STATISTICAL ANALYSIS PLAN

NCT #NCT04064411

A Randomized, Non-inferiority, Phase 3, Open-label, Multicenter Study to Evaluate the Efficacy and Safety of Abaloparatide-sMTS for the Treatment of Postmenopausal Women with Osteoporosis

Protocol BA058-05-021

Protocol Version and Date:	Amendment 3 (15 November 2019) Amendment 3.1 (EU Version 15 January 2020)
Name of Test Drug:	Abaloparatide-sMTS
Phase:	Phase 3
Methodology	International, Multicenter, Randomized, Active Control, Open-label
Study Sponsor:	Radius Health, Inc. (RADIUS) 22 Boston Wharf Road, 7th floor Boston, MA 02210 Tel: (617) 551-4700 Fax: (617) 551-4701
Analysis Plan Date:	17 September 2021
Analysis Plan Version:	Final V1.0

Protocol Title:	A Randomized, Non-inferiority, Phase 3, Open-label, Multicenter Study to Evaluate the Efficacy and Safety of Abaloparatide-sMTS for the Treatment of Postmenopausal Women with Osteoporosis
Protocol Number:	BA058-05-021, Amendment 3 (15 November 2019) BA058-05-021, Amendment 3.1 (EU Version 15 January 2020)
Study Drug:	Abaloparatide-sMTS
SAP Version:	Final V1.0 (17 September 2021)

Prepared By:

[redacted] Date
Radius Health, Inc.

[redacted] Date
Radius Health, Inc.

REVIEWED AND APPROVED BY:

[redacted] Date
Radius Health, Inc.

[redacted] Date
Radius Health, Inc.

[redacted] Date
Radius Health, Inc.

TABLE OF CONTENTS

TABLE OF CONTENTS 3
LIST OF TABLES 6
LIST OF FIGURES 6
1. BACKGROUND 10
1.1. Study Objectives 10
1.2. Study Design 10
1.2.1. Synopsis of Study Design 10
1.2.2. Randomization Methodology 11
1.2.3. Blinding 11
1.2.4. Study Procedures 12
1.2.5. Study Endpoints 17
1.2.5.1. Efficacy endpoints 17
1.2.5.2. Safety and Tolerability Endpoints 17
2. SUBJECT ANALYSIS POPULATIONS 18
2.1. Population Definitions 18
2.2. Protocol Deviation 18
3. GENERAL STATISTICAL METHODS 20
3.1. Sample Size Planned and Specified in the Protocol 20
3.2. General Methods 20
3.3. Computing Environment 21
3.4. Baseline Definitions 21
3.5. Data Pooling 21
3.6. Adjustments for Covariates 21
3.7. Multiple Comparisons 21
3.8. Subgroups 21
3.9. Anti-Abaloparatide Antibody Status 22
3.10. Withdrawals, Dropouts, and Loss to Follow-up 23
3.11. Missing, Unused, and Spurious Data 23
3.12. Visit Windows 23
3.13. Interim Analyses 24
4. STUDY ANALYSES 25

4.1. Subject Disposition 25
4.2. Demographics and Baseline Characteristics 25
4.3. Exposure to Study Drug 27
4.3.1. Extent of Exposure and Treatment Compliance 27
4.3.1.1. Residual Drug Amount of Abaloparatide-sMTS 31
4.3.2. Abaloparatide-sMTS Patch Adhesion Scores 32
4.3.3. Patch Wear Time 33
4.3.4. Nutritional Supplements Exposure 34
4.3.5. Concomitant Medication 34
4.3.6. Concomitant Procedures 34
4.4. Efficacy Analyses 35
4.4.1. Primary Efficacy Endpoint 35
4.4.1.1. Primary Analysis 35
4.4.1.2. Sensitivity Analysis 36
4.4.1.3. Subgroup Analysis 38
4.4.1.4. Analysis by Anti-Abaloparatide Antibody Status 39
4.4.2. Secondary Efficacy Endpoints 39
4.4.3. Additional Efficacy Endpoints 39
4.4.4. Other Endpoints 40
4.4.5. Exploratory Efficacy Endpoints 40
4.4.5.1. BMD Responders 40
4.4.5.2. Time to Clinical Fracture 41
4.4.5.3. Changes in Disease Status 41
4.5. Population Pharmacokinetics (PK) 42
4.6. Safety Analyses 43
4.6.1. Adverse Events 43
4.6.2. Adverse Events of Special Interest 45
4.6.3. Anti-Drug Antibody 47
4.6.4. Local Tolerance 47
4.6.4.1. Investigator Assessment of Local Tolerance 48
4.6.4.2. Subject Assessment of Local Tolerance 49
4.6.5. Laboratory Data 50

4.6.5.1. Potentially Clinically Significant Laboratory Values 51
4.6.5.2. Hypercalcemia 51
4.6.5.3. Hypercalciuria 52
4.6.5.4. Creatinine Clearance and Calcium Excretion 52
4.6.6. Vital Signs 53
4.6.6.1. Potentially Clinically Significant Vital Sign Values 54
4.6.7. Electrocardiogram 54
5. CHANGES FROM THE PROTOCOL-SPECIFIED ANALYSIS 56
6. REFERENCES 57
APPENDIX A. CONVENTIONS FOR IMPUTING MISSING DATES 58
APPENDIX B. ALGORITHM TO IDENTIFY CRITICAL PROTOCOL DEVIATIONS THAT EXCLUDE SUBJECTS FROM THE PER-PROTOCOL (PP) POPULATION 61
APPENDIX C. CRITERIA FOR POTENTIALLY CLINICALLY SIGNIFICANT (PCS) LABORATORY VALUES 62
APPENDIX D. NORMALIZATION OF CLINICAL LABORATORY VALUES 63

LIST OF TABLES

Table 1: Schedule of Assessments and Procedures 13
Table 2: Allowable Time Period for Collection of BMD Assessments.....23
Table 3: Analysis Visit Windows for BMD, s-PINP and s-CTX Assessments.....24
Table 4: Visit Windows for DXA Assessments for Sensitivity Analyses. 38
Table 5: Criteria for PCS of Vital Sign.....54

LIST OF FIGURES

Figure 1: Study BA058-05-021 Study Design.....11

LIST OF ABBREVIATIONS

Abbreviation	Term
µg	Microgram
µmol	Micromole
ADA	Anti-drug antibody
AE	Adverse event
AESI	Adverse event of special interest
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical class
BMD	Bone mineral density
BMI	Body mass index
BP	Blood pressure
bpm	Beats per minute
CI	Confidence interval
cm	Centimeters
CM	Concomitant medication
CRO	Contract research organization
CSR	Clinical study report
DBP	Diastolic blood pressure
DXA	Dual energy X-ray absorptiometry
ECG	Electrocardiogram
eCRF	Electronic case report form
EOS	End of Study
EOT	End of Treatment
FDA	Food and Drug Administration
FSH	Follicle-stimulating hormone
g	Gram
IRT	Interactive response technology
ITT	Intention-to-treat
IU	International units
kg	Kilograms
kg/m^2	Kilogram per square meter
L	Liter

Abbreviation	Term
LOCF	Last observation carried forward
MDR	Medical device reporting
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
mITT	Modified intention-to-treat
mL	Milliliter
mmHg	Millimeter of mercury
mmol	Millimoles
MMRM	Mixed-Effect Model Repeated Measures
msec	Millisecond
NAb	Neutralizing antibody
NI	Non-inferiority
PAP	Pharmacometric Analysis Plan
PCS	Potentially clinically significant
PD	Pharmacodynamics
PK	Pharmacokinetic(s)
PP	Per protocol
PT	Preferred Term
PTH	Parathyroid hormone
PTHrP	Parathyroid hormone related peptide
PTT	Partial thromboplastin time
Q1	First quartile
Q3	Third quartile
QT	Total depolarization and repolarization time
QTcF	Total depolarization and repolarization time corrected with heart rate (using Fridericia's correction formula)
REML	Restricted maximum likelihood
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SBP	Systolic blood pressure
SC	Subcutaneous
s-CTX	Serum carboxy-terminal cross-linking telopeptide of type I collagen
SD	Standard deviation
SE	Standard error
SI	Standard international

Abbreviation	Term
SMQ	Standardized MedDRA Query
sMTS	solid Microstructured Transdermal System
SOC	System Organ Class
s-PINP	Serum procollagen type I N-terminal propeptide
TEAEs	Treatment emergent adverse events
TSH	Thyroid stimulating hormone
UADE	Unanticipated adverse device effect
ULN	Upper limit of normal
USPI	United States Package Insert
WHO-DD	World Health Organization Drug Dictionary

1. BACKGROUND

This statistical analysis plan (SAP) enhances the statistical considerations specified in the protocol for non-pharmacokinetic (PK) data for Study BA058-05-021. If considerations are substantially different, they will be identified. Any post-hoc or unplanned analyses, or any significant changes from the planned analysis in this SAP will be clearly identified in Section 9.8 of the clinical study report (CSR). The final SAP will be included as an appendix to the final CSR.

The method for analyzing the PK data will be described in the Pharmacometric Analysis Plan (PAP) as a separate document.

1.1. Study Objectives

The objectives of this study are:

1. To evaluate the non-inferiority of abaloparatide-solid Microstructured Transdermal System (abaloparatide-sMTS) 300 µg compared to abaloparatide-SC 80 µg based on lumbar spine bone mineral density (BMD) at 12 months.
2. To evaluate the safety and tolerability of abaloparatide-sMTS in the treatment of postmenopausal women with osteoporosis.

1.2. Study Design

1.2.1. Synopsis of Study Design

This is a randomized, open-label, non-inferiority, multicenter study of abaloparatide-sMTS compared with abaloparatide-SC for the treatment of postmenopausal women with osteoporosis.

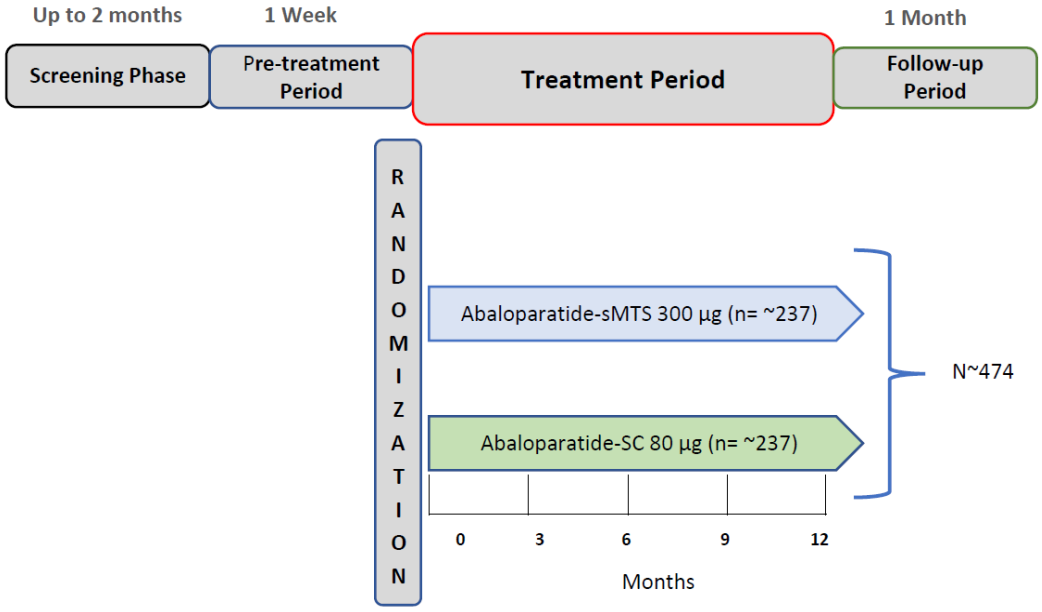
The study will consist of a Screening Period (up to 2 months), a Pretreatment Period (1 week), and a Treatment Period (12 months) with a final visit 1 month after the last dose of study drug (Follow-Up/End of Study [EOS] visit). 1 month is defined as 30 days in this study. Thus, subjects will participate in the study for up to 16 months.

Eligible subjects will be randomized in a 1:1 treatment ratio to either abaloparatide-sMTS or abaloparatide-SC using a permuted block randomization scheme. During the Treatment Period, subjects will self-administer study drug once a day for 12 months and visit the study site on Days 1 and 14 and Months 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and Month 13 for the follow-up visit.

All subjects will be provided calcium and Vitamin D to ensure that their daily intake is 1200 mg/day and 800 IU/day, respectively, or a dose determined by the investigator and agreed by the Sponsor Medical Monitor according to the subject's need.

The study design is presented in Figure 1.

Figure 1: Study BA058-05-021 Study Design



1.2.2. Randomization Methodology

It is planned to enroll approximately 474 subjects at approximately 125 study centers globally to receive either abaloparatide-sMTS 300 µg or abaloparatide-SC 80 µg in a 1:1 ratio on Day 1 using permuted block randomization.

Once a subject has been deemed eligible to be randomized into the study on Day 1 (Visit 3), the study center will use the Interactive Response Technology (IRT) system to assign a medication kit to the subject. Throughout the study, medication kits will be assigned to the subject based on information provided by the IRT.

1.2.3. Blinding

This is an open-label study; thus, study subjects and investigators will not be blinded to treatment assignment. However, to minimize bias in study conduct, RADIUS personnel performing statistical analyses will be blinded to treatment assignment and aggregated data by treatment assignment until after database lock. The Contract Research Organization (CRO) study team members and select RADIUS team members (including those responsible for trial management, data management, medical monitoring and pharmacovigilance) will not be blinded to an individual subject's treatment assignment during the conduct of the study but will be blinded to aggregated data by treatment assignment until after database lock.

In addition, the central imaging laboratory responsible for measuring BMD will be blinded to treatment assignment throughout the study. RADIUS will not receive post-baseline BMD measurements until database lock.

Further details regarding the blinding of study data are described in a separate Blind Management Plan document.

1.2.4. Study Procedures

The schedule of assessments, as outlined in the study protocol, is provided in Table 1.

Table 1: Schedule of Assessments and Procedures

Procedure / *Visit* / *Study Day/ Month (Mo.):*	1 Screen	2 Pre-TX	3 Day 1	4 Day 14	5 Mo. 1	6 Mo. 2	7 Mo. 3	8 Mo. 4	9 Mo. 5	10 Mo. 6	11 Mo. 7	12 Mo. 8	13 Mo. 9	14 Mo. 10	15 Mo. 11	16 Mo. 12/ EOT	17 Follow-up/ EOS[21]
Visit Window (Days)	N/A	N/A	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4
Duration	Up to 2 mo.	1 week	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Informed consent	X																
Verification of entry criteria	X	X															
Physical examination[1]	X																
Review of medical history[2]	X																
Symptom-directed physical examination		X	X		X		X			X			X			X	X
Vital signs[3]	X	X	X		X		X			X			X			X	X
Weight measurement	X		X		X					X						X	
Height measurement[4]	X	X	X		X		X			X			X			X	X
Electrocardiogram[5]	X		X													X	X
Urinalysis (dipstick)[6]	X		X		X					X						X	
Chemistry blood collection[7]	X		X		X					X						X	
Hematology blood collection[7]	X		X		X					X						X	
Coagulation (PT and PTT) blood collection[7]	X															X	
PTH(1–84) [7]	X															X	
25-hydroxyvitamin D level[7]	X															X	
1,25-dihydroxyvitamin D level			X													X	
Estradiol, FSH	X																
Thyroid stimulating hormone[8]	X																
Study drug assignment via IRT			X														

Procedure / *Visit* / *Study Day/ Month (Mo.):*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Screen	Pre-TX	Day 1	Day 14	Mo. 1	Mo. 2	Mo. 3	Mo. 4	Mo. 5	Mo. 6	Mo. 7	Mo. 8	Mo. 9	Mo. 10	Mo. 11	Mo. 12/ EOT	Follow-up/ EOS[21]
Visit Window (Days)	N/A	N/A	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4
Duration	Up to 2 mo.	1 week	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Subject training for study drug administration			X														
Calcium and vitamin D supplements		← Daily Administration →															
Study Drug administration			← Daily Administration →														
Serum markers of bone metabolism (s-PINP and s-CTX)			X		X		X			X						X	
Lumbar and thoracic spine radiographs[9]	X																
Serum calcium and albumin[10]	X		X		X		X			X			X				
24-hour urine collection (for calcium:creatinine and creatinine clearance)[11]			X				X										
Symptom-driven spinal radiologic assessment		← At any Time →															
Clinical assessment of new fractures[12]		X	X		X		X			X			X			X	
BMD of lumbar spine, total hip and femoral neck by DXA[13]	X						X			X						X	
ADA testing[14]			X	X	X		X			X			X			X	X
Assessment of local tolerance[15]			X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Assessment of patch adhesion[16]			← Daily Assessment →														
Subject diary review[17]			X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Document AEs and concomitant medication[18]	← At Any Time, Question Subjects at Every Visit →																
Dispense study drug			X		X	X	X	X	X	X	X	X	X	X	X		
Sparse PK sampling[19]			X	X	X		X			X			X				

Procedure	*Visit*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Study Day/ Month (Mo.):	Screen	Pre-TX	Day 1	Day 14	Mo. 1	Mo. 2	Mo. 3	Mo. 4	Mo. 5	Mo. 6	Mo. 7	Mo. 8	Mo. 9	Mo. 10	Mo. 11	Mo. 12/ EOT	Follow -up/ EOS[21]
Visit Window (Days)		N/A	N/A	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4	± 4
Duration		Up to 2 mo.	1 week	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Collect used patch and swab for abaloparatide-sMTS[20]				X	X	X		X			X			X				

1. A complete physical examination includes a review of the following systems: head/neck/thyroid, eyes/ears/nose/throat, respiratory, cardiovascular, lymph nodes, abdomen, skin, musculoskeletal, and neurological. Breast, anorectal, and genital examinations will be performed when medically indicated. After screening, any clinically significant abnormal findings in physical examinations should be reported as AEs.
2. Including alcohol and tobacco use assessment.
3. Blood pressure, pulse rate, body temperature, and respiration rate are to be recorded predose at each study visit. Only blood pressure, pulse rate, and respiration rate are to be recorded 1 hour after study drug administration at each study visit. All blood pressure assessments will be orthostatic.
4. Height is to be measured in the standing position using a medical stadiometer.
5. ECGs are to be performed predose and also at 1 hour after study drug administration on Day 1 and Month 12.
6. All routine urinalysis will be performed on a sample freshly voided during the visit and sent to a central lab for microscopy if test is positive for micro-organisms via dipstick.
7. Screening labs may be analyzed either at a local laboratory or sent to the central laboratory; central laboratory testing must be performed at all other study visits. Blood draws to be done prior to dosing during the treatment period.
8. Any subject with a TSH value outside of the normal range may have T3 and free T4 tested, with results within the normal range in order to be enrolled.
9. X-ray results will be evaluated by a qualified evaluator at the site, with prevalent fractures assessed using the Genant-semiquantitative scoring method.
10. Serum calcium and albumin will be measured predose from the standard chemistry panel on Screening, Day 1, Month 1 and Month 6 and from a separate blood draw predose at Month 3 and Month 9 and at 4 hours postdose at Month 1. Albumin corrected serum calcium will be reported using the serum calcium and albumin results on the following visits: Screening, Day 1, Month 1, Month 3, Month 6 and Month 9.
11. A 24-hour urine collection will be collected at Day 1 and Month 3 and will be used for urinary calcium and urinary creatinine measurements. Subjects will discard the 1st void and begin a 24-hour urine collection on the day prior to the study site visit.
12. Spine radiographs at screening will be assessed locally. If the subject reports that a fracture has occurred, remind the subject to bring X-rays and any medical reports of the fracture to the next study site visit. Documentation must be obtained on all new fractures that occur during the study. This documentation should be maintained in the source documents.
13. Each DXA for a given subject should be performed on the same machine, and if available, preferably by the same technician and will be assessed by a central imaging laboratory. For screening purposes, DXA scans of the lumbar spine, total hip, and femoral neck taken up to 35 days prior to the beginning of the Screening Period may be used to enroll a subject in the study if the BMD T-score criteria in protocol Section **Error! Reference source not found.** are met.
14. Samples for ADA will be drawn prior to treatment on Day 1, Day 14, Month 1, Month 3, Month 6, Month 9, Month 12/EOT, and 1 month following the last dose of study drug. All subjects who remain antibody positive will continue to be followed for ADA testing every 6 months until the result is negative.
15. Investigators will perform an assessment of the injection/application site at each study site visit prior to study drug administration, 5 minutes after study drug administration (ie, immediately after patch removal for subjects receiving abaloparatide-sMTS), and 1- hour following study drug administration. Signs of local tolerance will be assessed using a 4-point scale described in protocol Section **Error! Reference source not found.**. The subject will maintain a diary of their daily assessment of local tolerance; the application/administration site will be evaluated prior to study drug administration, 5 minutes after study drug administration (ie, immediately after patch removal for subjects receiving abaloparatide-sMTS), and 1 hour following study drug administration daily from Day 1 to the Month 12/EOT visit. Symptoms of local tolerance will be assessed using a 4-point scale described in protocol Section **Error! Reference source not found.**.

16. Patch adhesion will be assessed by the subject immediately prior to abaloparatide-sMTS patch removal each day during the Treatment Period.
17. The subject diary will be reviewed by study personnel at each study visit to ensure subject compliance.
18. AEs and SAEs will be recorded on the eCRF starting from the signing of the informed consent until 30 days after the last dose of study drug. All treatment-related AEs will be followed until resolution or stabilization. Any SAEs that occur at any time after completion of the study, which are considered by the Investigator to be related to study treatment, must be reported to the sponsor or its designee.
19. Blood samples for measurement of plasma concentrations of abaloparatide will be taken at Day 1, Day 14, Month 1, Month 3, Month 6, and Month 9. One sample per subject per visit will be collected at 1 of the following varying post-injection/application times: 1 hour to 2 hours (Day 1); 30 minutes to 1 hour (Day 14); 3 hours to 4 hours (Month 1); 2 hours to 3 hours (Month 3); 10 minutes to 30 minutes (Month 6) and predose (Month 9). The date and time of injection/patch application and the date and time of PK blood sample collection will be recorded on the CRF.
20. Used patches from the study site visits (Day 1, Day 14, Months 1, 3, 6, and 9) including the swab used will be placed in separate vials and shipped to 3M.
21. Follow-up/ EOS visit is 1 month (±4 days) after the Month 12/EOT Visit.

ADA = Anti-drug antibody; AE=adverse event; BMD=bone mineral density; DXA=dual energy X-ray absorptiometry; eCRF = Electronic case report form; EOS = End of study; EOT = End of treatment; FSH=follicle-stimulating hormone; IRT=Interactive Response Technology; NA=not applicable; PK = Pharmacokinetic; Pre-TX = Pretreatment; PT=prothrombin time; PTH=parathyroid hormone; PTT=partial thromboplastin time; SAEs = Serious adverse events; s-CTX=serum carboxy-terminal cross-linking telopeptide of type 1; sMTS=solid microstructured transdermal system; s-PINP=serum procollagen type 1 N peptide; TSH = Thyroid stimulating hormone.

1.2.5. Study Endpoints

1.2.5.1. Efficacy endpoints

Primary efficacy endpoint:

- Percent change from baseline in lumbar spine BMD at 12 months.

Secondary efficacy endpoints:

- Percent change from baseline in total hip BMD at 12 months
- Percent change from baseline in femoral neck BMD at 12 months

Additional efficacy endpoints:

- Percent change from baseline in
  - Lumbar spine BMD at 3 and 6 months
  - Total hip BMD at 3 and 6 months
  - Femoral neck BMD at 3 and 6 months

Other Endpoints:

- Log ratio of post-baseline over baseline in:
  - serum procollagen type I N-terminal propeptide (s-PINP) at 1, 3, 6, and 12 months
  - serum carboxy-terminal cross-linking telopeptide of type I collagen (s-CTX) at 1, 3, 6, and 12 months
- Plasma concentration of abaloparatide from sparse PK sampling

Exploratory efficacy endpoints:

- Proportion of subjects experiencing BMD increase from baseline of >0%, >3%, and >6% at the lumbar spine, femoral neck, and total hip at 3, 6, and 12 months
- Incidence of new clinical fracture at 12 months
- Changes in disease status - proportion of subjects converting from the categories of osteoporosis at baseline to osteopenia or to normal at end of study, where:
  - Osteoporosis is defined as lumbar spine or total hip BMD T-score ≤ -2.5,
  - Osteopenia is defined as:
    - lumbar spine T-score > - 2.5 and total hip BMD T-score > -2.5 and < -1.0, or
    - lumbar spine T-score > - 2.5 and < - 1.0 and total hip BMD T-score > -2.5,
  - Normal is defined as lumbar spine T-score ≥ -1.0 and total hip BMD T-score ≥ -1.0.

1.2.5.2. Safety and Tolerability Endpoints

Safety and tolerability evaluations will be based on treatment-emergent adverse events (TEAE), including AEs of special interest (AESI), vital signs (orthostatic blood pressure, heart rate, body temperature, and respiration rate), electrocardiograms (ECGs), laboratory tests (serum chemistry, hematology, coagulation, and urinalysis), investigator assessment of local tolerance, subject assessment of local tolerance, and presence of anti-drug antibodies.

In addition, the presence of hypercalcemia and of hypercalciuria based on laboratory values will be assessed.

2. SUBJECT ANALYSIS POPULATIONS

2.1. Population Definitions

Five subject populations will be considered in the statistical analyses of this study.

- **Intention-to-Treat (ITT) population**

  The ITT population consists of all subjects randomized into this study.

- **Modified Intention-to-Treat (mITT) population**

  The mITT population includes all randomized subjects who received at least one dose of study drug and had a baseline lumbar spine BMD measurement and at least one post-baseline lumbar spine BMD measurement.

- **Safety population**

  The Safety population consists of all randomized subjects who received at least one dose of study drug.

- **Per-Protocol (PP) population**

  The PP population consists of all mITT subjects who did not have any critical protocol deviations (described in Section 2.2).

- **Antibody Population**

  The Antibody population will include subjects in the Safety population who have at least one post-baseline anti-drug antibody assessment.

2.2. Protocol Deviation

A protocol deviation is any change, divergence, or departure from the study design or procedures defined in the protocol.

The protocol deviation categories related to study conduct are described in the Non-Compliance Management section of the Clinical Monitoring Plan and are designated as major or minor. Study conduct-related protocol deviations are recorded by the contract research organization (IQVIA Biotech) in their Clinical Trial Management System (CTMS). If an entire study visit is not done, IQVIA Biotech will record this major protocol deviation on CTMS but will not record the missed assessments associated with that missed visit as protocol deviations (e.g., if a Month 3 visit was not done, IQVIA Biotech will record the protocol deviation of Month 3 visit not done but will not record the protocol deviations for the following assessments that were subsequently not done at Visit 3: Dual energy X-ray absorptiometry (DXA) assessment, vital signs, investigator assessment of local tolerance, sample collection for: bone turnover markers, serum calcium and albumin, ADA, PK, and 24-hour urine). For protocol deviations associated with a study visit not done, the following rules apply:

- A study visit not done, according to the established schedule, cannot occur at the Screening visit (Visit 1) or Day 1 (Visit 3).
- If a protocol deviation of "Study visit not done" is reported, Biometrics will programmatically provide the other protocol deviations pertaining to missed assessments that are associated with the visit not done. These include missing assessments for the following, depending on the visit:

- DXA
- Local skin reaction scores
- ECG
- Safety labs (including serum calcium and albumin) and 24-hour urine collection
- All other blood sample collections, including for the following:
  - s-PINP and s-CTX
  - Anti-drug antibody
  - PK
  - Coagulation
  - 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D levels
- Vital signs, height, weight
- Clinical assessment of new fractures.

Subjects with protocol deviations, as recorded on the IQVIA Biotech CTMS or supplemented by Biometrics (as described above), will be presented in a data listing, with indicators for major and minor protocol deviations (as defined in the Clinical Monitoring Plan). A summary table of the major protocol deviations identified during the study will also be presented.

Critical protocol deviations are protocol deviations that have the potential to impact the efficacy of the treatment on the subject.

The algorithm to determine critical protocol deviations is in Appendix B of this SAP. Subjects with at least one critical protocol deviation will be excluded from the PP population. A Per-Protocol Memo with the algorithm in Appendix B together with the programming specification ("Per Protocol Population Data Set Development") will be developed and finalized prior to database lock to detail how each of the criteria in the algorithm can be programmed in SAS. The list of subjects with critical protocol deviations identified after applying the algorithm to the SDTM data after database lock will be reviewed and approved internally at Radius in a separate memo. A summary table of reasons for exclusion from the PP population will be presented. A data listing for subjects excluded from the PP population, together with the exclusion reason(s), will be provided.

3. GENERAL STATISTICAL METHODS

3.1. Sample Size Planned and Specified in the Protocol

The previous pivotal, Phase 3, multicenter, randomized, open-label active- and double-blind placebo-controlled study of abaloparatide-SC (Study BA058-05-003) showed that the placebo-adjusted effect of abaloparatide-SC 80 µg on the percent change from baseline in lumbar spine BMD based on a Mixed-Effect Model Repeated Measures (MMRM) analysis was 9.096% (95% CI: 8.557%, 9.634%) at 12 months. According to FDA's *Guidance for Industry: Non-Inferiority Clinical Trials to Establish Effectiveness* (FDA, 2016), the lower bound of this confidence interval (CI) can be considered as the historical treatment effect of abaloparatide-SC versus placebo, M_1 (=8.557%). Based on FDA's recommendation of a clinically meaningful difference of 2.0% between treatment groups, a non-inferiority (NI) margin, M_2, is selected at 2.0% to preserve approximately 77% of M_1, the historical treatment effect of abaloparatide-SC.

This preserved effect supports the superiority to placebo because the amount of preserved effect, 6.557% (=77% × M_1) is larger than the placebo-adjusted effect of alendronate 10 mg daily (5.4%; Fosamax® USPI) and of denosumab 60 mg once every 6 months (5.5%; Bolognese et al, 2013) on lumbar spine BMD at 12 months. Furthermore, the placebo-adjusted effect of teriparatide observed in the previous Phase 3 study (Study BA058-05-003) based on an MMRM analysis was 7.841% (95% CI: 7.384%, 8.297%). Thus, the proposed NI margin will preserve 89% (=6.557/7.384) of the effect of teriparatide, another market-approved anabolic agent.

Non-inferiority of abaloparatide-sMTS to abaloparatide-SC will be concluded if the lower bound of the two-sided 95% CI for the estimated treatment difference (abaloparatide-sMTS minus abaloparatide-SC) in the percent change from baseline in lumbar spine BMD at 12 months is above -2.0%, using an MMRM analysis.

A sample size of 426 subjects will provide at least 90% power to conclude the non-inferiority of abaloparatide-sMTS 300 µg to abaloparatide-SC 80 µg, assuming a true mean difference of 0% and a standard deviation of 6.35%. To ensure an analysis size of 426 subjects, an overall sample size of 474 subjects (237 subjects per group) will be randomized, anticipating that approximately 10% of treated subjects may not have both a baseline lumbar spine BMD measurement and at least one post-baseline lumbar spine BMD measurement.

3.2. General Methods

Summary tables will be presented by treatment group (abaloparatide-sMTS and abaloparatide-SC) and overall, when appropriate. For categorical data, summary tabulations of the number and percentage of subjects within each category of the parameter will be presented. When tabulating categorical data, "missing" will be included as a category and the number of subjects with missing data will be presented. For continuous data, the number of subjects, mean, standard deviation (SD), median, minimum, interquartile range (Q1 and Q3), and maximum will be presented.

All data listings that contain an evaluation date will contain a relative study day. Screening, pretreatment, and on-treatment study days will be numbered relative to the day of the first dose of study drug which is designated as Day 1. Study days prior to the first dose of study drug will be calculated as: [date of assessment – date of first dose of study drug]. Study days on or after

the first dose of study drug will be calculated as: [date of assessment – date of first dose of study drug + 1].

Unless otherwise specified, all statistical tests will be two-sided with a significance level of 5%.

Where applicable, CIs about a parameter estimate will be presented using the same number of decimal places as the parameter estimate, and p-values will be rounded to 4 decimal places prior to assessment of statistical significance.

The anti-drug antibody assessment in the serum samples for immunogenicity testing will be analyzed and reported separately. All the planned efficacy and safety analyses related to immunogenicity results and antibody status will also be reported separately.

This study is being conducted during the COVID-19 pandemic. The COVID-19 outbreak emerged in December 2019 and quickly became a global pandemic as declared by the World Health Organization in early March 2020. At that time, a variety of mitigations to assure safety of participants and address operational issues were implemented by RADIUS. To investigate the effect of the COVID-19 pandemic on study conduct, additional analyses will be conducted as described in the efficacy and safety analysis sections.

3.3. Computing Environment

All descriptive and statistical analyses will be performed using SAS statistical software Version 9.4 or later version. Medical history and AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), Version 24.0. Concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD) March 2018 B3.

3.4. Baseline Definitions

The baseline value is defined as the last non-missing value obtained prior to the first dose of study drug, including the screening value, if necessary. If a randomized subject is not treated, the baseline value is defined as the last non-missing assessment.

3.5. Data Pooling

Data from all study sites and countries will be pooled together for all analyses, unless otherwise specified.

3.6. Adjustments for Covariates

Adjustments for covariates will be employed in the analysis of covariance (ANCOVA) and MMRM analyses. Any adjustments will be pre-specified in the description of the statistical analysis (Section 4.4).

3.7. Multiple Comparisons

No adjustment for multiple comparisons is required for the primary efficacy endpoint. No p-value adjustments for multiple comparisons will be made for all other efficacy endpoints.

3.8. Subgroups

The following subgroups may be used in selected data analyses. Section 4.4 will describe the specific analyses to be performed for each subgroup.

- Age (< 65, 65 to < 75, ≥ 75 years)
- Race (White, Black/African-American, Asian, Other Races [i.e., American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Multiple (more than 1 race category chosen), Other])
- BMI (< 25, 25 - < 30, ≥ 30)
- Region (North America, Europe)
- Smoking status (Never, Current, Former)
- Lumbar spine BMD T-score at baseline (≤ -2.5, > -2.5)
- Total hip BMD T-score at baseline (≤ -2.5, > -2.5)
- Femoral neck BMD T-score at baseline (≤ -2.5, > -2.5)
- Prior clinical fractures (Yes = Subjects with a record in medical history with "fracture" in the preferred term, No)
- Prevalent vertebral fractures (Yes = Subjects with any vertebral fracture with a Genant semi-quantitative score > 0, No)
- Prior Fracture (Yes =Subjects with a prevalent vertebral fracture and /or a prior clinical fracture, No)
- s-PINP at baseline (tertile values calculated based on ITT population)

3.9. Anti-Abaloparatide Antibody Status

At Day 1, Day 14, Month 1, Month 3, Month 6, Month 9, Month 12/EOT, and Month 13/EOS, the anti-abaloparatide antibody (ADA) assessment in the serum samples for immunogenicity testing will be categorized by two outcomes:

- Negative for the ADA (ADA-)
- Positive for the ADA (ADA+)

For those subjects with a positive ADA (i.e., ADA is present), the serum samples will be further assessed for the presence of neutralizing anti-abaloparatide antibodies (NAb), and the results will be categorized by two outcomes:

- Negative for the NAb (NAb-)
- Positive for the NAb (NAb+)

Subject ADA status and presence of NAb by Month 12 (ADA-, ADA+/NAb+, ADA+/NAb-, All ADA+) will be defined as follows:

- Subjects will be classified as ADA- if they were ADA- in all post-baseline immunogenicity assessments during the 12-month treatment period.
- Subjects will be classified as ADA+ and NAb+ if they were ADA+ at any post-baseline immunogenicity assessment and NAb+ at any subsequent tests during the 12-month treatment period.
- Subjects will be classified as ADA+ and NAb- if they were ADA+ at any post-baseline immunogenicity assessment and NAb- at all subsequent tests during the 12-month treatment period.

3.10. Withdrawals, Dropouts, and Loss to Follow-up

Subjects who are withdrawn or discontinue from the study will not be replaced.

3.11. Missing, Unused, and Spurious Data

Statistical methods for addressing missing data for the analyses of the primary and secondary efficacy endpoints are described in Section 4.4.

Rules for imputation of missing dates for onset of AEs and start dates of concomitant medications are included in Appendix A of this SAP. Unless otherwise specified, missing dates for onset of AEs and start of concomitant medications will be handled according to the imputation rules.

In the event that incorrect data are discovered after database lock, the impact of such incorrect data will be assessed and a determination will be made as to whether the data need to be corrected in the database (and subsequently in the analysis). In cases where it is decided that the data need to be corrected, this decision will be documented, and the documentation will be saved in the Trial Master File.

All available data will be included in data listings.

3.12. Visit Windows

In general, the nominal visit will be used for all efficacy and safety data summaries and analyses.

During the COVID-19 pandemic, if a subject's DXA assessment cannot be performed during the protocol specified visit windows, the site will contact the study Medical Monitor, and the subject will be allowed to make up the visits within an allowed time period. The allowable time period for collection of DXA assessments is shown in Table 2, where the ideal visit day is defined as the ideal number of days from Day 1, when the first dose of study drug was administered, to the nominal visit.

Table 2: Allowable Time Period for Collection of BMD Assessments

Month	Ideal Visit Day	BMD Assessment
Month 3	90	Up to 30 days after ideal visit day
Month 6	180	Up to 120 days after ideal visit day
Month 12	360	± 4 days of the ideal visit day

Due to the expanded time periods for collection of efficacy assessments, analysis visit windows are defined in Table 3 for the analyses of BMD and serum markers of bone metabolism (s-PINP and s-CTX) data to determine the assessment that will be used for a given study visit.

Table 3: Analysis Visit Windows for BMD, s-PINP and s-CTX Assessments

Scheduled Visit / Month	Ideal Visit Day	Visit Window in Study Days	
		BMD	s-PINP and s-CTX
Visit 3 / Baseline	≤ 1	≤ 1	≤ 1
Visit 4 / Month 1	30		2-60
Visit 5 / Month 3	90	2 -120	61-120
Visit 6 / Month 6	180	121 - 300	121-300
Visit 8 / Month 12	360	≥ 301	≥ 301

It is possible that more than 1 assessment for a parameter may be obtained within the same analysis visit window. In such circumstances, the assessment with the date closest to the ideal visit day will be used. In the event that 2 assessments are equidistant to the ideal visit day, the later of the assessments will be used.

The analysis visit window convention will not be applied to the data listings. The data listings will display all data as collected with the corresponding nominal visit.

3.13. Interim Analyses

No interim analysis for efficacy is planned for this study.

4. STUDY ANALYSES

4.1. Subject Disposition

Subject disposition will be tabulated by treatment group, including the number screened, the number randomized, the number treated, the number in each analysis population, the number who completed the study, the number who discontinued study drug with the primary reason for discontinuation of the treatment, and the number who discontinued from the study together with the primary reason for discontinuation as recorded in the eCRF. The number of subjects excluded from the PP population together with the exclusion reasons will also be summarized.

A Kaplan-Meier plot of the time from randomization to withdrawal from study will be presented by treatment group. Subjects who complete the study will be censored at their date of study completion as recorded on the EOS eCRF page.

The total number of subjects randomized will also be tabulated by region (North America, Europe), country, study site, and treatment group.

Subject disposition data will be carefully reviewed with respect to the impact of the COVID-19 pandemic on study conduct. A summary of subject disposition by subject's randomization date relative to COVID-19 pandemic onset (subjects randomized prior to 01 March 2020 vs subjects randomized on or after 01 March 2020) will be presented. The number (%) of subjects with assessments impacted by COVID-19 as recorded on the COVID-19 Impact Log eCRF page will be summarized by visit and by category. The categories are:

- Altered Protocol Assessment/Procedure
- Discontinuation
- IP – Missed/Modified
- Missing Data
- Missed Visit
- Out of Window Visit
- Remote /Virtual Visit
- Assessment Not Done
- Other

A listing of subjects with assessments impacted by COVID-19 as indicated on the COVID-19 Impact Log eCRF page will be provided. The listing will include subject ID, age, race, site, visit, CRF form used, impact category, and comment.

A by-subject data listing of study completion information, including the primary reasons for premature study withdrawal as recorded in the eCRF, will be presented. A by-subject data listing of subjects excluded from PP population and the reasons for exclusion will also be presented.

A summary of screening failures and listings of inclusion/exclusion criteria responses for screening failure subjects will also be presented.

4.2. Demographics and Baseline Characteristics

Demographic and baseline characteristics will be summarized and presented by treatment group and overall for the mITT, ITT, PP and Safety populations. These data include:

- Age, age group (<65, 65-<75, ≥75 years), race, ethnicity, region (North America, Europe), height, weight, body mass index (BMI), and BMI group (BMI<25, 25≤BMI<30, BMI≥30). Subjects who report more than one race category will be counted under the race category "Multiple".
- Smoking, drinking status and other substances use:
  - Smoking status (Never, Current, Former): Smoking status is based on the two questions asked during screening: 'Has the subject ever smoked cigarettes?' and 'Has the subject ever smoked tobacco?'. If both are answered 'Never', then smoking status is 'Never'. If either one is answered 'Current', then smoking status is 'Current'. If either question is answered 'Former' and the answer to the other question is not 'Current', then smoking status is 'Former'.)
  - Has the subject ever smoked cigarettes? (Never, Current, Former)
    - Number of cigarettes per day
    - Number of years smoked
  - Has the subject ever smoked tobacco? (Never, Current, Former),
    - 'How long ago did you stop?'
  - Has the subject ever drunk alcohol? (Never, Current, Former)
    - Number of years of drinking alcohol
    - How many drinks per week?
  - Has the subject ever used other substances? (Never, Current, Former).
- Fracture history of:
  - Prior clinical fracture (Yes, No): Yes - for any subjects with a record in medical history with "fracture" in the preferred term. No - otherwise.
  - Prevalent vertebral fracture (Yes, No): Yes - for any subjects with any vertebra with a Genant semi-quantitative score > 0 on the Spine X-ray Evaluation eCRF page. No - otherwise.
  - Prior fracture (Yes, No): Yes -for any subjects with a prevalent vertebral fracture and/or a prior clinical fracture. No - otherwise.
  - Prior clinical vertebral fracture (Yes, No): Yes -for any subjects with a record in medical history with "vertebral fracture" or "spinal fracture" in the preferred term. No - otherwise.
  - Prior vertebral fracture (Yes, No): Yes -for any subjects with a prior clinical vertebral fracture or a prevalent vertebral fracture. No - otherwise.
- Baseline values of BMD measurements and biomarker parameters including:
  - Lumbar spine BMD T-score and T-score group (≤ -2.5, > -2.5)
  - Total hip BMD T-score and T-score group (≤ -2.5, > -2.5)
  - Femoral neck BMD T-score and T-score group (≤ -2.5, > -2.5)
  - s-PINP and s-PINP category (tertile values calculated based on ITT population)
  - s-CTX.

The continuous variables including age (years), height (cm), weight (kg), BMI (kg/m^2), number of years smoked cigarettes, number of cigarettes per day, number of years of drinking alcohol, number of drinks per week, s-PINP, s-CTX and BMD T-score will be summarized using descriptive statistics (number of subjects, mean, SD, median, interquartile range, minimum and maximum). The categorical variables including ethnicity and race, subgroups (see Section 3.8) of baseline BMD measurements, history of fracture and fracture location, baseline s-PINP category, and history of use of cigarettes, tobacco, alcohol, and other substances will be presented by the number and percentage of subjects.

Medical history will be presented by MedDRA system organ class (SOC), preferred term (PT), and treatment group, summarizing the proportion of subjects in each treatment group who have a condition noted.

Physical examination results at baseline will be summarized by body system as recorded in the eCRF and treatment group.

No statistical tests will be performed for the demographic and baseline variables.

All analyses described in this section will also be repeated for the Antibody population. The presentation of these analyses will be based on subjects' ADA status and presence of neutralizing antibody (NAb) during the study with columns for ADA-, ADA+/NAb+, ADA+/NAb-, All ADA+, and Overall.

Additionally, demographics data and baseline characteristics will be tabulated by all subgroups listed in Section 3.8 using the ITT, mITT and PP populations. To address the effect of the COVID-19 pandemic, summaries of demographic data and baseline characteristics by subject's randomization date relative to COVID-19 pandemic onset (subjects randomized prior to 01 March 2020 vs subjects randomized on or after 01 March 2020) will be presented using the ITT and mITT populations.

Demographics including race, ethnicity, age, and country for screen failure subjects will also be summarized descriptively.

All subject data collected for demographics and baseline characteristics will be presented by treatment, study site, and subject number in data listings.

4.3. Exposure to Study Drug

4.3.1. Extent of Exposure and Treatment Compliance

Subjects were trained by study personnel on how to self-administer study drug with the abaloparatide-SC injection pen or the abaloparatide-sMTS patch. On the days of study site visits, study drug must be administered at the study site to accommodate pre-administration and post-administration procedures.

The overall duration of study drug exposure will be calculated using the data recorded on the eCRF as follows:

- Overall duration of study drug exposure (days) = (date of last dose – date of first dose) + 1.

If the date of last dose is missing, then the earlier of the following 2 dates will be used as the last dose date for analysis and summary purposes: the final visit/early termination visit date or the last drug dispense date + 30 days.

For the abaloparatide-SC group, the total number of doses delivered and study drug administration compliance will be calculated in two ways as follows:

- **Based on missed doses in subject diary:**
  - Number of doses delivered = Overall duration of study drug exposure – Total number of missed doses per subject diary
  - Percent compliance (%) to study drug administration = (Number of doses delivered / Overall duration of study drug exposure) × 100%
- **Based on measurement of drug remaining in pen:**
  - Number of doses delivered = [(45.5 mm × Number of used pens) - Sum of measurements of drug remaining in all used pens)] /1.08 mm. Here the 45.5mm is the initial measurement of unused study drug in an unused injection pen and 1.08mm is the distance the plunger travels for each 80 ug dose from the pen.
    - Note: If the measurement of drug remaining in any used pen is missing, then 100% compliance will be assumed for the intended treatment period for which the pen with missing measurement of drug remaining was dispensed. For each pen dispensed prior to Month 11 visit, the intended treatment period is a 30-day period starting from the day when first dose was taken from the pen. For the pen dispensed at Month 11 visit, the intended treatment period is a period less than or equal to 30 days starting from the day when first dose was taken from the pen to the EOT day. The number of doses taken from the pen with missing measurement of drug remaining is a value less than or equal to 30, which will be based on the number of doses taken from the pen(s) with non-missing measurement of drug remaining and the duration of the treatment period. The number of doses delivered would be the minimum of the following:
      - Total doses taken from pens with non-missing measurements of drug remaining (calculated as: [45.5 mm × Number of pens with non-missing measurements of drug remaining - Sum of measurements of drug remaining from all pens with non-missing measurements of drug remaining] /1.08 mm) + 30 × number of pens with missing measurement of drug remaining
      - Overall duration of study drug exposure (last dose date - first dose date + 1)
  - Percent compliance (%) to study drug administration = (Number of doses delivered / Overall duration of study drug exposure) × 100%

For the abaloparatide-sMTS group, the total number of doses delivered and study drug compliance will be calculated in two ways as follows:

- **Based on missed doses in subject diary:**
  - Number of doses delivered = Overall duration of study drug exposure - Total number of missed doses per subject diary
  - Percent compliance (%) to study drug administration = (Number of doses delivered / Overall duration of study drug exposure) × 100%
- **Based on number of patches returned:**
  - Number of doses delivered = 20 × Number of kits dispensed – Sum of patches returned in all dispensed kits. The number of patches dispensed per kit is 20.

- Note: If the number of patches returned in any dispensed kit is missing, then 100% compliance will be assumed for the treatment period for which the unreturned kit was dispensed. The number of patches used from the unreturned kit is a value less than or equal to 20, which will be based on the number of patches used from the returned kit(s) dispensed at the same visit as the unreturned kit. The number of doses delivered would be the minimum of the following:
  - Total patches taken from kits with non-missing number of patches returned (calculated as: 20 × Number of non-missing kits - Sum of number of patches returned from all non-missing kits) + 20 × number of kits with missing number of patches returned
  - Overall duration of study drug exposure (last dose date - first dose date + 1)

– Percent compliance (%) to study drug administration = (Number of doses delivered / Overall duration of study drug exposure) × 100%

The duration of study drug exposure, duration category in months (≤ 1 month, 1< to ≤ 3 months, 3 to ≤ 6 months, 6 to ≤ 9 months, and > 9 months), number of doses delivered, and percent compliance to study drug administration will be summarized by treatment group for the Safety, mITT and PP populations. The summaries will not be provided for the ITT population since the Safety and ITT populations represent the same subjects with respect to study drug exposure.

The study drug exposure and compliance will also be summarized by region (North America and Europe) for the Safety, mITT and PP populations.

To evaluate the relationship between study drug exposure and the antibody status, the duration of study drug exposure and percent compliance to study drug administration will also be summarized by antibody status (ADA-, ADA+/Nab+, ADA+/Nab-, All ADA+, and Overall) using the Antibody Population.

Two different patches will be dispensed to subjects in the abaloparatide-sMTS group. The ultra-low bioburden patches will be dispensed from Day 1 to all subjects. Sterile patches (the patches intended for commercial use) will be dispensed to subjects after confirmation that the sterile product is bioequivalent to the ultra-low bioburden product (Study 058-05-023). For the abaloparatide-sMTS group, the duration of study drug exposure, duration category in months, number of doses delivered and percent compliance with study drug administration using two calculations of percent compliance will also be summarized by patch type (ultra-low bioburden, sterile) for the Safety, mITT and PP populations.

If a subject is shown to be <80% compliant with study drug administration using both calculations of study drug administration compliance for abaloparatide-sMTS group, or is <80% compliant with study drug administration using both calculations of study drug administration compliance for the abaloparatide-SC group, this subject will be considered as having a critical protocol deviation and thus excluded from the PP Population.

Subjects' drug exposure and compliance will be carefully reviewed against the onset of the COVID-19 pandemic. Each subject's study drug exposure and compliance prior to 01 March 2020 and on or after 01 March 2020 will be calculated based on information collected from the subject diary:

- For subjects with last dose date prior to 01 March 2020 and subjects with first dose date on or after 01 March 2020, their duration of study drug exposure (days), number of doses delivered, and percent compliance to study drug administration relative to 01 March 2020 will be calculated as above based on the missed doses in subject diary
- For subjects who started taking the study drug prior to 01 March 2020 and continued on or after 01 March 2020, their duration of study drug exposure and compliance before 01 March 2020 and on or after 01 March 2020 will be calculated as follow:
  - Durations of study drug exposure relative to 01 March 2020:
    - Duration before 01 March 2020 = 29 February 2020 – date of first dose + 1
    - Duration on or after 01 March 2020 = Overall duration of study drug exposure – duration before 01 March 2020; where overall duration of study drug exposure = last dose date – first dose date + 1
  - The numbers of doses delivered relative to 01 March 2020 (The number of doses provided to a subject is different at each visit between the abaloparatide-SC group and abaloparatide-sMTS group. Therefore, different formulas are used to calculate the number of doses delivered before 01 March 2020 for the two treatment groups):
    - Number of doses delivered before 01 March 2020 (abaloparatide-SC group) = A – (B + [C × min {1, D / E}]), and will be rounded to the nearest integer; where:
      - A = Duration of study drug exposure before 01 March 2020
      - B = Total number of missed doses from pens dispensed before 01 March 2020 and with scheduled return date before or on 01 March 2020
      - C = Number of missed doses from pens dispensed before 01 March 2020 but with scheduled return date after 01 March 2020
      - D = 01 March 2020 – last drug dispensing date prior to 01 March 2020
      - E = First drug dispensing date on or after 01 March 2020 – last drug dispensing date prior to 01 March 2020
        - Note: if there is no drug dispense on or after 01March 2020, then E = Last dose date - last drug dispensing date prior to 01 March 2020
    - Number of doses delivered before 01 March 2020 (abaloparatide- sMTS group) = A – (B + C × D), and will be rounded to the nearest integer; where:
      - A = Duration of study drug exposure before 01 March 2020
      - B = Total number of missed doses from kits dispensed before 01 March 2020 and with scheduled return date before or on 01 March 2020
      - C = Number of missed doses from kits dispensed during the last drug dispensing visit before 01 March 2020 but with scheduled return date after 01 March 2020
      - D = (01 March 2020 – last drug dispensing date prior to 01 March 2020)/30
    - Number of doses delivered on or after 01 March 2020 = Number of doses delivered during whole study – number of doses delivered before 01 March 2020.
  - The percent compliance (%) to study drug administration relative to 01 March 2020:

- Percent compliance before 01 March 2020 = (Number of doses delivered before 01 March 2020 / Duration of study drug exposure before 01 March 2020) × 100%
- Percent compliance on or after 01 March 2020 = (Number of doses delivered on or after 01 March 2020 / Duration of study drug exposure on or after 01 March 2020) × 100%.

The study drug exposure and percent compliance to study drug administration before 01 March 2020 and on or after 01 March 2020 will be summarized by treatment group for the Safety, mITT and PP populations. Similar summaries will not be provided by patch type (ultra-low bioburden, sterile) since the switch from ultra-low bioburden patch to sterile patch for subjects in abaloparatide- sMTS group did not occur until after 01 March 2020.

The number (%) of subjects with any missed dose or dose interruption as recorded on the Dose Interruption Log eCRF will be summarized by treatment group, along with the reason(s) for missed/interrupted dose(s) for the mITT, PP, and Safety populations. If a subject reported a start date for the missed/interrupted dose but no corresponding start date for the resumption of study drug, the occurrence of the missed dose/dose interruption will not be included in the summary. The total number of days with missed/interrupted dose for each subject will be calculated and summarized by descriptive statistics and in the following categories: 1-7 days, 8-14 days, 15-28 days, >28 days by treatment group.

Study drug exposure and study drug administration compliance data will be listed by treatment (including patch type for abaloparatide-sMTS group), study site, and subject number.

4.3.1.1. Residual Drug Amount of Abaloparatide-sMTS

For the abaloparatide-sMTS group, residual drug amount from patches and skin swabs obtained at study site visits on Days 1 and 14 and Months 1, 3, 6, and 9 will be provided by Kindeva (previously 3M). The residual drug amount from patches and skin swabs will be used to determine the patch released dose, defined as:

Patch Released Dose = (initial patch content) – (patch residual drug amount) – (swab residual drug amount).

Initial patch content will be measured by Kindeva at the time of lot release and will be based on the average of 6 samples per lot. If the swab residual drug amount is missing, the patch released dose will be calculated based on initial patch content and patch residual drug amount. If the patch residual drug amount is missing, the patch released dose will be considered as missing.

Descriptive statistics for the safety population will be used to summarize the patch released dose, the percentage of released dose to nominal dose, the percentage of released dose to initial patch content, the residual drug amount from patch, the residual drug amount from swab, and the residual drug amount from patch and swab, by visit for Days 1 and 14, Months 1, 3, 6, and 9, and overall for the abaloparatide-sMTS group.

The summaries will also be presented by patch type (ultra-low bioburden, sterile).

4.3.2. Abaloparatide-sMTS Patch Adhesion Scores

Patch adhesion will be assessed daily by subjects randomized to the abaloparatide-sMTS treatment group using the following scoring system:

0 = ≥ 90% adhered (essentially no lift off the skin)

1 = ≥ 75% to < 90% adhered (some edges only lifting off the skin)

2 = ≥ 50% to < 75% adhered (less than half of the patch lifting off the skin)

3 = > 0% to < 50% adhered but not detached (more than half of the patch lifting off skin without falling off)

4 = 0% adhered – patch detached (patch completely off the skin)

A summary table using descriptive statistics will be provided combining all patch adhesion scores collected in the study. Frequency counts (and percentages) of all patch adhesion scores will be presented. The summaries will also be presented by patch type (ultra-low bioburden, sterile).

The mean of the patch adhesion scores collected from Day 1 to Month 12/EOT will be calculated for each subject and summarized using descriptive statistics. Each subject's best adhesion score and worst adhesion score will also be summarized. In addition, frequency counts (and percentages) for each subject's best adhesion score and for each subject's worst adhesion score from Day 1 to Month 12/EOT will be provided. For subjects treated with both patch types (ultra-low bioburden, sterile), the mean patch adhesion scores will also be calculated for each patch type for each subject for the respective period of treatment of the patch type. Each subject's best adhesion score and worst adhesion score by patch type will also be summarized for the respective period of treatment of the patch type. Frequency counts (and percentages) for each subject's best adhesion score and for each subject's worst adhesion score by patch type will be presented for the respective period of treatment of the patch type.

The percentage of patches that are ≥ 75% adhered (adhesion score ≤ 1) will be summarized by patch type and overall for abaloparatide-sMTS treatment group. Statistical assessment of patch adhesion will be evaluated via the following hypothesis test:

- H_0: $p \leq 80\%$ versus H_1:$p > 80\%$
- H_0 is rejected if the 95% lower confidence limit for p is greater than 80%

where p is the estimated percentage of patches that are ≥ 75% adhered. Two-sided 95% CIs for p will be provided for each patch type and overall using a model-based method as primary analysis and using Wilson's Score method as sensitivity analysis. Considering that for each subject, the patch adhesion score is collected daily using the subject diary during the treatment period, the model-based method will use a GEE (Generalized Estimating Equation) model assuming binomial distribution and including treatment group as fixed categorical effect with an unstructured correlation matrix to model the within-subject errors. For patch type comparison, the model will include patch type as fixed categorical effect. The p-value will be reported for patch type difference (ultra-low bioburden vs sterile).

All analyses described in this section will be presented for the Safety, mITT and PP populations. The analysis will not be provided for the ITT population since untreated subjects in the ITT population will not have exposure data.

4.3.3. Patch Wear Time

Patch wear time will be summarized for subjects randomized to abaloparatide-sMTS. Although the subject diary includes a field for "Time patch was removed/detached", the information from this field was recorded on the database as time of "5 Minutes Post-dose Evaluation". The patch wear time will be calculated as the difference between 'Time Patch Applied to Thigh' (as recorded on the subject diary and the database) and time of '5 Minutes Post-dose Evaluation'.

For each patch, the subject diary allows for only one date to be recorded. This results in calculated patch wear times that are negative (< 0 minutes) for patches applied before midnight but removed at or after midnight. To ensure appropriate calculation of patch wear time in such circumstances, 24 hours will be added to the patch removal time so that the patch wear time will be re-calculated using the updated patch removal time for records that meet all of the following conditions:

1. Patch application time is after 12:00 pm; and
2. Patch removal time is before 9:00 am; and
3. Original calculated patch wear time $\leq -12 \times 60$ minutes

where original calculated patch wear time = patch removal time minus patch application time, as recorded on the subject diary.

Negative patch wear times that cannot be corrected by the algorithm above will be excluded from the analysis.

Patch wear time collected at each study visit will be summarized using descriptive statistics by visit. Patch wear time collected over the entire treatment period will also be summarized using descriptive statistics. In addition, the longest wear time and shortest wear time for each subject over the entire treatment period will be summarized.

Patch wear time will also be summarized using the following categories:

- $<$ 4 minutes
- $\geq$ 4 minutes – 6 minutes
- $>$ 6 minutes – 7 minutes
- $>$ 7 minutes

The number (%) of subjects with patch wear time in each of the above categories at each study visit will be presented. The number (%) of patches belonging to each of the above categories over the entire treatment period will also be presented. In addition, similar categorical summaries will be provided for the longest wear time and the shortest wear time for each subject over the entire treatment period.

All analyses described in this section will be performed for both patch types (ultra-low bioburden, sterile) and overall using the Safety population.

4.3.4. Nutritional Supplements Exposure

Subject exposure to Vitamin D and calcium supplements will be tabulated separately in terms of total amount of IU (for Vitamin D) and mg (for calcium) received (equal to total dispensed minus total returned) and duration of exposure (Duration of Supplement Exposure (days) = date of last dose – date of first dose + 1) by treatment group using the Safety population.

Summary statistics for the average Vitamin D supplement exposure per day (= ([total dispensed minus total returned] / duration of exposure)) will also be presented separately by treatment group. Similarly, summary statistics will be provided for the average calcium supplement exposure per day.

4.3.5. Concomitant Medication

Concomitant medications will be coded using WHO-DD. The number (and percentage) of subjects taking medications will be tabulated by Anatomical Therapeutic Chemical (ATC) class, PT, and treatment group for the Safety population.

The tabulations will include concomitant medications taken from the date of the first dose of study drug (Day 1) until 30 days after the last dose of study drug. Medications that did not end prior to the first dose of study drug will be included in the summary.

Prior medications (those with start date prior to the first dose of study drug) will also be summarized by ATC, PT, and treatment group.

If both start date and stop date are missing, the medication will be assumed to occur both prior and concomitantly.

All recorded concomitant medications will be presented in a by-subject listing.

To explore the effect of the COVID-19 pandemic, summaries for the prior medications and concomitant medications will also be presented for the following:

- By region (North America and Europe).
- By medication start dates (medications start before 01 March 2020, vs medications start on or after 01 March 2020). For the calculation of percentages, the denominator includes all subjects in the safety population for the relevant treatment group.

4.3.6. Concomitant Procedures

All surgical, therapeutic, or diagnostic procedures performed during the study as a result of an (other) action taken due to an AE were captured on the Concomitant Procedures Log eCRF. These procedures will be coded using MedDRA Version 24.0. The number (and percentage) of subjects that had procedures will be tabulated by procedure class, PT, and treatment group for the Safety population.

The tabulations will include concomitant procedures, defined as those procedures with a start date on or after the date of the first dose of study drug (Day 1) until 30 days after the last dose of study drug. Procedures that did not end prior to the first dose of study drug will be included in the summary.

Prior procedures (those with start date prior to the first dose of study drug) will be summarized separately by procedure class, PT, and treatment group.

If both start date and stop date are missing, the procedures will be assumed to occur both prior and concomitantly.

All recorded concomitant procedures will be presented in a by-subject listing.

To explore the effect of the COVID-19 pandemic, summaries for the prior procedures and the concomitant procedures will also be presented for the following:

- By region (North America and Europe).
- By procedure start dates (procedures with start date before 01 March 2020, vs procedures with start date on or after 01 March 2020). For the calculation of percentages, the denominator includes all subjects in the safety population for the relevant treatment group.

4.4. Efficacy Analyses

The mITT population will be the primary population for the analysis of all efficacy endpoints. All efficacy endpoints will also be analyzed with the ITT and PP populations. Subjects will be analyzed according to their randomized treatment group.

All BMD-related endpoints will be derived from the Calyx-corrected BMD measurements.

BMD scans may be deemed unevaluable at the first reading and a repeat BMD scan may be performed. The re-scan BMD assessment will be recorded as an unscheduled visit for the corresponding nominal visit. To determine the assessment that will be used for a given study visit by the analysis window convention, all BMD assessments (including unscheduled BMD assessments) will be considered. In the efficacy analyses by nominal visit, unscheduled BMD assessments will be used if the BMD assessment at the nominal visit is not available but an unscheduled BMD assessment for the same nominal visit is available.

Analyses of the BMD-related efficacy endpoints and the serum markers of bone metabolism (s-PINP and s-CTX) will be conducted using the analysis visit windows defined in Table 3 of Section 3.12 as the primary analysis. Analyses of the BMD-related efficacy endpoints and the serum markers of bone metabolism (s-PINP and s-CTX) using the nominal visits will also be presented as supportive analyses. Study sites collected samples for bone turnover marker assessment in 4 aliquots per subject at each collection timepoint. When more than 1 aliquot was analyzed for a given timepoint for a subject, the average of the results for that timepoint for that subject will be used in the analysis.

Unless otherwise specified, all CIs, statistical tests, and p-values will be reported as nominal 2-sided from testing for treatment difference, with a significance level of 0.05.

4.4.1. Primary Efficacy Endpoint

4.4.1.1. Primary Analysis

The primary efficacy endpoint is the percent change from baseline in lumbar spine BMD at 12 months.

The primary hypothesis is that the effect of abaloparatide-sMTS 300 µg worn 5 minutes on the thigh on the percent change from baseline in lumbar spine BMD at 12 months is no worse than the effect of abaloparatide-SC 80 µg by a non-inferiority margin of 2.0%. Non-inferiority of

abaloparatide-sMTS to abaloparatide-SC will be concluded if the lower bound of the 2-sided 95% CI for the estimated treatment difference (abaloparatide-sMTS minus abaloparatide-SC) in the percent change from baseline in lumbar spine BMD at 12 months is above -2.0%, using an MMRM analysis.

The primary efficacy endpoint will be analyzed using an MMRM with fixed effects of treatment, DXA instrument manufacturer, visit, and treatment-by-visit interaction, and with baseline lumbar spine BMD as a covariate. Imputation for missing data is not necessary for the MMRM model because the model uses a restricted maximum likelihood (REML) based on the repeated-measures approach. An unstructured variance-covariance matrix will be used to model the within- and between-subject variabilities over the visits. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom and to adjust standard errors (Kenward and Roger, 1997). Analyses will be implemented using the SAS PROC MIXED procedure. The least square (LS) means for the treatment groups and LS means and corresponding 95% CIs for the treatment differences (abaloparatide-sMTS minus abaloparatide-SC) at Month 12 will be derived from the MMRM model.

4.4.1.2. Sensitivity Analysis

The primary analysis using the MMRM model for the primary efficacy endpoint will be repeated using the ITT and PP populations to support the finding from the mITT population. Additional sensitivity analyses described below will also be performed in the mITT, ITT, and PP populations.

The primary analysis using MMRM model is based on the missing at random (MAR) assumption. To assess the impact of deviations from the MAR assumption for the primary analysis, a sensitivity analysis using the tipping-point approach will be conducted on the primary efficacy endpoint to assess the robustness of the primary analysis result.

The tipping point approach assesses how severe departures from MAR must be to overturn conclusions from the primary analysis. If implausible departures from MAR to change the results from being favorable to the experimental treatment to being unfavorable, the results will be said to be robust to the departure from MAR assumption. This provides additional confidence in the results obtained based on the MMRM model with the MAR assumption.

The tipping point analysis will only be performed if NI is concluded based on the primary analysis using MMRM.

Implementation of the tipping point analysis requires performing multiple imputation of missing lumbar spine BMD data under the missing not at random (MNAR) assumption with a sequence of shift parameters, which adjust the imputed values for the abaloparatide-sMTS treatment group.

Implementing the tipping point analysis include the following steps with the first three steps being the standard multiple imputation (MI) steps:

1. The missing data in both arms will be imputed 100 times to generate 100 complete data sets. The following process will be followed within Step 1:
   a) A seed value of 940938 will be used in SAS PROC MI.

b) The Markov Chain Monte Carlo (MCMC) method will be used to impute the intermediate missing values, so that the missing values follow a monotone pattern. One hundred data sets will be generated.
c) For each of the 100 generated data sets, the monotone missing values will be imputed using the pattern-mixture model approach with a regression method. At this step, a shift parameter of zero will be applied. The SAS PROC MI methodology with the MNAR statement for imputation of monotone missing data patterns will be used in this step.

2. Percent changes from baseline will be calculated and each of the 100 complete data sets will be analyzed using the MMRM model as described in Section 4.4.1.1.
3. The results from the 100 complete data sets will be combined for the statistical inference using SAS PROC MIANALYZE.
4. The imputation at Step 1 above will start with 0 shift in both arms. If the lower bound of the 95% CI for the difference between abaloparatide-sMTS and abaloparatide-SC at Step 3 is greater than –2.0%, repeat Step 1c) to generate multiple imputed data sets, with a specified non-zero shift parameter that adjusts the imputed values in the abaloparatide-sMTS group. Otherwise, no further imputation will be performed, and it will be concluded that the results based on MMRM model with the MAR assumption is not robust.
5. Repeat Step 2 for the imputed data sets with a shift parameter applied.
6. Repeat Step 3 to obtain the 2-sided 95% CI for the difference between abaloparatide-sMTS and abaloparatide-SC to see if the lower bound of the 95% CI is still > -2.0%.
7. Repeat the Steps 4-6 with a more stringent shift parameter applied until the lower bound of the 95% CI ≤ - 2.0%.

Once the NI is established by the primary analysis, the lower bound of the 95% CI should be > -2.0%, and the tipping point (TP) that reverses the study conclusion will be determined. The analysis results with the 95% CI from Step 6 will be presented for the shift parameters between (TP – 5i) and min [0, (TP + 5i)], where i is the increment of the shift parameter, with a value of 0.01 or less.

Another sensitivity analysis based on the washout imputation method using the observed baseline values will be performed on the primary efficacy endpoint to address the possibility of data missing not at random. The following steps will be followed in this analysis:

1. Calculate the baseline mean and Month 12 standard deviation (SD) of BMD values in each treatment group.
2. Impute the missing BMD values at Month 12 in each arm using a normal random variable with mean and SD respectively equal to the baseline mean and Month 12 SD in that treatment group. A seed value of 940938 will be used at this step.
3. After the missing data imputation for month 12 is completed using the above steps, the percent change from baseline to month 12 will be calculated in each of the imputed datasets.
4. 100 datasets will be generated and will be analyzed using an ANCOVA model with treatment and DXA instrument manufacturer as fixed effects, and the baseline lumbar spine BMD as a covariate for the comparison of percent change of BMD from baseline to Month 12.

5. Treatment effects (difference in LS means between treatments) from these 100 analyses will then be combined using Rubin's Method via SAS PROC MIANALYZE procedure for overall statistical inference.

As a supportive analysis, the primary efficacy endpoint will be analyzed using an ANCOVA model using the last observation carried forward (LOCF) method. The ANCOVA model will include treatment and DXA instrument manufacturer as fixed effects, and baseline lumbar spine BMD as covariate. The LS mean and corresponding 95% CI for the difference between treatment groups (abaloparatide-sMTS minus abaloparatide-SC) at Month 12 will be presented. Missing data imputation based on the LOCF method will be performed as follows:

a. If baseline BMD is missing, no imputation for post-baseline value will be done and this subject's data will be excluded from the analysis.

b. If baseline BMD is observed, missing post-baseline BMD will be imputed by the last observed value prior to the missing assessment. That is, if a value is missing at a post-baseline time point, the post-baseline BMD will be imputed using the last value observed at any time (could be a value obtained at an unscheduled visit) between baseline (inclusive) and the specified post-baseline time point.

To investigate the effect of the COVID-19 pandemic, the following sensitivity analyses will also be performed:

a. MMRM analysis using the nominal visit

b. MMRM analysis excluding BMD measurements collected outside the allowable time period for data collection. In this analysis, BMD assessments are allocated to a Visit based on the inclusion rules shown in Table 4 below for Sensitivity Analysis b).

c. MMRM analysis excluding BMD measurements which are more than 30 days after the ideal visit date. In this analysis, BMD assessments are allocated to a Visit based on the inclusion rules shown in Table 4 below for Sensitivity Analysis c).

Table 4: Visit Windows for DXA Assessments for Sensitivity Analyses.

Visit	Ideal Visit Day	Sensitivity Analysis b): Allowable days from first dose of study drug	Sensitivity Analysis c): Allowable days from first dose of study drug
Month 3	90	2- 120 days	2 - 120 days
Month 6	180	121 - 300 days	121 - 210 days
Month 12	360	356 - 364 days	301 – 390 days

If there is more than 1 assessment for a parameter within the same visit window, the assessment with the date closest to the ideal visit day will be used. In the event that 2 assessments are equidistant to the ideal visit day, the later of the assessments will be used.

4.4.1.3. Subgroup Analysis

Subgroup analyses for each of the subgroups defined in Section 3.8 will be performed on the primary efficacy endpoint using MMRM analysis for mITT, ITT, and PP populations. The LS means and corresponding 95% CIs for the treatment differences (abaloparatide-sMTS minus

abaloparatide-SC) at Month 12 for each subgroup will be derived. A forest plot of the estimated treatment difference (and the corresponding 95% CI) will be presented for each subgroup. Additional exploratory analyses examining subgroup effect and treatment by subgroup interaction may be performed, if warranted.

4.4.1.4. Analysis by Anti-Abaloparatide Antibody Status

The analysis of the primary efficacy endpoint using MMRM model will also be repeated for the Antibody population within each treatment group, in which the subjects who are ADA+ at any post-baseline time point will be compared with subjects who are ADA- at all post-baseline time points. The MMRM model will include the fixed effects of ADA status, DXA instrument manufacturer, visit, and ADA status-by-visit interaction, with the baseline lumbar spine BMD as covariate.

4.4.2. Secondary Efficacy Endpoints

The secondary efficacy endpoints are:

- Percent change from baseline in total hip BMD at 12 months
- Percent change from baseline in femoral neck BMD at 12 months

The secondary efficacy endpoints will be analyzed using the MMRM model similar to the analysis for the primary efficacy endpoint, with the appropriate baseline BMD as covariate for the mITT, ITT, and PP populations. The LS means of the secondary efficacy endpoints for the treatment groups and the treatment differences will be estimated. The corresponding 95% CIs will be presented.

The sensitivity analyses using washout imputation using the observed baseline values described in Section 4.4.1.2 will also be applied to these 2 endpoints.

In addition, sensitivity analyses for addressing the effect of the COVID-19 pandemic and supportive analyses using ANCOVA model with LOCF imputation method, as with the analyses for the primary efficacy endpoint, will also be conducted with the appropriate baseline BMD as covariate. Subgroup analyses using the MMRM model will be performed on the secondary efficacy endpoints using mITT, ITT and PP populations for each of the subgroups defined in Section 3.8. The analyses will use MMRM analysis models as described above. A forest plot of the estimated LS mean (and the corresponding 95% CI) of the difference between the treatment groups will be presented by subgroup for each of the secondary efficacy endpoints.

The analysis of the secondary efficacy endpoints using MMRM model will also be repeated for the Antibody population within each treatment group, in which the subjects who are ADA+ at any post-baseline time point will be compared with subjects who are ADA- at all post-baseline time points, with fixed effects of ADA status (ADA+ vs. ADA-), DXA instrument manufacturer, visit, and ADA status-by-visit, and with the appropriate baseline BMD as a covariate.

4.4.3. Additional Efficacy Endpoints

Additional efficacy endpoints are:

- Percent change from baseline in
  - Lumbar spine BMD at 3 and 6 months

- Total hip BMD at 3 and 6 months
- Femoral neck BMD at 3 and 6 months

Additional efficacy endpoints will be analyzed using the MMRM model similar to the analysis for the primary efficacy endpoint for the mITT and PP populations, with the appropriate baseline BMD as covariate. The LS means of additional efficacy endpoints for the treatment groups and the treatment differences will be estimated. The corresponding 95% CIs will be provided.

Percent (%) changes from baseline in BMD will be summarized using descriptive statistics by visit (Month 3, 6, and 12) for each treatment group for the mITT and PP populations.

Plots will be presented for the LS mean (± SE) estimated from MMRM models of the percent change in BMD of the lumbar spine, total hip, and femoral neck over the 12-month treatment period for each of the treatment groups.

4.4.4. Other Endpoints

The primary population for the analysis of the bone formation marker, s-PINP, and the bone resorption marker, s-CTX, will be the mITT population. The same analyses will be repeated for the ITT and PP populations. Analysis of these bone turnover markers will be based on the ratio of the post-baseline value relative to the baseline value at each visit (using the analysis visit windows defined in Table 3). The transformation of the $\log_e$ ratio of post-baseline versus baseline value (derived by dividing the post-baseline value by the baseline value and then applying the natural log transformation) will be used to normalize the distributions of the s-PINP and s-CTX parameters. The analysis comparing abaloparatide-sMTS with abaloparatide-SC will use a MMRM model similar to the MMRM model for BMD analysis, with the appropriate natural-log transformed baseline bone turnover marker value as covariate. The analysis will be repeated based on the nominal visits.

Geometric mean (±SE) values, geometric mean values (±SE) relative to baseline, mean (±SD) change from baseline values, and mean (±SD) percent change from baseline values at each visit for s-PINP and s-CTX will be tabulated using descriptive statistics.

Plots will be presented for the geometric mean (±SE) of the ratio relative to baseline, as well as mean (±SE) percent changes from baseline, for s-PINP and s-CTX values over time by treatment group. Plots will also be presented for the median (± interquartile range) for the raw values of s-PINP and s-CTX over time by treatment group.

Any value of s-PINP or s-CTX that is provided as "< *numeric value*" or "> *numeric value*" will be imputed as the *numeric value* in all summary tables and analyses, but will be shown with the actual result in the listings.

4.4.5. Exploratory Efficacy Endpoints

Analysis of the exploratory efficacy endpoints described in this section will be performed using the mITT population as the primary analysis and repeated for the ITT and PP populations.

4.4.5.1. BMD Responders

Three categories of BMD response will be evaluated: % BMD increase > 0%, > 3%, and > 6%. The by-visit responses (based on the analysis visit windows described in Table 3) will be

assessed at the lumbar spine, total hip, and femoral neck. For the "all sites" analysis, subjects who have the pre-specified percentage BMD increase at all 3 anatomical sites at the same visit will be considered responders. A subject must have non-missing BMD results at all 3 anatomical sites at the same visit to be included in the "all sites" analysis. No imputation of missing data will be implemented. For the "all sites" analysis, the BMD response status at end of study will also be summarized. In this analysis, the "all sites" response status will be based on the subject's last BMD assessment during the study. If the last BMD assessment is missing the result for one or two anatomical site(s), the subject will be classified as a non-responder at End of Study. The denominator for the percentage of BMD responders at End of Study will be the number of subjects in the analysis population who had both a baseline and a post-baseline BMD.

BMD responder analyses will also be performed at each of the 3 anatomical sites separately, using the same 3 categories of BMD response used in the "all sites" analysis. A subject must have non-missing BMD results at a given visit to be included in the by-visit analysis. No imputation of missing BMD data will be implemented.

The primary time point for all BMD responder analyses is Month 12. The 95% CI of the response rate in each treatment group will be provided using Wilson's method. The Pearson chi-square test and 95% CI using a normal approximation will be used to explore the difference in the number (%) of responders between the 2 treatment groups at each visit (Months 3, 6 and 12) for each category of response. If the number of responders is less than 5 in any treatment group, the Fisher's exact test will be used. No multiplicity adjustment to the p-values for the BMD responder analysis will be implemented.

4.4.5.2. Time to Clinical Fracture

Analysis of the time to new clinical fracture by Month 12 will be performed for exploratory purposes. Duration in days from the date of randomization to the first incidence of clinical fracture will be derived.

The crude incidence rates, Kaplan-Meier estimates of the incidence rate at Month 12, hazard ratio (95% CI) based on a Cox proportional hazard model without covariates, and Kaplan-Meier curves, and p-values from the log-rank test for the comparison of the treatment groups will be generated for this time-to-event variable.

If a subject does not experience any clinical fracture over the 12 months of treatment plus the 30-day follow-up, this subject will be censored at the last known day in the study up to the follow-up visit.

4.4.5.3. Changes in Disease Status

Disease status categories are:

- Osteoporosis, defined as lumbar spine or total hip BMD T-score ≤ -2.5
- Osteopenia, defined as:
  - lumbar spine T-score > - 2.5 and total hip BMD T-score > - 2.5 and < -1.0, or
  - lumbar spine T-score > - 2.5 and < -1.0 and total hip BMD T-score > - 2.5
- Normal, defined as lumbar spine T-score ≥ -1.0 and total hip BMD T-score ≥ -1.0

The number and percentage of subjects for whom disease status has changed from osteoporosis at baseline to osteopenia, normal, or no change at end of study will be tabulated and analyzed using the CMH test stratified by baseline disease status. No imputation of missing BMD data will be implemented.

4.5. Population Pharmacokinetics (PK)

PK data will be analyzed and reported separately. Details of the PK data summaries and analysis methods will be provided in a separate analysis plan.

4.6. Safety Analyses

All safety and tolerability analyses will be conducted using the Safety population, except the summary of TEAE by ADA status, which will be conducted using the Antibody population. Treatment classification will be based on the actual treatment received. Unless otherwise specified, no formal statistical hypothesis testing will be performed for safety endpoints, and the presentations will be based on descriptive summaries. For summaries involving visit, the nominal visit will be used, unless otherwise specified.

The impact of the COVID-19 pandemic and the implementation of a variety of mitigations for study conduct will be reviewed and evaluated carefully for the safety assessment. The pandemic-related protocol deviations including missed or delayed visits, missed safety related tests, use of alternative assessment procedures for visit execution, use of alternative data collection or alternative modalities, use of rescue medications, or non-adherence to the planned course of treatment will be identified. Additional data summaries and analyses addressing the impact of the pandemic may be performed if necessary.

4.6.1. Adverse Events

All AEs will be coded to SOC and PT using MedDRA Version 24.0.

Analyses of AEs will be performed for those events that are considered to be TEAEs. A TEAE is defined as any AE that was absent (i.e., had not occurred) or had resolved prior to the start of study drug, and which occurred on or after the date of the first dose of study drug and within 30 days of the last dose of study drug; or any AE that started prior to the first dose of study drug, was ongoing after study treatment started, and increased in severity after the start of study drug and within 30 days of the last dose of study drug. AEs with missing onset dates will be handled according to the Conventions for Imputing Missing Dates in Appendix A.

TEAEs reported as unanticipated adverse device effects (UADE) or medical device reports (MDR) are those AEs with “Yes” to all of the following three questions on the AE eCRF: “Serious Event?”, “Was AE related to device?”, and “Would you consider this AE unexpected?”. For North American sites, unanticipated device-related SAEs will be reported as UADE for abaloparatide-sMTS, and will be reported as MDR for subjects on abaloparatide-SC. For European sites, unanticipated device-related SAEs will be reported as UADE for subjects on either abaloparatide-sMTS or abaloparatide-SC.

In any tabulation of TEAEs, a subject contributes only once to the count for a given PT and/or SOC, or for a time period category. For summaries by severity, a subject with multiple occurrences of a TEAE will be represented under the most severe occurrence. For summaries by relationship to study drug, a subject with multiple occurrences of a TEAE will be represented under the most related occurrence.

An overall summary of AEs will be presented by treatment group and overall, including the number and percentage of subjects reporting AEs, TEAEs, study treatment-related TEAEs (with probable or possible relationship to study drug), severe TEAEs, treatment-related severe TEAEs, serious TEAEs, treatment-related serious TEAEs, TEAEs leading to study drug interruption, TEAEs leading to study drug withdrawal, TEAEs leading to discontinuation from study, TEAEs related to the device, and TEAEs resulting in death.

The following summaries will be presented:

- TEAEs by SOC and PT
- TEAEs by SOC, PT, and region (North America, Europe)
- TEAEs by SOC, PT and Maximum Severity (mild, moderate, or severe)
- TEAEs by SOC, PT and relationship to the study drug (related or not related)
- TEAEs by PT
  Note: which will be sorted in descending order of total number of subjects with the preferred term in the overall column.
- Treatment-related TEAEs by SOC and PT
- Most Common (≥5% in either treatment group) TEAEs by SOC and PT
- Severe TEAEs by SOC and PT
- Treatment-related Severe TEAEs by SOC and PT
- Serious TEAEs by SOC and PT
- Treatment-related Serious TEAEs by SOC and PT
- TEAEs leading to study drug withdrawal by SOC and PT
- TEAEs leading to discontinuation from study by SOC and PT
- TEAEs leading to death by SOC and PT
- TEAEs leading to study drug interruption by SOC and PT
- Non-TEAEs by SOC and PT.

SOCs will be sorted in alphabetical order. Within an SOC, AEs will be sorted in descending order of the total number of subjects with the PT in the overall group.

To explore the relationship between TEAEs and the duration of study drug exposure, the incidence of TEAEs by time of onset will be presented for the following categories: ≤ 1 month, > 1 to ≤ 3 months, > 3 to ≤ 6 months, > 6 to ≤ 9 months, and > 9 months, by treatment group, SOC, and PT. Subjects with multiple occurrences of an AE within a time period will be counted once in each time period that an onset of the AE occurs.

To evaluate the relationship between TEAEs and antibody status, the incidence of TEAEs will also be summarized by antibody status (ADA-, ADA+/Nab+, ADA+/Nab-, All ADA+, and Overall), SOC, and PT using the Antibody population. A summary of TEAEs by PT and antibody status will also be presented by treatment group, in which TEAEs will be sorted in descending order of total number of subjects with the preferred term.

To explore the effect of the COVID-19 pandemic on reported AEs, summaries of TEAEs by SOC and PT, TEAEs leading to treatment discontinuation, and serious TEAEs will also be presented:

- by AE onset date (AE occurs before 01 March 2020, vs AE occurs on or after 01 March 2020)
- by region (North America and Europe)

All TEAE summaries, including AEs of special interest (AESI, Section 4.6.2) will be presented by patch type (ultra-low bioburden, sterile) for the abaloparatide-sMTS group. The exposure-adjusted incidence rate of TEAEs will also be presented by patch type for the abaloparatide-sMTS group. The exposure-adjusted incidence rate is calculated as number of subjects with TEAEs divided by the total on-treatment exposure in subject-years.

The number of subjects with any TEAEs, the number of subjects with any SAEs, and the number of subjects with TEAEs leading to study drug withdrawal will also be summarized by treatment group, SOC, PT and age group (< 65, 65 - < 75, and ≥ 75 years). Similar tables will be provided by treatment group, SOC, PT and BMI group (BMI < 25, 25 ≤ BMI < 30, BMI ≥ 30). Both summaries by age group and by BMI group will also be presented by patch type (ultra-low bioburden, sterile) for the abaloparatide-sMTS group.

By-subject listings will be provided for all AEs, including an indicator for TEAE/non-TEAE and for patch type (ultra-low bioburden, sterile). Listings will also be provided for deaths, serious TEAEs, TEAEs leading to study drug withdrawal, TEAEs leading to discontinuation from study, AEs of special interest, and severe AEs.

A data listing for COVID-19 related AEs (AE PTs related to COVID-19 using MedDRA Version 24.0) will be prepared by SOC and PT. A listing of subjects who experience an UADE or MDR will be provided from the clinical database.

4.6.2. Adverse Events of Special Interest

An adverse event of special interest (AESI) is an AE that is designated to be of special medical or scientific interest to the Sponsor. There are four major categories of AESIs as follows:

- Administration site reaction AEs -- Skin reaction AEs at the injection/application site
- Skin AESIs -- AEs of eschar, ulcer, or non-healing wounds at the injection/application site
- Hypersensitivity AESIs -- TEAEs in the Hypersensitivity Standardized MedDRA Query (SMQ narrow)
- Other AESIs:
  - Orthostatic hypotension
  - Tachycardia
  - Nausea
  - Hypercalcemia
  - Hypercalciuria
  - Hypophosphataemia

The MedDRA PTs, High Level Terms (HLT), and SMQs used to identify the AESIs are stored in the study folder in the Biometrics SAS environment.

An overall summary of the treatment emergent AESIs will be provided, in which the number of subjects with any adverse events, number of subjects with any TEAEs, number of subjects with any treatment-related TEAEs, number of subjects with any severe TEAEs, number of subjects with any treatment-related severe TEAEs, number of subjects with any serious TEAEs, number of subjects with any treatment-related serious TEAEs, number of subjects with any TEAES leading to study drug withdrawal, number of subjects with any TEAEs leading to study drug interruption, number of subjects with any TEAEs leading to study discontinuation, number of subjects with any TEAEs related to the device, and number of subjects with any TEAEs leading to death will be summarized by treatment group and overall.

The treatment emergent AESIs will be summarized by PT and SOC for each of the AESIs. Summary of the onset time and duration of the treatment emergent AESIs will be provided by

treatment group using descriptive statistics (number of subjects, mean, SD, median, interquartile range, minimum and maximum).

AESI onset time groups include:

- 1 day
- 2-3 days
- 4-7 days
- 8-15 days
- 16-29 days
- 1 to <2 months
- 2 to <6 months
- 6 to <10 months
- ≥10 months

AESI duration groups include:

- For Administration site reaction AEs, Skin AESI and Hypersensitivity AESIs:
  - < 1 hour
  - 1-<2 hours
  - 2-<4 hours
  - 4-<24 hours
  - 1-<2 day
  - 2-3 days
  - 4-7 days
  - 8-15 days
  - 16-29 days
  - >29 days
- For other AESIs:
  - 1-<2 day
  - 2-3 days
  - 4-7 days
  - 8-15 days
  - 16-29 days
  - >29 days

For AESI duration, if there is more than one occurrence of the AESI for a subject, then the average duration for all occurrences for the subject will be used. If AE stop date is missing, then the last study date will be used. If both AE start date and AE stop date are missing, the AE duration will be missing.

To evaluate the relationship between AESIs and the antibody status, the incidence of treatment emergent AESIs will also be summarized by antibody status (ADA-, ADA+/Nab+, ADA+/Nab-, All ADA+, and Overall), SOC, and PT for both treatment groups.

An overall summary of the treatment emergent AESIs, as described above, will also be presented by treatment group (and overall) for the following subgroups:

- Age group (< 65, 65 - < 75, and ≥ 75 years)
- BMI group (BMI < 25, 25 ≤ BMI < 30, and BMI ≥ 30)
- Region (North America and Europe)

Subject listings will be provided for each of the AESIs.

4.6.3. Anti-Drug Antibody

The ADA assessment in the serum samples for immunogenicity testing will be categorized by 2 outcomes:

- Negative for anti-drug antibodies
- Positive for anti-drug antibodies

The number (%) of subjects in the above categories of abaloparatide antibodies will be summarized at Day 1 (Visit 3), Day 14 (Visit 4), Month 1 (Visit 5), Month 3 (Visit 7), Month 6 (Visit 10), Month 9 (Visit 13), Month 12/EOT (Visit 16), and 30 days following the last dose of study drug (Month 13/Visit 17/EOS) by treatment group using the Antibody population.

For those subjects with anti-abaloparatide antibody present (i.e., are ADA positive), the titer level, the presence of neutralizing anti-abaloparatide antibodies (negative or positive), and the assessment outcomes for cross-reactivity to parathyroid hormone (PTH) (negative or positive) and parathyroid hormone-related peptide (PTHrP) antibodies (negative or positive) together with the neutralizing antibody status in subjects that are positive for PTH (negative or positive for the presence of PTH-neutralizing antibodies) or PTHrP (negative or positive for the presence of PTHrP-neutralizing antibodies) will be summarized by study visit. Figures for the incidence of antibody and neutralizing antibody for the anti-abaloparatide, anti-PTH, and anti-PTHrP will be presented by nominal visit.

All antibody information will be presented in a data listing.

The results of antibody analyses will be presented in a separate report.

4.6.4. Local Tolerance

Local tolerance will be assessed by the investigator and by the subject separately using a 4-point scale (0 = None, 1 = Mild, 2 = Moderate, 3 = Severe).

4.6.4.1. Investigator Assessment of Local Tolerance

Investigators will assess each of the 12 signs of local skin reaction. The 12 signs for local skin reactions are as follows:

- erythema
- edema
- vesiculation
- glazed appearance
- erosions
- crusting
- hyperpigmentation
- hypopigmentation
- scarring
- atrophy
- bruising
- bleeding

Investigators will evaluate the injection/application site at each study site visit prior to study drug administration, 5 minutes after study drug administration (i.e., immediately after patch removal for subjects receiving abaloparatide-sMTS), and 1 hour after study drug administration.

For each local skin reaction symptom, descriptive statistics (number of subjects, mean, SD, median, interquartile range, minimum and maximum) will be used to summarize the responses on the 4-point scale (0=None, 1=Mild, 2=Moderate,3=Severe) by treatment, visit and timepoint. In addition, for each symptom, the number (%) of subjects in each response category (none, mild, moderate, and severe) will be provided by treatment per visit and timepoint.

The maximum severity of each symptom of local skin reaction per subject will be summarized with the number (%) of subjects in each response category (none, mild, moderate, and severe) by treatment and visit. The maximum severity of each symptom of local skin reaction per subject will also be summarized with the number (%) of subjects in each response category by treatment and timepoint (and at post-dose timepoint).

The maximum severity of symptom scores will also be summarized descriptively (number of subjects, mean, SD, median, interquartile range, minimum and maximum) for each response on the 4-point scale (0=None, 1=Mild, 2=Moderate, 3=Severe) by treatment and visit, and by treatment and timepoint, respectively.

All summary tables described above will be repeated by patch type (ultra-low bioburden, sterile) for all subjects in the abaloparatide-sMTS group.

During the COVID-19 pandemic, the investigator's assessment of the symptoms of local skin reactions may be assessed through alternative methods other than direct on-site assessment. To explore the impact of the COVID-19 pandemic, summaries of the investigator's assessment of local tolerance will also be presented by:

- region (North America and Europe)
- excluding local tolerance assessments made via alternative methods other than direct on-site assessment

All local tolerance scores assessed by the investigator will be presented in a listing.

4.6.4.2. Subject Assessment of Local Tolerance

Subjects will be instructed to perform a daily self-assessment of symptoms of local skin reactions at the administration site beginning on Day 1 through the Month 12/EOT visit. The assessments include:

- pain
- itching
- burning
- tenderness
- swelling

Subjects will evaluate the injection/application site prior to study drug administration, 5 minutes after study drug administration (i.e., immediately after patch removal for subjects receiving abaloparatide-sMTS), and 1 hour after study drug administration.

For each subject, the following local tolerance scores will be derived for each skin reaction symptom:

- the mean of the local tolerance scores across the entire treatment period by time point
- the maximum severity of local tolerance scores across the entire treatment period by time point

For each local skin reaction symptom, descriptive statistics (number of subjects, mean, SD, median, interquartile range, minimum and maximum) will be used to summarize both derived scores above:

- by treatment, visit and timepoint
- by treatment and visit
- by treatment and timepoint

In addition, for each symptom, the number (%) of subjects in each response category (none, mild, moderate, and severe) will be provided for the maximum severity of local tolerance scores:

- by treatment, visit and timepoint
- by treatment and visit
- by treatment and timepoint

For subjects in the abaloparatide-sMTS group, descriptive analysis described above for both mean and maximum severity of local tolerance scores will be performed by timepoint and patch type (ultra-low bioburden, sterile).

In addition, the mean of the (subject) local tolerance scores will be calculated for each month (visit period). Data for the Month 1 visit period will consist of subject assessments from Day 1 to the day of the Month 1 visit; data for the Month 2 visit period will consist of assessments from the day after the Month 1 visit to the day of the Month 2 visit; data for the Month 3 visit period will consist of assessments from the day after the Month 2 visit to the day of the Month 3 visit; and similarly for the visit periods for Months 4, 5, 6, 7, 8, 9, 10, 11, and 12. If the subject missed a visit, the ideal visit date will be used as a cutoff date for calculating the subject's local tolerance scores for the corresponding visit periods.

All the analysis for subject's local tolerance scores will be repeated with only the days of site visits included.

All local tolerance scores by the subject will be presented in a listing.

All summary tables described above will be repeated by patch type (ultra-low bioburden, sterile) for all subjects in the abaloparatide-sMTS group.

4.6.5. Laboratory Data

Clinical laboratory values will be provided in conventional units and in standard international (SI) units by the central laboratory, unless otherwise specified.

In this study, local labs were allowed for screening. In addition, the original central lab was [redacted] Midway through the study, [redacted] was replaced with Medpace. Normal values of laboratory parameters in this study were provided by different local and central laboratories with inconsistent units and normal ranges. Thus, all numerical values will be converted to SI units and normalized to a standard set of reference/normal ranges as described by Chuang-Stein (1992 and 2001). The normalization process will be performed separately for each of the laboratory parameters as described in Appendix D, using the Medpace normal ranges as the reference ranges.

Summaries of laboratory data will be presented using normalized laboratory values. Summaries using descriptive statistics by treatment group and study visit will be presented based on the SI units, including absolute results and changes from baseline. This includes serum chemistry, hematology, coagulation, urinalysis (for quantitative urinalysis assays), hormones and Vitamin D. In the event of repeat assessments, the last non-missing value per study day/time will be used in the summary statistics. Results from unscheduled visits will not be included in the descriptive summary tables.

Tables for the shift of laboratory data from baseline to the worst post-baseline value and from baseline to Month 12 (Visit 16) will be presented by treatment group. Results for serum chemistry, hematology, coagulation, and urinalysis will be presented by category (above normal limit, within normal limit, below normal limit). For shift tables from baseline to worst post-baseline value, and for baseline to Month 12, percentages will only be calculated based on the number of subjects with valid data for baseline and post-baseline; subjects who are missing either assessment will not be included in the percentage calculation (numerator or denominator). Results from repeat/unscheduled visits will be included in the shift analyses, as applicable.

Summaries of all laboratory parameters will also be presented by region (North America and Europe).

Any laboratory parameter value that is provided as "< *numeric value*" or "> *numeric value*" will be imputed as the *numeric value* in all summary tables and analyses, but will be shown with the actual result in the listings.

All laboratory data, including repeat and unscheduled assessments, will be presented in data listings, with indication of higher or lower than the associated normal range of each laboratory test, where applicable. Lab listings will show values in original results (in conventional units), results in SI units, and normalized values.

4.6.5.1. Potentially Clinically Significant Laboratory Values

Criteria for determining values of laboratory tests that are considered potentially clinically significant (PCS) are listed in Appendix C of this SAP. The number (%) of subjects with post-baseline PCS values who did not have PCS values at baseline will be presented for each laboratory parameter with PCS criteria. Results from unscheduled visits will be included in identification of subjects with post-baseline PCS values, as applicable. For the calculation of the percentages of subjects, the denominator will be based on the number of subjects who had no PCS values at baseline and had at least one post-baseline assessment for the specific laboratory parameter of interest. The numerator will be based on the number of subjects from the denominator who had at least one PCS value post-baseline for the specific laboratory parameter of interest.

A listing of subjects with any PCS laboratory values will be provided.

4.6.5.2. Hypercalcemia

A hypercalcemic event (hypercalcemia) is defined as an elevated serum calcium (albumin-corrected) value ≥ 0.3 mg/dL (or ≥ 0.08 mmol/L) above the upper limit of the normal range (ULN).

Three cutoff levels of serum calcium (albumin-corrected) values will be defined:

a. ≥ ULN
b. ≥ ULN + 0.3 mg/dL (or ≥ ULN + 0.08 mmol/L)
c. ≥ ULN + 1.0 mg/dL (or > ULN + 0.25 mmol/L)

The following summaries and analyses will be provided by treatment group for each of the cutoff levels above:

- Number (%) of subjects with at least one occurrence of a serum calcium (albumin-corrected) value above the cutoff level at each visit, timepoint, and overall (entire treatment period):
  - The corresponding 95% CI based on Wilson's Score method will be provided per treatment group.
  - For the number of subjects with at least one occurrence of a serum calcium (albumin-corrected) value above the cutoff level during the entire treatment period:
    - The treatment groups will be compared using the chi-square test.
    - The risk reduction between the treatment groups (calculated as Abaloparatide-sMTS − Abaloparatide-SC) will be provided together with the 95% CI based on Newcombe's method.
    - The relative risk reduction between the treatment groups (calculated as (Abaloparatide-sMTS − Abaloparatide-SC)/Abaloparatide-SC) will be provided together with the 95% CI based on Wald's method.
- Number of occurrences of a serum calcium (albumin-corrected) value above the cutoff level per subject during the entire treatment period
- The ratio of the total number of occurrences of a serum calcium (albumin-corrected) value above the cutoff level to the total number of serum calcium (albumin-corrected) blood samples per subject during the entire treatment period.

Mean (±SE) serum calcium (albumin-corrected) values will be plotted over study visit by treatment group. All serum calcium (albumin-corrected) results will be included in the plot, which will indicate the results at pre-dose and 4-hours post-dose at each visit, as applicable.

To evaluate the relationship between age and hypercalcemia, the number (%) of subjects with at least one occurrence of hypercalcemia at each visit and overall (entire treatment period) will be summarized by treatment group and age group (< 65, 65 - < 75, and ≥ 75 years).

4.6.5.3. Hypercalciuria

The number (%) of subjects with at least one occurrence of hypercalciuria (urine calcium/creatinine ratio > 0.4 mg/mg (> 1.131 mmol/mmol)) at Baseline (Day 1), Month 3, and overall will be summarized by treatment group and visit. A 95% CI for the percentage based on Wilson's Score method will be calculated for each treatment group at each visit and overall. This analysis will be repeated for subjects with at least one occurrence of urine calcium/creatinine ratio > 0.3 mg/mg (> 0.848 mmol/mmol).

To evaluate the relationship between age and hypercalciuria, the number (%) of subjects with at least one occurrence of hypercalciuria at each visit and overall will be summarized by treatment group and age group (< 65, 65 - < 75, and ≥ 75 years).

4.6.5.4. Creatinine Clearance and Calcium Excretion

From the 24-hour urine collection at Day 1 and Month 3, urine sample values will be adjusted for the time period for the sample collection to derive the urine result that is standardized for a 24-hour period. The 24-hour urine results will be summarized descriptively by treatment group and visit. If the end of the collection time of the 24-hour urine is on the same day as the first dose date or within 24 hours from the time of first dose, then the 24-hour urine sample will be considered as a Day 1 sample and will be treated as the baseline sample. Otherwise, the 24-hour urine data will be considered as a post-Day 1 (i.e. post-baseline) sample.

The parameters to be summarized include:

- Urine calcium: creatinine ratio
- 24-hour urine calcium excretion (calcium concentration × 24-hour urine volume)
- 24-hour urine creatinine excretion (creatinine concentration × 24-hour urine volume)
- Creatinine clearance (CCr) calculated in two ways:
  a. Based on 24-hour urine collection
  b. Based on serum creatinine, calculated using the Cockcroft-Gault estimate, CC_r:

$$CC_r\ (mL/min) = \{[(140 - age) \times weight \times 1.23] \div SCr\} \times F$$

where:

- *age* = in years
- *weight* = in kilograms
- *F* (constant) = 0.85 for females
- *SCr* = serum creatinine in µmol/L.

The following categories of renal function will be defined:

	Cockcroft-Gault Estimated Serum Creatinine Clearance
Normal	≥ 90 mL/min (≥ 1.5 mL/s)
Mild	≥ 60 mL/min to < 90 mL/min (≥ 1.0 mL/s to < 1.5 mL/s)
Moderate	≥ 30 mL/min to < 60 mL/min (≥ 0.5 mL/s to < 1.0 mL/s)
Severe	≥ 15 mL/min to < 30 mL/min (≥ 0.25 mL/s to < 0.5 mL/s)

The number (%) of subjects in each renal function category at each visit and overall (across all visits) will be presented by treatment group. A similar summary table will be presented by treatment group and age group (< 65, 65 - < 75, and ≥ 75 years).

4.6.6. Vital Signs

Vital signs data, including orthostatic blood pressure, heart rate, body temperature, and respiratory rate, will be summarized by treatment group, visit, and time point using descriptive statistics. Summaries will be presented for the observed value, the change from baseline value, the change from baseline to the worst post-baseline value during the treatment period, and the change from baseline to EOS. If a subject has repeat assessments on the vital signs at a particular scheduled visit/time point, the last non-missing value will be used for the by-visit summary. Unscheduled assessments will be included in the summaries for the worst post-baseline value or PCS, they will not be included in the descriptive by-visit summary tables.

The number (%) of subjects experiencing post-dose orthostatic hypotension will be summarized by treatment group, visit (and across all visits), and time point (including pre-dose). Orthostatic hypotension will be defined as a decrease in systolic blood pressure (SBP) of ≥ 20 mmHg from supine to standing or in diastolic blood pressure (DBP) of ≥ 10 mmHg from supine to standing. All blood pressure data, including repeat and unscheduled assessments, will be included in determining the incidence of orthostatic hypotension.

The number (%) of subjects experiencing heart rate increase from pre-dose to post-dose and the number of occurrences per subject will be summarized with various threshold values (ranging from >5 bpm to >40 bpm, with threshold values increasing by 5 bpm). The median, minimum and maximum heart rate increase (in bpm) will be provided per threshold value. The number of occurrences per subject will be summarized using the median, minimum and maximum values. All heart rate data, including repeat and unscheduled assessments, will be included in determining the increase from pre-dose in this summary.

The above planned summaries will also be presented by region (North America and Europe).

To evaluate the relationship between age and orthostatic hypotension, the number (%) of subjects with at least one occurrence of post-dose orthostatic hypotension at each visit, timepoint, and overall (across all visits) will be summarized by treatment group and age group (< 65, 65 - < 75, and ≥ 75 years).

To evaluate the relationship between age and heart rate increase, the number (%) of subjects with at least one occurrence of heart rate increase from pre-dose to post-dose will be summarized with various threshold values (ranging from >5 bpm to >40 bpm, with threshold values increasing by 5 bpm) by treatment group and age group (< 65, 65 - < 75, and ≥ 75 years). Descriptive statistics for heart rate increase per threshold value and the number of occurrences of heart rate increase per subject will be per threshold value will be provided.

All vital signs data will be presented for each subject in a data listing. A listing of subjects with orthostatic hypotension will be presented.

4.6.6.1. Potentially Clinically Significant Vital Sign Values

Criteria for determining vital signs values that are considered PCS are listed in Table 5.

All vital signs data, including repeat and unscheduled assessments, will be included in determining the incidence of PCS vital sign values. The number (%) of subjects who have at least 1 PCS vital sign values for each vital sign parameter of interest after the first dose of study drug, but had no PCS values at baseline, will be presented by treatment. For the calculation of the percentages of subjects, the denominator will be based on the number of subjects who had no PCS values at baseline and had at least one post-baseline assessment for the specific vital sign parameter of interest.

The above planned summaries will also be presented by region (North America and Europe).

Table 5: Criteria for PCS of Vital Sign

Vital Sign	Criterion
Body Temperature (°C)	≥ 39.0
Supine Heart Rate (bpm)	> 100 > 130 < 45
Supine Systolic BP (mmHg)	> 155 < 80
Supine Diastolic BP (mmHg)	> 100
Standing Systolic BP (mmHg)	> 155 < 80
Standing Diastolic BP (mmHg)	> 100

A listing of subjects with any PCS vital sign values will be provided.

4.6.7. Electrocardiogram

The electrocardiogram (ECG) values collected are heart rate, PR interval, QRS duration, QT, and QTcF (QT corrected by Fridericia's formula). The observed ECG values and change from baseline values, and the change from baseline to the worst post-baseline value during the treatment period will be summarized descriptively for the quantitative ECG results by treatment group, visit, and time point. The overall ECG assessments will be summarized categorically (normal, abnormal without clinical significance, and abnormal with clinical significance) as a shift table by treatment group, study visit, and timepoint. With the exception of change from baseline to the worst post-baseline value during the treatment period, results from unscheduled visits will not be included in the summaries.

Summaries will also be provided for the change from predose values for ECG heart rate, where the predose value is the last value obtained prior to dosing at each visit where ECG is assessed. If a subject has multiple ECG results at a particular visit/time point, the repeat measurement reason will be checked to determine if the first or the last non-missing value will be used for the

summary. If the reason for the repeat assessment is abnormal value, then the first value will be used. If the reason is misplacement of the limb lead electrodes, the last value will be used. Results from unscheduled visits will not be included in these descriptive summary tables.

The number (%) of subjects experiencing heart rate increase from pre-dose to post-dose and the number of occurrences per subject will be summarized with various threshold values (ranging from >5 bpm to >40 bpm, with threshold values increasing by 5 bpm). The median, minimum, and maximum heart rate increase (in bpm) will be provided per threshold value. The number of occurrences per subject will be summarized using the median, minimum and maximum values. All ECG heart rate data, including repeat and unscheduled assessments, will be included in determining the increase from pre-dose in this summary. To evaluate the relationship between age and heart rate increase, the above summary of heart rate increase with various threshold values will also be repeated by treatment and age group (< 65, 65 - < 75, and ≥ 75 years).

The number and percentage of subjects with the following QTcF values will be presented by treatment group, visit and time point. Results from repeat/unscheduled visits will be included in shift analyses, as applicable:

- post-baseline QTcF > 450 msec
- post-baseline QTcF > 480 msec
- post-baseline QTcF > 500 msec
- change from baseline > 30 msec in QTcF
- change from baseline > 60 msec in QTcF
- post-baseline QTcF > 500 msec and change from baseline > 60 msec in QTcF

Shift tables from baseline to post-baseline QTcF values using the following categories will be presented for each post-baseline visit and time point. Results from repeat/unscheduled visits will be included in shift analyses, as applicable:

- ≤ 450 msec and > 450 msec
- ≤ 480 msec and > 480 msec
- ≤ 500 msec and > 500 msec.

All ECG data for each subject will be provided in a data listing. Listings will also be provided for those subjects who have any QTcF values greater than 450 msec, 480 msec, and 500 msec, and those with QTcF change from baseline values >30 msec and >60 msec.

Summaries for ECG parameters will also be presented by region (North America and Europe).

5. CHANGES FROM THE PROTOCOL-SPECIFIED ANALYSIS

The following changes from the protocol-specified analyses are included in this SAP.

- Antibody Population was added and defined in Section 2.1. All analyses related to the Antibody Population and ADA status will be reported separately.
- The following exploratory efficacy endpoints were added for analysis:
  - BMD responder - proportion of subjects experiencing BMD increase from baseline of >0%, >3%, and >6% at the lumbar spine, femoral neck, and total hip at 3, 6, and 12 months. For the "all sites" analysis by visit and at last assessment in study, subjects who have the pre-specified percentage BMD increase at all 3 anatomical sites at the same visit will be considered responders. A subject must have non-missing BMD results at all 3 anatomical sites at the same visit to be included in the "all sites" analysis. See Section 4.4.5.1.
  - Time to new clinical fracture by 12 months in Section 4.4.5.2.
  - Changes in disease status – The number and percentage of subjects for who disease status has changed from osteoporosis at baseline to osteopenia, normal, or no change at end of study will be tabulated and analyzed using the chi-square test. See definition of osteoporosis, osteopenia and normal in Section 1.2.5.1.

- Section 3.12 defines visit windows for the analyses of BMD, and serum markers of bone metabolism (s-PINP and s-CTX) data to determine the assessment that will be used for a given study visit.
- Section 4.3.2 was added for the analysis of patch adhesion scores for subjects randomized to abaloparatide-sMTS.
- Section 4.3.3 was added for the analysis of patch wear time for subjects randomized to abaloparatide-sMTS.
- Section 4.3.4 was added for the summary of subject exposure to Vitamin D and Calcium supplements.
- Section 4.3.5 was added for the summary of concomitant procedures, which were captured on the Concomitant Procedures Log eCRF as a result of an (other) action taken due to an AE.
- Sensitivity analyses for the primary endpoint were added in Section 4.4.1.2. These sensitivity analyses were also added for the secondary endpoints described in Section 4.4.2.
- Additional analyses to evaluate the impact of the COVID-19 pandemic on study conduct were added for safety and efficacy data, where applicable.

All other statements in this SAP are enhancements to the statistical considerations described in the protocol.

6. REFERENCES

Bolognese MA, Teglbjærg CS, Zanchetta JR, et al. Denosumab significantly increases DXA BMD at both trabecular and cortical sites: results from the FREEDOM study. *J Clin Densitom.* 2013 Apr-Jun;16(2):147-53.

Chuang-Stein, C., "Some Issues Concerning the Normalization of Laboratory Data Based on Reference Ranges," *Drug Information Journal*, 2001; 35:153-156.

Chuang-Stein, C., "Summarizing Laboratory Data with Different Reference Ranges in Multicenter Clinical Trials," *Drug Information Journal*, 1992; 26:77-84.

Document 3.2.P.7 Container Closure System –Medical Device [Abaloparatide, Vetter, Subcutaneous]

Food and Drug Administration (FDA) Guidance for Industry: Non-Inferiority Clinical Trials to Establish Effectiveness. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER) November 2016

Kenward MG, Roger JH, Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 1997; 53: 983-997.

ICH E9 (R1) Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials, May 2021, Revision 1.

Ratitch B, O'Kelly M, Tosiello R. "Missing Data in Clinical Trials: From Clinical Assumptions to Statistical Analysis Using Pattern Mixture Models." Pharm Stat. 2013;12:337- 347.

Appendix A. Conventions for Imputing Missing Dates

1. Adverse event (AE) start (onset) date imputation

AE start (onset) date is separated into 3 different data fields, including day, month and year. Completely missing AE start date (i.e., missing day, missing month and missing year) will not be imputed. Partial AE start date will be imputed according to treatment start date as follows in a sequential fashion.

- If AE year is missing, then no imputation
- If AE month and day are missing and
  - if AE year < Treatment year then AE month/day = '01JUL'
  - if AE year = Treatment year then AE month/day = treatment start date+1
  - if AE year > Treatment year then AE month/day = '01JAN'
- If AE day is missing and
  - if AE year < Treatment year then AE day = '15'
  - if AE year = Treatment year and
    - if AE month < Treatment month then AE day = '15'
    - if AE month = Treatment month then AE date = treatment start date+1
    - if AE month > Treatment month then AE day = '01'
  - If AE year > Treatment year then AE day = '01'
- If AE start date is still missing after the above imputation and AE end date is non-missing then AE start date = AE end date.
- If AE start date is still missing after the above imputation then count this AE as a treatment-emergent AE.

2. Concomitant medication (CM) start and end date imputation

The imputation of missing medication start and/or end dates is performed for the following reasons:

- To determine if a medication is a prior medication or a concomitant medication.
- To determine if treatment with the medication satisfies any of the medication-related exclusion criteria #16, #17, #18, #19, #20, #21, #22, #23, #24 and #26 for subject eligibility.

The imputation conventions described below for concomitant medication also apply to concomitant procedure (CP).

The medication start and end dates are separated into 3 different data fields, including day, month, and year. Completely missing medication start and/or end date (i.e., missing day, missing month, and missing year) will not be imputed. Partially missing medication start/end date will be imputed according to treatment start date. The rules on how to determine if a medication is a

prior medication only or both prior and concomitant medication and to determine if subjects violated any protocol specified medication-related exclusion criteria are specified below.

The prior and concomitant medications data listings will display all data as collected; imputed values will not be shown.

- **Medication Start Date Imputation**
  a) If medication start year is missing, then no imputation
  b) If medication start year is reported while medication start month and day are missing, then medication start month/day = '01JAN'
  c) If medication start year and month are reported while medication start day is missing, then medication start day = '01'

- **Medication End Date Imputation**
  a) If medication start date and medication end date are completely missing; or medication end date is completely missing while medication start date is either a complete date or partially missing and the observed/imputed medication start date < treatment start date, then no imputation:
     - If Ongoing = Yes, the medication is considered to be both prior and concomitant medication
     - If Ongoing = No or missing, the medication is considered to be a prior medication only
     - Subject will be considered as violating the medication-related exclusion criteria if the medication is one of the prohibited medications according to the medication-related exclusion criteria.

  b) If medication start date is completely missing while medication end date is partially missing; or medication end date is partially missing while medication start date is either a complete date or partially missing and the observed/imputed medication start date < treatment start date, then the missing medication end date will be imputed as follows:

	MMM Missing	MMM < TRT Month	MMM = TRT Month	MMM > TRT Month
YYYY < TRT Year	31DEC	Last day of MMM	Last day of MMM	Last day of MMM
YYYY = TRT Year	TRTDT *	Last day of MMM	TRTDT *	01MMM
YYYY > TRT Year	31JAN	Last day of MMM	Last day of MMM	Last day of MMM

* If Ongoing = Yes, then no imputation.

Where: YYYY = Medication end year; MMM = Medication end month; TRT Year = Treatment start year; TRT Month = Treatment start month; and TRTDT = Treatment start date.

- The observed and/or imputed medication end date will be used to determine if a medication is a prior medication only; or is both a prior and a concomitant medication.
- If the medication is one of the prohibited medications according to the medication-related exclusion criteria, the observed and/or imputed medication end date will also be used to determine if a subject meets any medication-related exclusion criteria.

c) If the medication start date is either a complete date or partially missing with the observed/imputed medication start date >= treatment start date, then no imputation for medication end date, and:

- The medication is considered to be a concomitant medication only.
- Subject doesn't meet any medication-related exclusion criteria.

3. **Other date imputations**

Any event with a partial date that requires imputation but has no specific imputation rule will be imputed as follows:

- If the day is missing, then impute the day as '01'
- If the month is missing, then impute the month as 'JUL'
- If the imputation of a start date causes the start date to be post the non-missing end date, then the start date will be imputed as the end date.
- If the imputation of a start date causes the start date to be prior to the start of treatment (but the event should happen post-baseline), then the start date will be imputed as "treatment start date + 1."
- Any date cannot be imputed past the date of death. When this happens, the date will be imputed as the date of death.

Appendix B. Algorithm to Identify Critical Protocol Deviations that Exclude Subjects from the Per-Protocol (PP) Population

The below algorithm outlines the criteria for identifying critical protocol deviations. Subjects with at least one of these critical protocol deviations will be excluded from the PP population.

A. Deviations of subject eligibility (Inclusion/Exclusion criteria) that will result in exclusion from PP population (refer to the protocol for the complete text of each inclusion/exclusion criterion below):

   1. Inclusion Criterion #2: Less than 5 years postmenopausal.
   2. Inclusion Criterion #3: A BMD T-score outside the required range or fracture criteria not met.
   3. Exclusion Criterion #1: History of more than 4 spine fractures, mild or moderate, or any severe fractures.
   4. Exclusion Criterion #2: Presence of abnormalities of the lumbar spine that would prohibit assessment of spinal bone mineral density (BMD).
   5. Exclusion Criterion #3: Unevaluable hip BMD or subjects who have undergone bilateral hip replacement.
   6. Exclusion Criteria #4, #5, #6: Diseases and treatments affecting bone strength.
   7. Inclusion Criterion #5: Laboratory tests affecting the bone strength.
   8. Exclusion Criteria #16, #17, #18, #19, #20, #21, #22, #23, #24 and #26: Prior treatment with bone-acting agents within the specified time periods.

B. Deviations of study drug administration that will result in exclusion from PP population:

   1. Duration of study drug exposure is less than 3 months (90 days)
   2. Compliance to study drug administration < 80%

C. Deviations of efficacy measurements that will result in exclusion from PP population:

   1. Lumbar spine BMD is missing at baseline
   2. No post-baseline lumbar spine BMD assessments recorded

D. Other critical protocol deviations

   1. Any protocol deviation with potential major impact on the efficacy endpoint will exclude a subject from the PP population. This will include subjects taking any prohibited medications during study.

Appendix C. Criteria for Potentially Clinically Significant (PCS) Laboratory Values

Lab Parameter	Conventional Unit		SI Unit	
	Low	**High**	**Low**	**High**
Hematology				
Absolute Eosinophils		>5000 cells/mm^3		>5.00 10^9/L
Absolute Lymphocytes	≤499 cells/mm^3		≤0.499 10^9/L	
Absolute Neutrophils	≤999 cells/mm^3		≤0.999 10^9/L	
Hemoglobin	≤9.4 gm/dL (female)	change from baseline ≥2.1 g/dL (female)	≤94 g/L (female)	change from baseline ≥21 g/L (female)
Platelets	≤99000 cells/mm^3		⩽99 10^9/L	
White Blood Cells	≤1499 cells/mm^3	≥20001 cells/mm^3	⩽1.499 10^9/L	≥20.0 10^9/L
Activated Partial Thromboplastin Time		≥1.41*ULN		≥1.41*ULN
Prothrombin Time		≥1.21*ULN		≥1.21*ULN
Serum Chemistry				
Alanine Aminotransferase		≥5.1*ULN		⩾5.1*ULN
Albumin	<2.5 g/dL		<25 g/L	
Alkaline Phosphatase IFCC		≥3.1*ULN		⩾3.1*ULN
Aspartate Aminotransferase		≥5.1*ULN		⩾5.1*ULN
Bilirubin		≥1.51*ULN (with ALT and/or AST>ULN) ≥2.0*ULN (with normal liver function tests)		⩾1.51*ULN (with ALT and/or AST>ULN) ≥2.0*ULN (with normal liver function tests)
Calcium	≤7.4 mg/dL	≥11.6 mg/dL	⩽1.85 mmol/L	⩾2.9 mmol/L
Cholesterol Total		>226 mg/dL		>5.8534 mmol/L
Creatine Kinase		≥3.1*ULN		⩾3.1*ULN
Creatinine		≥2.1 mg/dL		⩾ 185.64 umol/L
Glucose	≤54 mg/dL	>200 mg/dL (random)	≤ 2.997 mmol/L	> 11.1 mmol/L
Phosphorus	<2.0 mg/dL	>5.0 mg/dL	<0.645 mmol/L	>1.613 mmol/L
Potassium	≤3.2 mEq/L	≥5.5 mEq/L	⩽3.2 mmol/L	⩾5.5 mmol/L
Protein Total	<5 g/dL		<50 g/L	
Sodium	≤129 mEq/L	≥148 mEq/L	⩽129 mmol/L	⩾148 mmol/L
Urine				
Glucose		2+ (≥100 mg/dL)		2+ (≥5.55 mmol/L)
Protein		2+ (≥100 mg/dL)		2+ (≥ 1.0 gm/L)
Blood		>50 rbc/hpf		>50 rbc/hpf

Source: *FDA Guidance for Industry 2007, Toxicity Grading Scale for Health Adult and Adolescent Volunteers Enrolled in Preventive Vaccine Clinical Trials*

Appendix D. Normalization of Clinical Laboratory Values

Because normal values of laboratory parameters in this study were provided from different local or central laboratories with inconsistent units and normal ranges, all numerical values will be converted to the Standard International (SI) units and normalized to a standard set of reference/normal ranges as described by Chuang-Stein (1992 and 2001). The normalization process will be performed separately for each of the laboratory parameters.

For numerical laboratory values, the normalized values will be proportional to the ratio of the widths of reference ranges and aligned with the corresponding lower limits. For example, let (L_S, U_S) denote the standard reference range, which is the Medpace normal ranges in this study. Let (L_A, U_A) represents the reference range from a central or local Laboratory A. Let φ denote the laboratory value provided from Laboratory A, which is needed to be normalized. The normalized value, φ_S, can be derived from the formula,

$$\varphi_S = L_S + \beta \times (\varphi - L_A),$$

where $\beta = (U_S - L_S) / (U_A - L_A)$ equals the ratio of the widths of the reference ranges.

If the normalized value, φ_S, becomes negative, the value will be replaced by zero. If one, say the lower, of the limits of a reference range is missing then the missing limit will be imputed according to the ratio of the limits from one of the closest reference ranges provided in this study. For example, let L_M denote the missing lower limits from Laboratory M. Its corresponding higher limit, U_M, is found to be closest to the upper limit, U_C from Laboratory C among all the reference upper limits in the study. Then, L_M will be imputed by the formula,

$$L_M = L_C \times (U_M / U_C),$$

where L_C denotes the lower limit from Laboratory C. For cases where only the upper limits are missing (i.e., the lower ones are non-missing), same imputation process will be performed. For cases where both limits of a reference range are missing, the missing upper and lower limits will be imputed by the largest upper and lowest lower limit among all the observed reference ranges, respectively.

Signature Page for RAD-CLIN-002892 v1.0

Approval	[redacted] 17-Sep-2021 15:59:22 GMT+0000

Approval	[redacted] 17-Sep-2021 16:22:11 GMT+0000

Approval	[redacted] 20-Sep-2021 18:16:50 GMT+0000

Signature Page for RAD-CLIN-002892 v1.0