 56 visit, which is the first visit of the follow-up treatment period (Period 3).

20. Patients can be rescreened for study participation.

21. All patients will have a whole blood sample collected, preferably at the screening/baseline visit, for future ACVR1 gene sequencing. However, the sample can be collected at any time during the study period, as the result of this test is not to be used for assessment of study eligibility.

22. Depending upon the level of musculoskeletal compromise for a patient, the measurement of the patient's height may not be precise or sometimes not possible to be assessed.

23. A urinalysis and culture will be collected/performed if there is an AESI (only applicable to male patients).

24. With the exception of the Daily Pain NRS and FOP disease activity e-diary, if a study questionnaire is not available in a patient's native language at the time of enrollment that questionnaire will not be part of the required study procedures for that patient.

25. For patients already enrolled in the study, the blood sample for these assessments will be collected at their next visit, unless these assessments have been performed in the past year. These assessments will be conducted locally.

10.3. Criteria for Treatment-Emergent Potentially Clinically Significant Value

Parameter	Treatment Emergent PCSV	Comments
Clinical Chemistry		
ALT*	>3 and ≤ 5 ULN and baseline ≤ 3 ULN*	Enzyme activity must be expressed in ULN, not in IU/L.
	>5 and ≤ 10 ULN and baseline ≤ 5 ULN	Concept paper on DILI – FDA draft Guidance Oct 2007.
	>10 and ≤ 20 ULN and baseline ≤ 10 ULN	Each category is calculated independently.
	>20 ULN and baseline ≤ 20 ULN	* At least one level is required, multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution across the different PCSV levels, additional shift table on ≤3, >3 to ≤5, > 5 to ≤10, >10 to ≤20, and > 20 category for baseline vs. post baseline may be provided
AST*	>3 and ≤ 5 ULN and baseline ≤ 3 ULN*	Enzyme activity must be expressed in ULN, not in IU/L.
	>5 and ≤ 10 ULN and baseline ≤ 5 ULN	Concept paper on DILI – FDA draft Guidance Oct 2007.
	>10 and ≤ 20 ULN and baseline ≤ 10 ULN	Each category is calculated independently.
	>20 ULN and baseline ≤ 20 ULN	* At least one level is required, multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution across the different PCSV levels, additional shift table on ≤3, >3 to ≤5, > 5 to ≤10, >10 to ≤20, and > 20 category for baseline vs. post baseline may be provided
Alkaline Phosphatase	>1.5 ULN and baseline ≤ 1.5 ULN	Enzyme activity must be expressed in ULN, not in IU/L.
		Concept paper on DILI – FDA draft Guidance Oct 2007.
Total Bilirubin*	>1.5 and ≤ 2 ULN and baseline ≤ 1.5 ULN*	Must be expressed in ULN, not in µmol/L or mg/L. Categories are cumulative.
	>2 ULN and baseline ≤ 2.0 ULN	Concept paper on DILI – FDA draft Guidance Oct 2007.
		* At least one level is required, multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution of significant level, additional shift table on ≤1.5, >1.5 to ≤2.0 and > 2.0 category for baseline vs. post baseline may be provided

Parameter	Treatment Emergent PCSV	Comments
Conjugated Bilirubin	(Direct Bilirubin >35% Total Bilirubin and Total Bilirubin >1.5 ULN) and (Direct Bilirubin <=35% Total Bilirubin or Total Bilirubin <=1.5 ULN) at baseline	Conjugated bilirubin dosed on a case-by-case basis.
ALT and Total Bilirubin	(ALT>3 ULN and TBILI>2 ULN) and baseline (ALT <=3 ULN **or** TBILI <=2 ULN)	Concept paper on DILI – FDA draft Guidance Oct 2007.
CPK*	>3 and ≤ 10 ULN and baseline ≤ 3ULN*	FDA Feb 2005.
	>10 ULN and baseline ≤ 10ULN	Am J Cardiol April 2006.
		Categories are cumulative.
		* At least one level is required, multiple levels are optional for phase 2/3 studies. If it is desirable to get the distribution of significant level, additional shift table on ≤3, >3 to ≤10, and > 10 category for baseline vs. post baseline may be provided
Creatinine	≥150 µmol/L (Adults) and baseline < 150 µmol/L	Benichou C., 1994.
	>=30% change from baseline and <100% change from baseline	3 independent criteria
	≥100% change from baseline	
Uric Acid		Harrison- Principles of internal Medicine 17th Ed., 2008.
Hyperuricemia	>408 µmol/L and <=408 µmol/L at baseline	Two independent criteria
Hypouricemia	<120 µmol/L and >= 120 µmol/L at baseline	
Blood Urea Nitrogen	≥17 mmol/L and <17 mmol/L at baseline	Two independent criteria
Chloride		Two independent criteria
Hypochloremia	<80 mmol/L and baseline ≥ 80 mmol/L	
Hyperchloremia	>115 mmol/L and baseline ≤ 115 mmol/L	

Parameter	Treatment Emergent PCSV	Comments
Sodium		Two independent criteria
Hyponatremia	≤129 mmol/L and baseline > 129 mmol/L	
Hypernatremia	≥160 mmol/L and baseline <160 mmol/L	
Potassium		FDA Feb 2005.
Hypokalemia	<3 mmol/L and baseline ≥ 3 mmol/L	Two independent criteria
Hyperkalemia	≥5.5 mmol/L and baseline <5.5 mmol/L	
Total Cholesterol	≥7.74 mmol/L and < 7.74 mmol/L at baseline	Threshold for therapeutic intervention.
Triglycerides	≥4.6 mmol/L and < 4.6 mmol/L at baseline	Threshold for therapeutic intervention.
Glucose		
Hypoglycaemia	(≤3.9 mmol/L and <LLN) and (>3.9 mmol/L or >=LLN) at baseline	ADA May 2005.
Hyperglycaemia	≥11.1 mmol/L (unfasted); ≥7 mmol/L (fasted) and < 11.1 mmol/L (unfasted); <7 mmol/L (fasted) at baseline	ADA Jan 2008.
HbA1c	>8% and <= 8% at baseline	
Albumin	≤25 g/L and >25 g/L at baseline	
CRP	>2 ULN or >10 mg/L (if ULN not provided) and <=2 ULN or <=10 mg/L (if ULN not provided) at baseline	FDA Sept 2005.

Parameter	Treatment Emergent PCSV	Comments
Hematology		
WBC	<3.0 Giga/L and >=3.0 Giga/L at baseline (Non-Black); <2.0 Giga/L and >=2.0 Giga/L at baseline (Black)	Increase in WBC: not relevant.
	≥16.0 Giga/L and < 16 Giga/L at baseline	To be interpreted only if no differential count available.
Lymphocytes	>4.0 Giga/L and <= 4.0 Giga/L at baseline	
Neutrophils	<1.5 Giga/L and >=1.5 Giga/L at baseline (Non-Black); <1.0 Giga/L and >=1.0 Giga/L at baseline (Black)	International Consensus meeting on drug-induced blood cytopenias, 1991. FDA criteria.
Monocytes	>0.7 Giga/L <= 0.7 Giga/L at baseline	
Basophils	>0.1 Giga/L <= 0.1 Giga/L at baseline	
Eosinophils	(>0.5 Giga/L and >ULN) and (<=0.5 Giga/L or <= ULN at baseline)	Harrison- Principles of internal Medicine 17th Ed., 2008.
Hemoglobin	≤115 g/L and > 115 g/L at baseline for male; ≤95 g/L and> 95 g/L at baseline for Female.	Three criteria are independent.
	≥185 g/L and <185 g/L at baseline for Male; ≥165 g/L and < 165 g/L at baseline for Female	Criteria based upon decrease from baseline are more relevant than based on absolute value. Other categories for decrease from baseline can be used (≥30 g/L, ≥40 g/L, ≥50 g/L).
	Decrease from Baseline ≥20 g/L	

Parameter	Treatment Emergent PCSV	Comments
Hematocrit	≤0.37 v/v and > 0.37 v/v at baseline for Male ; ≤0.32 v/v and > 0.32 v/v at baseline forFemale ≥0.55 v/v and < 0.55 v/v at baseline for Male ; ≥0.5 v/v and < 0.5 v/v at baseline for Female	Two Criteria are independent
RBC	Female <3 Tera/L and baseline ≥3 Tera/L ≥6 Tera/L and baseline < 6 Tera/L Male <4 Tera/L and baseline ≥4 Tera/L ≥7 Tera/L and baseline < 7 Tera/L	Unless specifically required for particular drug development, the analysis is redundant with that of Hb. Otherwise, consider FDA criteria.
Platelets	<100 Giga/L and >=100 Giga/L at baseline ≥700 Giga/L and < 700 Giga/L at baseline	International Consensus meeting on drug-induced blood cytopenias, 1991. Two independent criteria
Urinalysis		
pH	≤4.6 and > 4.6 at baseline ≥8 and < 8 at baseline	Two independent criteria
Vital signs		
HR	≤50 bpm and decrease from baseline ≥20 bpm ≥120 bpm and increase from baseline≥20 bpm	To be applied for all positions (including missing) except STANDING.
SBP	≤95 mmHg and decrease from baseline ≥20mmHg ≥160 mmHg and increase from baseline ≥20 mmHg	To be applied for all positions (including missing) except STANDING.

Parameter	Treatment Emergent PCSV	Comments
DBP	≤45 mmHg and decrease from baseline ≥10 mmHg	To be applied for all positions (including missing) except STANDING.
	≥110 mmHg and increase from baseline ≥10 mmHg	
Weight	≥5% increase from baseline	FDA Feb 2007.
	≥5% decrease from baseline	
ECG		Ref.: CPMP 1997 guideline.
HR	≤50 bpm and decrease from baseline ≥20 bpm	
	≥120 bpm and increase from baseline ≥20 bpm	
PR	≥220 ms and increase from baseline ≥20 ms	
QRS	≥120 ms & < 120 ms at baseline	

Parameter	Treatment Emergent PCSV	Comments
QTc	Absolute values (ms)	To be applied to any kind of QT correction formula.
Borderline	Borderline:	
Prolonged*	431-450 ms and < 431ms at baseline for Male;	
Additional	451-470 ms and < 451 ms at baseline for Female	
	Prolonged:	
	>450 to <500 ms and <= 450 ms at baseline for Male;	*QTc prolonged and ΔQTc>60 ms are the PCSA to be identified in individual subjects/patients listings.
	>470 to <500 ms and <= 470 ms at baseline for Female	
	≥500 ms and < 500 ms at baseline	5 independent criteria
	Increase from baseline	
	Borderline: Increase from baseline 30-60 ms	
	Prolonged: Increase from baseline >60 ms	

10.4. NRS Questionnaire

These are the three questions for the Pain NRS:

1. Please rate your current level of pain due to FOP.
2. Please rate the least level of pain you experienced due to FOP over the past 24 hours.
3. Please rate the worst level of pain you experienced due to FOP over the past 24 hours.

For each of these questions, the response is a rating on a scale from 0 to 10, where 0 is "no pain" and 10 is "worst possible pain".

The average of the 3 ratings will be used to represent the patient's level of pain over the past 24 hours.

10.5. FOP Disease Activity Assessment (DAA) Questionnaire

Patient Instructions:

Think about how active your FOP disease has been in the last 12 months. You will be asked about your FOP disease during a flare-up and without a flare-up.

In this study, an FOP flare-up is defined as experiencing at least 2 of the following: new onset of pain, swelling, joint stiffness, decrease in movement, or detection of HO.

In this first entry of the FOP Disease Activity diary, you will be asked about flare-ups you experienced in the last 12 months.

You will record the area(s) of the flare-ups in a body map and note the start and end date for each flare-up event. If you experienced more than one flare-up at the same area at different times you will record them as different flare-up events. If you experienced a flare-up in multiple adjacent areas, you may select more than one area and record it as a single flare-up event.

If you are currently experiencing an ongoing flare-up, you will be able to select "ongoing" for the end date. You will be asked to record the symptoms you are experiencing and to rate the intensity of each of four key symptoms: pain, swelling, stiffness, and decreased movement.

When rating the intensity of your symptoms, please consider how much the symptom affects daily activities that you usually could do independently:

0: no symptom (the diary will automatically record this as 0)

1: mild symptom: does not affect my usual activity

2: moderate symptom(s): limits my usual activity

3: severe symptom(s): disabling; I am not able to do what I normally could do independently

Explanation of Body Map with pictures* (CRFH will include this)

Date: *scroll down choices YYYY MM DD (unknown allowed for day)*

Subject Number: ___ ___ __ *system auto populate*

This entry is completed by (baseline and daily)

- ☐ Self
- ☐ With the help of a caregiver

↓

In the last 12 months, have you experienced at least one FOP flare-up?

- ☐ Yes
- ☐ No (go to the last question)

↓

If yes, how many flare-up events did you experience?
(Select a number from a scroll down list of numbers)

↓

Now please tell us about each of your flare-ups in the last 12 months
Body area(s), dates, outcome
Symptoms, severity (for ongoing flare-ups)

Daily: for patient who had at least one ongoing flare-up

Now please tell us about your previously ongoing flare-up
(Outcome: symptom, severity if still ongoing)

Daily: In the last 24 hours, have you experienced any new flare-ups?

- ☐ Yes, *please tell us about your new flare-up, populate body area(s), symptom, severity*
- ☐ No (go to the last daily question)

(See screen concept on next page)

Body Area:

Select on body map (Figure 1)

In the last 12 months, have you had new bone formation or decrease of joint movement without noticeable pain, swelling, warmth, or redness in that area of the body?

- ☐ If yes, please describe the location *(Select the body area on a body map)*
- ☐ If no—diary ends.

Daily: In the last 24 hours, have you had new bone formation or decrease of joint movement without noticeable pain, swelling, warmth, or redness in that area of the body?

- ☐ If yes, please describe the location *(Select the body area on a body map)*
- ☐ If no—diary ends.

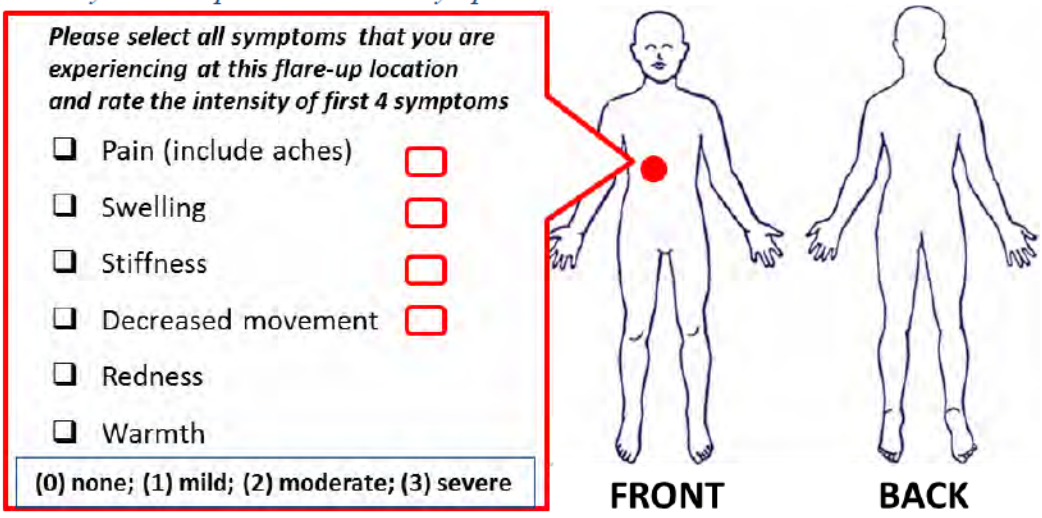
- If end date is selected, area and outcome screen will be presented
- If patient selected ongoing for end date – area, symptom, and severity screens are presented

Outcome of the flare-up: *(Scroll and select from drop-down box)*

- ☐ Flare-up is ongoing
- ☐ Flare-up is resolved and new bone formed
- ☐ Flare-up is resolved but no new bone formed
- ☐ Flare-up is resolved and joint movement is affected
- ☐ Flare-up is resolved and joint movement is not affected

Symptoms and severity (for ongoing flare-up)

Severity scale is presented when symptom is selected



10.6. SAS Code for EQ-5D Index Utility Scoring

```
/*=================================================================*/;
/* Aim : Derive the EQ5-D index (utility)
*/;
/* Source for the algorithm : scoring EQ-5D health states [redacted]
*/;
/* (cf.[redacted])
*/;
/* Note : UK based population [redacted] 1997)
*/;
/* Author : [redacted]
*/;
/*=================================================================*/;
data EUROQOL; set temp;
profil= (10000* mob_std)+(1000* self_std)+(100* usual_std)+(10*
pain_std)+anxiety_std;
eq5d=1;
**********Mobility*********;
if mob_std=2 then eq5d=eq5d-0.069;
if mob_std=3 then eq5d=eq5d-0.314;
********Self-care**********;
if self_std=2 then eq5d=eq5d-0.104;
if self_std=3 then eq5d=eq5d-0.214;
******Usual activities*****;
if usual_std=2 then eq5d=eq5d-0.036;
if usual_std=3 then eq5d=eq5d-0.094;
*****Pain/discomfort******;
if pain_std=2 then eq5d=eq5d-0.123;
if pain_std=3 then eq5d=eq5d-0.386;
*****Anxiety/depression*****;
if anxiety_std=2 then eq5d=eq5d-0.071;
if anxiety_std=3 then eq5d=eq5d-0.236;
if (mob_std ne 1 or self_std ne 1 or usual_std ne 1 or pain_std ne 1 or
anxiety_std ne 1)
then eq5d=eq5d-0.081;
if (mob_std=3 or self_std=3 or usual_std=3 or pain_std=3 or anxiety_std=3)
then eq5d=eq5d-0.269;
if (mob_std=. or self_std=. or usual_std=. or pain_std=. or anxiety_std=.)
then eq5d=.
```

10.7. Cumulative Analogue Joint Involvement Scale (CAJIS) for FOP

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

Area Assessed	Area Assessed/Coded	Axial	Upper Limb	Lower Limb
Neck	1	*		
Thoracic & Lumbar spine	2	*		
Jaw	3	*		
Right shoulder	4		*	
Left shoulder	5		*	
Right elbow	6		*	
Left elbow	7		*	
Right wrist	8		*	
Left wrist	9		*	
Right hip	10			*
Left hip	11			*
Right knee	12			*
Left knee	13			*
Right ankle	14			*
Left ankle	15			*

Signature Page for VV-RIM-00130464 v1.0

ESig Approval	[redacted]00
ESig Approval	[redacted]2020 15:33:16 GMT+0000
ESig Approval	[redacted]00
ESig Approval	[redacted]00
ESig Approval	[redacted]2020 12:34:45 GMT+0000
ESig Approval	[redacted]2020 14:21:26 GMT+0000

Signature Page for VV-RIM-00130464 v1.0 Approved

Signature Page for VV-RIM-00130464 v1.0

ESig Approval	[redacted]20 14:54:36 GMT+0000

Signature Page for VV-RIM-00130464 v1.0 Approved