Statistical Analysis Plan

NCT Number: NCT03324880

Title: A Randomized, Double-blind, Placebo-controlled, Adaptive Study to Evaluate Symptom Improvement and Metabolic Control Among Adult Subjects With Symptomatic Hypoparathyroidism Treated With Recombinant Human Parathyroid Hormone [rhPTH(1-84)]

Study Number: SHP634-401

Document Version and Date: Statistical Analysis Plan Version 1.0, 16-MAY-2022

Certain information within this document has been redacted (ie, specific content is masked irreversibly from view) to protect either personally identifiable information or company confidential information.

STATISTICAL ANALYSIS PLAN

Study Number: SHP634-401

Study Title: A Randomized, Double-blind, Placebo-controlled, Adaptive Study to Evaluate Symptom Improvement and Metabolic Control Among Adult Subjects With Symptomatic Hypoparathyroidism Treated With Recombinant Human Parathyroid Hormone [rhPTH(1-84)]

Phase: 3b-4

Version: 1.0

Date: 16 May 2022

Prepared by: [redacted]

Based on:

Protocol Version: Amendment 3

Protocol Date: 11 Nov 2020

REVISION HISTORY

Version Number	Issue Date	Approval Date	Primary Rationale for Revision
Original			[Not Applicable]

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

ABBREVIATIONS

ACSC	Albumin-corrected serum calcium
AE	adverse event
CI	confidence interval
CRF	case report form
CS	Compound Symmetry
ECG	electrocardiogram
eGFR	estimated glomerular filtration rate
EOT	end-of-treatment
EQ-5D	EuroQol five dimensions questionnaire
ET	early termination
FACIT-Fatigue	Functional Assessment of Chronic Illness Therapy-Fatigue
FACT-Cog	Functional Assessment of Cancer Therapy-Cognitive Function
FGF	fibroblast growth factor
FSH	follicle-stimulating hormone
HPT	Hypoparathyroidism
HPT-SD	Hypoparathyroidism Symptom Diary
HRQoL	health-related quality of life
IRT	interactive response technology
ITT	Intention-to-treat
LOCF	Last Observation Carried Forward
MCID	Minimal Clinical Important Difference
MCS	mental component summary
MMRM	mixed-effect model for repeated measures
PCS	physical component summary
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PPS	Per-protocol Set
PRO	patient-reported outcome
PTH	parathyroid hormone
QTC	QT corrected for heart rate
RBC	red blood cell
SAE	serious adverse event
SAP	statistical analysis plan
SD	Standard Deviation
SE	Standard Errors
SF-36v2	36-Item Short Form Health Survey version 2
TEAE	treatment-emergent adverse event
TSH	thyroid-stimulating hormone
UI	Utility Index

ULN	upper limit of normal
UN	Unstructured
VAS	visual analog scale
WBC	white blood cell
WPAI:SHP	Work Productivity and Activity Impairment Questionnaire: Specific Health Problem

1.0 OBJECTIVES, ENDPOINTS AND ESTIMANDS

1.1 Objectives

1.1.1 Primary Objective

The primary objective is to test the hypothesis that rhPTH(1-84) treatment can result in superior improvements in the symptoms of hypoparathyroidism assessed by the Hypoparathyroidism Symptom Diary (HPT-SD) symptom subscale compared with standard therapy.

1.1.2 Key Secondary Objectives

The key secondary objectives are to test the hypotheses that rhPTH(1-84) treatment can result in superior improvements in:

- Fatigue as assessed by the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) compared with standard therapy.
- The physical component summary (PCS) derived from the 36-Item Short Form Health Survey version 2 (SF-36v2) acute version compared with standard therapy.

1.1.3 Secondary Objectives

The secondary objectives are as follows:

- To evaluate the efficacy of rhPTH(1-84) as assessed by the following patient-reported outcomes
  - HPT-SD impact subscale, total and individual items
  - HPT-SD symptom item anxiety and symptom item sadness or depression individual items
  - Individual symptom items of the HPT-SD symptom subscale
  - Evaluate response to the HPT-SD symptom subscale (as measured by a ≥30% reduction in symptom subscale score)
  - The most bothersome symptom of hypoparathyroidism
  - Functional Assessment of Cancer Therapy-Cognitive Function (FACT-Cog)
  - Individual domains and mental component summary (MCS) of the SF-36v2
  - Work Productivity and Activity Impairment Questionnaire: Hypoparathyroidism (WPAI:Hypoparathyroidism)
  - Patient Global Impression of Severity (PGI-S) for Hypoparathyroidism
  - Patient Global Impression of Change (PGI-C) for Hypoparathyroidism.
- To evaluate the effect of rhPTH(1-84) on neurocognitive performance as assessed by the following neurocognitive assessments (CogState Brief Battery; CogState Groton Maze Learning Test, CogState International Shopping List Task, CogState International Shopping List Task-Delayed Recall).

- To evaluate the effect of rhPTH(1-84) on metabolic control as assessed by serum albumin-corrected calcium, serum phosphate, and 24-hour urine calcium excretion, as well as active vitamin D and calcium supplement doses.
- To evaluate the effect of rhPTH(1-84) on bone turnover.
- To assess safety and tolerability of rhPTH(1-84).

1.2 Endpoints

1.2.1 Primary Endpoint(s)

The primary efficacy endpoint is:

- Change in the HPT-SD symptom subscale score from baseline to Week 26

1.2.2 Secondary Endpoint(s)

1.2.2.1 Key Secondary Endpoints(s)

The key secondary efficacy endpoints are:

- Change in FACIT-Fatigue score at Week 26
- Change in the PCS derived from SF-36v2 scores at Week 26

1.2.2.2 Other Secondary Endpoint(s)

The other secondary efficacy endpoints assessed at Week 26 are:

- Change in the HPT-SD impact subscale score
- Change in score of individual HPT-SD impact items
- Change in the HPT-SD symptom item anxiety and symptom item sadness or depression individual item score
- Change in score of individual HPT-SD symptom items
- Response as defined as a 30% reduction in HPT symptom subscale score from Baseline to Week 26
- Change in the most bothersome symptom score
- Change in FACT-Cog score (Perceived Cognitive Impairment, Impact on Quality of Life domains)
- Change in score of individual domains of SF-36v2
- Change in score of mental component summary (MCS) of SF-36v2
- Change in Work Productivity and Activity Impairment Questionnaire: Hypoparathyroidism (WPAI:Hypoparathyroidism) score
- Change in scores of patient's assessment of overall health status using:
    - Patient Global Impression of Severity (PGI-S) for Hypoparathyroidism

  - Patient Global Impression of Change (PGI-C) for Hypoparathyroidism

- Change in in-clinic neurocognitive assessment scores (assessed through Week 24) (CogState Brief Battery; CogState Groton Maze Learning Test, CogState International Shopping List Task, CogState International Shopping List Task-Delayed Recall)
- Change in at-home neurocognitive assessment scores (CogState Brief Battery)
- Change in 24-hour urine calcium excretion
- Change in serum phosphate level
- Changes in doses of active vitamin D and calcium supplements
- Albumin-corrected serum calcium control, defined as a concentration between 1.87 mmol/L (7.5 mg/dL) and ULN for the central laboratory normal range
- Composite endpoint, as defined as achieving all of the following:
  - Albumin-corrected serum calcium between 1.87 mmol/L (7.5 mg/dL) and the ULN for the central laboratory normal range
  - Dose of prescribed active vitamin D decreased by 50%
  - At least a 50% reduction from the baseline oral calcium supplement dose (this criterion will be considered met if the subject's baseline calcium dose is <1000 mg and it does not increase during the study).
- Changes in bone turnover markers

1.2.2.3 Exploratory Endpoint(s)

The exploratory efficacy endpoints are:

- Change in plasma FGF-23 levels
- Change in 24-hour urine phosphate, sodium, citrate, magnesium excretion
- Change in creatinine clearance
- Change in estimated glomerular filtration rate
- Change in EuroQol five dimensions questionnaire 5-level version (EQ-5D-5L): EQ-5D-5L index value, descriptive system, and EQ VAS measure
- Changes in measures of healthcare resource utilization: frequency of encounters (outpatient visits, laboratory tests, and procedures not scheduled in the protocol; emergency room visits; hospitalization), length of stay (hospitalization), and reasons/diagnoses for the encounters.
- Change in dose of native vitamin D supplements (recorded as concomitant medications).

1.2.3 Safety Endpoints

The following safety variables constitute the safety endpoints measured in this study:

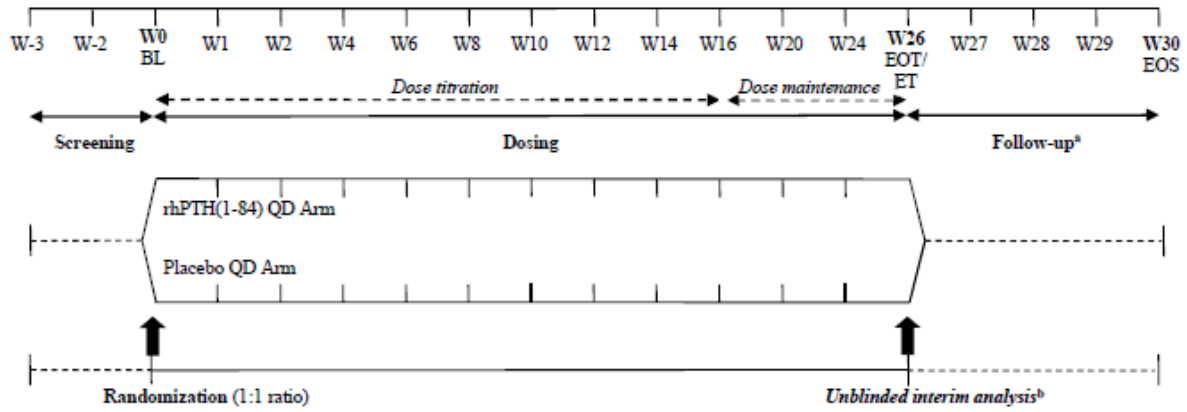
- Adverse events (AEs)
- Hypocalcemic AEs
- Vital signs
- Laboratory safety data (serum chemistry, hematology, urinalysis)
- Electrocardiograms (ECGs)
- Anti-PTH antibodies

2.0 STUDY DESIGN

This is a randomized, double-blind, placebo-controlled, 2-arm, adaptive study in a minimum of 92 and no more than 150 adult subjects with symptomatic chronic hypoparathyroidism on standard therapy. The study periods will be as follows: a 3-week screening period; a 16-week dose-titration period; a 10-week maintenance-dosing period with minimal change in the investigational product dose; and a 4-week safety follow-up period that includes an end-of study (EOS) contact for all subjects 30 days following the last dose of investigational product. The safety follow-up period includes weekly visits for subjects discontinuing rhPTH(1-84) treatment after the end-of-treatment (EOT) visit (Week 26) or early termination visit. Subjects transferring immediately to commercial rhPTH(1-84) who experienced a treatment gap of >7 day after EOT (Week 26) visit will proceed with weekly follow-up visits for every week that there is an interruption in rhPTH(1-84) dosing until receiving commercial rhPTH(1-84) or until a maximum of 30 days has elapsed.

A schematic representation of the study design and titration period is displayed in Figure 1. Specific procedures at each study visit are summarized in Table 1.

Figure 1 Study Design Flow Chart



Dosing consists of a 16-week dose-titration period (Weeks 1-16) and a 10-week maintenance-dosing period (Weeks 17-26) with minimal change in investigational product dose.

BL=baseline; EOT=end of treatment; ET=early termination; FU=follow-up; QD=once daily; rhPTH(1-84)=recombinant human parathyroid hormone; W=week.

[a] All subjects will complete an end-of-study (Week 30) contact 30 days following the last dose of investigational product. Weekly follow-up visits (Weeks 27-30) applies to subjects discontinuing treatment with rhPTH(1-84) after Week 26 (EOT/ET) visit. Subjects transferring immediately to commercial rhPTH(1-84) after the EOT (Week 26) visit will continue to be monitored under the care of a physician according to rhPTH(1-84) labeling instructions. Subjects transferring to commercial rhPTH(1-84) who experienced a treatment gap of >7 day after the EOT (Week 26) visit will proceed with weekly follow-up visits for every week that there is an interruption in rhPTH(1-84) dosing until receiving commercial rhPTH(1-84) or until a maximum of 30 days has elapsed. See protocol Section 7.1.4 for additional details related to the follow-up period. The EOS contact (Week 30 visit) is a safety follow-up telephone call initiated by the site staff to query for adverse events (AEs) and concomitant treatments. Subjects who will be returning to the site for their last weekly follow-up visit can complete the EOS contact (Week 30 visit) at the site.

[b] The interim analysis will be performed when approximately 68 subjects are enrolled and complete 26 weeks of treatment or prematurely withdrawn from the study, whichever comes first. This unblinded interim analysis will be performed by a third party to reassess the assumptions used for the sample size and assess for futility. The total number of subjects enrolled in the study may be maintained at 92, increased from 92 up to 150 subjects or the study may be stopped due to futility.

Table 1 Schedule of Assessments

Visit	Screening		Baseline	Dose-titration Period (Weeks 1-16)[a]									Maintenance-dosing Period (Weeks 17-26)[b]			Follow-up[c]			
															End of Treatment/ET				EOS/
Study Week	-3	-2	0	1	2	4	6	8	10	12	14	16	20	24	26	27	28	29	30
Study Procedures		±2-day window											±3-day window			±2-day window			
Visit Type (S=site/P=phone)	S	S	S	S	S	S	P	S	P	S	P	S	S	S	S	S	S	S	S/P
Informed consent	X																		
Inclusion/exclusion	X	X	X																
Medical history and demography	X																		
Prior/concomitant treatments	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Physical examination	X														X				
Vital signs[d]	X	X	X	X	X	X		X		X		X	X	X	X	X	X	X	X
ECG (12-lead)	X									X					X				
Dietary evaluation and recommendations		X					X			X									
X-ray to ensure epiphyseal closure[e]	X																		
Hematology	X									X					X				
Serum chemistry[f]	X		X							X					X				
Serum TSH	X														X				
Serum pregnancy test[g]	X																		
Urine pregnancy test[g]	X		X	X	X	X		X		X		X	X	X	X				X
FSH levels[h]	X																		
Serum PTH	X		X																
In-clinic predose nadir level: serum calcium, albumin, phosphate, magnesium[i]	X		X	X	X	X		X		X		X	X		X	X	X	X	X
Between visit predose nadir levels (serum calcium, albumin, phosphate, magnesium)[i,j]			As needed following adjustment of investigational product, active vitamin D and/or calcium supplements																
Between visit post-dose peak levels (serum calcium, albumin, phosphate, magnesium)[i,j]			As needed following adjustment of investigational product, active vitamin D and/or calcium supplements																
Serum 25-hydroxyvitamin D	X		X		X	X		X		X		X	X		X				
Serum 1,25-dihydroxyvitamin D			X							X					X				
Plasma FGF-23			X												X				

Table 1 Schedule of Assessments (continued)

	Screening		Baseline	Dose-titration Period (Weeks 1-16)[a]									Maintenance-dosing Period (Weeks 17-26)[b]			Follow-up[c]			
Visit															End of Treatment/ET				EOS/
Study Week	-3	-2	0	1	2	4	6	8	10	12	14	16	20	24	26	27	28	29	30
Study Procedures		±2-day window													±3-day window	±2-day window			
Visit Type (S=site/P=phone)	S	S	S	S	S	S	P	S	P	S	P	S	S	S	S	S	S	S	S/P
Bone turnover markers			X							X					X				
Anti-PTH antibodies[k]			X							X					X				X
Urinalysis	X														X				
24-hour urine collection[l]		X			X			X		X			X		X				
Randomization			X																
Review Hypoparathyroidism Symptom Diary[m]		X	X	X	X	X	X	X	X	X	X	X	X	X	X				
Review Most Bothersome Hypoparathyroidism-related Symptom		X[n]																	
Review subject records of IP, active vitamin D, and calcium supplement doses[o]		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Neurocognitive assessment training		X																	
Neurocognitive assessments – in clinic		X												X					
Remind subject of at-home neurocognitive assessments completion[p]		X							X					X					
Phone call to remind subjects to complete PROs prior to the baseline (Week 0) visit			X[r] (Day -1)												X[r] (Day prior to the visit)				
Review at-home neurocognitive assessments[q]			X							X					X				
Review focused hypoparathyroidism symptom assessments (FACT-Cog, FACIT-Fatigue)[r]			X			X				X				X	X				
Review health-related quality of life (SF-36v2, EQ-5D-5L)[r]			X			X				X					X				
Review work productivity (WPAI:Hypoparathyroidism)[r]			X			X				X				X	X				
Review Patient Global Impression of Severity for Hypoparathyroidism[r]	X	X	X			X				X				X	X				

Table 1 Schedule of Assessments (continued)

	Screening		Baseline	Dose-titration Period (Weeks 1-16)[a]									Maintenance-dosing Period (Weeks 17-26)[b]			Follow-up[c]			
Visit															End of Treatment/ET				EOS/
Study Week	-3	-2	0	1	2	4	6	8	10	12	14	16	20	24	26	27	28	29	30
Study Procedures		±2-day window											±3-day window			±2-day window			
Visit Type (S=site/P=phone)	S	S	S	S	S	S	P	S	P	S	P	S	S	S	S	S	S	S	S/P
Review Patient Global Impression of Change for Hypoparathyroidism[r]		X	X							X					X				
Healthcare resource utilization[s]			X		X	X		X		X		X	X	X	X				X
AE/SAE monitoring	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
IP self-administration training and review			X																
Dispense/account for IP and supplements[t]			X		X	X		X		X		X	X	X	X				
Collection of unused IP, supplements, injection pen, mixing apparatus, and PRO and neurocognitive devices															X				

Note: See protocol Section 7.2 for guidance about standardized meals and timing of active vitamin D, calcium supplements, and or investigational product doses on study visit days.

AE=adverse event; ECG=electrocardiogram; EOS=end of study; EQ-5D-5L= EuroQol five dimensions questionnaire 5-level version; ET=early termination; FACIT Fatigue=Functional Assessment of Chronic Illness Therapy-Fatigue; FACT-Cog=Functional Assessment of Cancer Therapy-Cognitive Function; FGF-23=fibroblast growth factor-23; FSH=follicle-stimulating hormone; HPT=hypoparathyroidism; IP=investigational product; PTH=parathyroid hormone; SAE=serious adverse event; SF-36v2=36-Item Short Form Health Survey version 2; TSH=thyroid-stimulating hormone; WPAI: Hypoparathyroidism=Work Productivity and Activity Impairment Questionnaire: Hypoparathyroidism

[a] Changes in investigational product dose can occur approximately every 4 weeks up to and including Week 16. Active vitamin D and calcium supplements should be taken as determined by the investigator in order to achieve albumin-corrected serum calcium levels in the target range (see protocol Appendix 2). Native vitamin D should be taken as determined by the investigator to target a serum 25-hydroxyvitamin D concentration between 75 nmol/L (30 ng/mL) and the upper limit of normal (ULN). Magnesium supplements should be given, as appropriate, to achieve a serum magnesium concentration within the laboratory normal range

[b] Following Week 16, investigational product doses are intended to remain relatively stable during the last 10 weeks of the treatment period; however, if the subject has not reached the optimal dose of investigational product, adjustments can be made following discussion with the medical monitor (see protocol Appendix 2). At any time during the study as needed for safety reasons, investigational product doses may be decreased in 25 µg decrements to a minimum of 25 µg QD.

[c] All subjects will have an end-of-study (Week 30) contact 30 days following the last dose of investigational product. Weekly follow-up visits will be performed for subjects discontinuing treatment with rhPTH(1-84) after Week 26 (end-of-treatment or early termination) visit. These subjects will be prescribed appropriate oral calcium and/or active vitamin D supplements to compensate for the cessation of the rhPTH(1-84). Subjects transferring immediately to commercial rhPTH(1-84) after Week 26 visit (end-of-treatment visit) will continue to be monitored under the care of a physician according to rhPTH(1-84) labeling instructions. Subjects transferring to commercial rhPTH(1-84) who experienced a treatment gap of >7 days after the EOT (Week 26) visit will proceed with weekly follow-up visits for every week that there is an interruption in rhPTH(1-84) dosing until receiving commercial rhPTH(1-84) or until a maximum of 30 days has elapsed (see protocol Section 7.1.4 for additional details). If a subject transfers to commercial rhPTH(1-84) treatment before the EOS (Week 30) visit, assessments scheduled at the EOS (Week 30) visit should be performed at their last follow-up visit before transferring to commercial rhPTH(1-84); these subjects will have an EOS contact 30 days following their last dose of investigational product. The EOS contact (Week 30 visit) is a safety follow-up telephone call initiated by the site staff to query for adverse events (AEs) and concomitant treatments. Subjects who will be returning to the site for their last weekly follow-up visit can complete the EOS contact (Week 30 visit) at the site.

[d] Vital signs include blood pressure, pulse, body temperature, and respiratory rate.

[e] For subjects 18-25 years of age, bone age X-ray (single posteroanterior X-ray of the left wrist and hand).

[f] Serum chemistry panel excludes albumin, calcium, phosphate, and magnesium, which are noted separately.

Table 1 Schedule of Assessments (continued)

[g] Pregnancy tests for female subjects of childbearing potential. The urine pregnancy test results obtained at screening will be confirmed by a serum pregnancy test.
[h] FSH levels required for newly menopausal women.
[i] In addition to protocol specified laboratory testing processed in the central laboratory, blood for the evaluation of serum calcium, albumin, phosphate, and magnesium levels obtained on scheduled visit days can be analyzed locally at the sites or other local laboratories that are authorized by the investigator. These values may be used for titration of investigational product and supplements. When these samples are drawn locally, these samples can be analyzed per local laboratory standards (eg, ionized or total serum or plasma calcium) and should be interpreted using the local laboratory normal values. If a local laboratory is used on scheduled study site visit days (ie, Weeks -3, -2, 0, 1, 2, 4, 8, 12, 16, 20, 24, 26, 27, 28, 29, 30), a duplicate sample must be collected and provided to the central laboratory. See protocol Appendix 2 for testing between scheduled visits.
[j] It is recommended that between visit predose nadir and post-dose peak levels should be assessed following adjustment the of investigational product, active vitamin D, and/or calcium supplements to aid in titration. If between visit predose nadir and post-dose peak levels are not assessed, the reason for not completing the measurements should be documented. Blood for the assessment of these levels should be collected approximately 2-5 days following any adjustment(s). The post-dose peak level blood draw for the evaluation of serum calcium, albumin, phosphate, and magnesium should be drawn approximately 8-12 hours following administration of the investigational product dose. The predose nadir level blood draw for the evaluation of serum calcium, phosphate, albumin, and magnesium should be drawn in the morning prior to the administration of the dose of investigational product that day, within 24 hours of the last dose of investigational product. Blood for predose nadir and post-dose peak levels can be drawn on the same day or different days as long as the blood draws are completed within the 2-5 day window and within the window specified for each level. Local laboratories can be used for these assessments; duplicate samples do not need to be sent to the central laboratory. Additional between visit blood draws can be completed at any time at the discretion of the investigator.
[k] Blood samples for the detection of anti-PTH antibodies must be collected prior to dosing with the investigational product (at least 14 hours after the prior dose of investigational product).
[l] Subjects should strive to the greatest extent possible to maintain consistent daily dietary (non-supplement) intake of calcium, phosphate, and sodium during 24-hour urine collections. The 24-hour urine collection is to be completed the morning of the specified visits. Subjects taking vitamin C or a multivitamin containing vitamin C must stop this medication for the 5 days prior to and during each 24-hour urine collection period. Subjects should take their usual calcium, vitamin D (both native and active), and investigational product doses. On urine collection days, subjects should drink enough water to ensure that at least 1 liter of urine is collected over 24 hours.
[m] Subjects will be asked to complete a Hypoparathyroidism Symptom Diary at a consistent time every day from the day of screening (Week -3) until the end-of-treatment/ET visit. During treatment with investigational product, it will be expected that subjects record responses approximately 5-12 hours after morning investigational product doses. Data collected during Week -2 to baseline (Week 0) will be used to confirm eligibility to proceed in the study.
[n] Subjects will be asked to complete the question once at home within the 2 days before the Week -2 visit; if the question was not completed at home prior to the Week-2 visit, it can be completed once at home within the 2 days after the scheduled visit. The question cannot be completed during the Week -2 visit.
[o] Subjects will be asked to record their doses of active vitamin D and calcium supplements each day from screening (Week -3) until the safety follow-up visit at end-of-study visit (Week 30) or end-of-treatment/ET visit, as applicable. Subjects will also be asked to record their doses of investigational product each day from the baseline visit (Week 0) until the end-of-treatment/ET visit.
[p] Subjects will be reminded to complete the neurocognitive assessments (CogState Brief Battery) at home twice daily on each of the 14 days before the visit, ending 1 day before the visit. For the baseline visit, the assessments are to be completed twice daily on each of the days between the Week -2 visit and baseline visit (Week 0). Subjects will complete the testing once each morning after their morning routine (eg, dressing, eating, and taking medication) and once in the evening before dinner.
[q] Review subject's compliance with completion of at-home neurocognitive assessment.

Table 1 Schedule of Assessments (continued)

[r] At the Week -3 visit, subjects should complete the PGI-S for hypoparathyroidism at the clinic after the subject has received and has been trained on the PRO device. PRO instruments (including PGI-S for hypoparathyroidism and PGI-C for hypoparathyroidism, Fact-Cog, FACIT-Fatigue, SF-36v2, EQ-5D-5L, and WPAI:hypoparathyroidism) should be completed once at home within the 2 days prior to baseline (Week 0) and EOT (Week 26) visits. On the day before the baseline (Week 0) and the EOT (Week 26) visits, subjects should receive a phone call to remind them to complete their PRO instruments. Subjects who do not complete their PRO instruments prior to the baseline (Week 0) visit will not be able to be randomized and can be rescreened starting at the Week -3 visit per the discretion of the investigator. Subjects should complete the PRO instruments for the Week -2 visit (including the most bothersome hypoparathyroidism-related symptom, PGI-S for hypoparathyroidism, and PGI-C for hypoparathyroidism) and for applicable visits between Week 1 and Week 24 (including PGI-S for hypoparathyroidism, PGI-C for hypoparathyroidism, Fact-Cog, FACIT-Fatigue, SF-36v2, EQ-5D-5L, and WPAI:hypoparathyroidism) once at home within the 2 days prior to the scheduled visit. If the PRO instruments are not completed prior to the scheduled visit, they can be completed once at home within the 2 days after the scheduled visit. PRO instruments cannot be completed during scheduled visits. During treatment with investigational product, the instruments should be completed approximately 5-12 hours after investigational product dosing. Review subject's compliance with completion of indicated instruments at specified visits.

[s] Baseline Visit version to be used at the baseline visit (Week 0) only. Follow-up version to be used at all subsequent visits.

[t] Subjects will use their pre-existing active vitamin D and calcium supplements during the screening period. Investigational product and active vitamin D and calcium supplements to be taken after randomization will be dispensed at the baseline (Week 0) visit.

3.0 STATISTICS HYPOTHESES AND DECISION RULES

3.1 Statistical Hypotheses

The study is designed to test the hypothesis that treatment with rhPTH(1-84), a recombinant human parathyroid hormone (PTH), can result in superior improvements in the symptoms of hypoparathyroidism compared with standard therapy.

3.2 Statistical Decision Rules

Not applicable.

3.3 Multiplicity Adjustment

In order to maintain study-wide Type I error control, a hierarchical testing procedure will be used in the comparisons between rhPTH(1-84) and placebo on the primary and key secondary efficacy endpoints. Specifically, the testing will be conducted in the following order: primary endpoint (HPT-SD), first key secondary endpoint (FACIT-Fatigue), and second key secondary endpoint (PCS from the SF-36v2). A latter test can only be reported as significant if all prior 1-sided tests are also found significant at the 0.025 level of significance. If prior 1-sided tests are not found to be statistically significant, p-values generated for latter analyses will be descriptive and described as nominally significant if less than or equal to 0.025.

Multiplicity is not adjusted for other efficacy endpoints in this study. Summary statistics including nominal p-values will be reported.

4.0 SAMPLE-SIZE DETERMINATION

A total 92 subjects (46 subjects per arm) are planned to be randomized in a 1:1 ratio to the rhPTH(1-84) and placebo treatment groups in order to have a total of approximately 78 subjects (39 subjects per arm) completing the study.

The sample size was estimated for the primary efficacy analysis on change from baseline to Week 26 in the HPT-SD symptom subscale. We assumed a mean difference of 0.4 units with a standard deviation of 0.6 units. A total of 39 completed subjects per treatment group will provide 82.5% power to detect a treatment difference of 0.67 between the rhPTH(1-84) and placebo treatment groups, based on a 1-sided, 2-sample t-test at the 0.025 level of significance. This takes into account a single interim analysis for futility at approximately 74% information with a non-binding stopping rule at conditional power under the alternative hypothesis less than approximately 10%. The power analysis and sample size estimation are calculated using East Version 6.

Based on cross-sectional data from the HPT-SD psychometric evaluation study, a difference of 0.4 in HPT-SD would roughly correspond to a between-person one-unit difference in the PGI-S for hypoparathyroidism (ie, the difference between 'mild' versus 'moderate' or 'moderate' versus 'severe'). Based on data from HPT-SD psychometric evaluation study, an effect size of 0.67 is achieved approximately by a clinically meaningful between treatment group difference of 0.4 in mean change in the HPT-SD. Approximately 92 subjects are planned to be randomized to compensate for the 15% of randomized subjects who will not complete the treatment phase.

This proposed sample size would also provide >70% power to detect a difference of 3.5 in change from baseline in the physical component summary (PCS) derived from the SF-36v2, a key secondary endpoint, between the rhPTH(1-84) and placebo treatment groups and a standard deviation of 6.0.

As described in SAP Section 6.10, an unblinded interim analysis for the primary endpoint will be performed to reassess the appropriateness of assumptions used for the sample size calculation of the primary efficacy endpoint when the study was designed and to assess for futility.

5.0 ANALYSIS SETS

5.1 Screened Set

The Screened Set will consist of all subjects who have signed informed consent.

5.2 Enrolled Subjects

Enrolled subjects consist of all subjects who have signed informed consent and some study procedures have begun.

5.3 Intention-to-treat (ITT) Set

Intention-to-treat (ITT) Set consists of all randomized subjects. Subjects in the ITT Set will be analyzed in the treatment group assigned at randomization, regardless of the actual treatment received (analyzed as randomized).

5.4 Safety Analysis Set

The Safety Set will consist of all subjects in the ITT Set who have taken at least 1 dose of investigational product (study drug or placebo). Subjects in the Safety Set will be analyzed in the treatment group corresponding to actual treatment received (analyzed as treated).

5.5 Per-protocol Set (PPS)

Per-protocol Set (PPS) will consist of all subjects in the ITT Set who complete the study and who do not have predefined protocol deviations that impact the primary efficacy assessment. Subjects in the PPS will be analyzed by actual treatment received.

6.0 STATISTICAL ANALYSIS

6.1 General Considerations

Baseline is defined as the last available pre-dose value except for diary variables (e.g., HPT-SD), for which baseline will be derived as the average of the values over the 14 days (Day -14 to Day -1) preceding baseline. Additionally, the baseline of ECG is the average of triplicated pre-dose ECG assessment results. Calculation of average is specified in relevant analysis sections.

All summaries will be presented by treatment group, unless otherwise specified.

All listings will be displayed by treatment group and subject number, based on the Safety Analysis Set for safety listings and ITT analysis set for other listings, unless otherwise specified.

All efficacy analyses will be based on the ITT set unless otherwise specified. In addition, the primary endpoint analysis and analyses for key secondary endpoints will be repeated over the PPS set. All confidence intervals (CIs) will be 2-sided 95% CIs, unless stated otherwise. All p-values reported will be 1-tailed and rounded to 3 decimal places prior to assessment of statistical significance. Multiplicity adjustments will be performed on the primary and key secondary efficacy endpoints only per SAP Section 3.3.

Means and medians will be presented to 1 more decimal place than the recorded data. The standard deviations (SDs) will be presented to 2 more decimal places than the recorded data. CIs will be presented using the same number of decimal places as the parameter estimate.

Where applicable, variables will be summarized descriptively by study visit. For the categorical variables, the counts and proportions of each possible value will be tabulated by treatment group. The denominator for the proportion will be based on the number of subjects who provided non-missing responses to the categorical variable. For continuous variables, the number of subjects with non-missing values, mean, median, SD, minimum, and maximum values will be tabulated.

6.1.1 Handling of Treatment Misallocations

Efficacy endpoints will be analyzed by randomized treatment (i.e., based on the randomization number assigned by Interactive Response Technology [IRT]). Safety endpoints will be analyzed by actual treatment received. Any subject who received both treatments (rhPTH[1-84] and placebo) will be assigned to the actual treatment group with longer duration of treatment.

6.1.2 Handling of Dropouts or Missing Data

Imputation will be performed to explore the impact of missing data on the primary endpoint per SAP Section 6.5.1.2. For all other endpoints, only the observed data from the subjects will be used in the statistical analyses, i.e., there is no plan to estimate (impute) missing data.

6.1.3 Visit Window

For diary data including primary efficacy endpoint HPT-SD, visit windows will be defined for every 14-day period prior to the actual visit date of each subject.

For other endpoints analyzed by visit, assessments will be assigned to visits based on the completed CRF page.

6.2 Disposition of Subjects

The number of subjects included in each subject set will be summarized by treatment group. The number and percentage of subjects who completed and prematurely discontinued from study will be presented for each treatment group and overall for ITT Set. Reasons for premature discontinuation from study as recorded on the early termination page of the CRF will be summarized (number and percentage) by treatment group. Subject disposition, subjects completing and prematurely discontinued during the study, and study analysis sets will be listed by subject for ITT Set.

The number of subjects enrolled, randomized and completed will be tabulated by site for enrolled subjects. Duration of enrollment will be calculated as (last date of contact for any subject at that site - the first date of informed consent for any subject at that site + 1).

Subjects with at least 1 protocol deviation for the deviation category below will be summarized for ITT set by treatment group. Protocol deviations will be identified programmatically as well as based on protocol deviations collected by the study monitors. After the last subject exits the trial and prior to database lock, protocol deviations will be reviewed to determine which subjects fall into categories listed below. Significant deviations which may impact the efficacy results will be excluded from PPS analysis. These deviations will be decided prior to database lock and unblinding of treatment codes.

- Did not meet inclusion/exclusion criteria
- Randomized prior to the date of the baseline visit
- Administered the incorrect study treatment
- Non-compliance with protocol specified visit windows
- Non-compliance with study drug
- Received disallowed concomitant medications post-baseline
- Developed withdrawal criteria but were not withdrawn
- Other (as applicable)

All protocol deviations will be listed.

6.3 Demographic and Other Baseline Characteristics

6.3.1 Demographics

Descriptive summaries of demographic [age (years), age category (18-<45, 45-64, >=65), sex, race, ethnicity, educational level] will be presented by treatment group and overall for the Safety Analysis Set, PPS, and ITT Set. Demographic will be listed for the ITT Set.

6.3.2 Medical History and Concurrent Medical Conditions

Medical and surgical history will be coded per documented MedDRA Dictionary version 20.1 or above in the Data Validation Manual. The history will be summarized by system organ class (SOC) and preferred term (PT) within system organ class for the Safety analysis set, with SOC sorted alphabetically and PT sorted by decreased frequency within SOC. Medical history will be listed for the ITT Set.

6.3.3 Baseline Characteristics

The following baseline characteristics will be summarized using Safety Analysis Set, PPS, and ITT Set: height (cm), weight (kg), BMI (kg/m^2), baseline calcium supplement dose (as continuous parameter and by categories of 0-2000 mg/day, >2000 mg/day), active vitamin D dose (as continuous parameter and by categories of low dose 0-0.25 µg/day, medium dose >0.25-0.5 µg/day, high dose >0.5 µg/day), history of thyroid hormone replacement (as continuous parameter in years and by categories of yes vs no) and most bothersome symptom. Baseline characteristics will be listed for the ITT Set.

6.4 Medication History and Concomitant Medications

Classification of prior and concomitant medications by therapeutic class will be accomplished per documented World Health Organization (WHO) Drug Dictionary version Sep 2017 or above.

6.4.1 Prior and Concomitant Medications

Prior medication is defined as any medication with the start date prior to the date of the first dose of double-blind investigational product.

Concomitant medication is defined as any medication with a start date prior to the date of the first dose of double-blind investigational product and continuing after the first dose of investigational product or with a start date between the dates of the first and last doses of investigational product, inclusive. Any medication with a start date after the date of the last dose of investigational product will not be considered a concomitant medication.

Partial date imputation for medications is described in SAP Section 9.5.2.

Both prior and concomitant medication usage will be summarized by the number and proportion of subjects receiving each medication within each level of therapeutic class and preferred term for ITT Set. Medications can be counted both as prior and concomitant medication. Multiple medication usage by a subject in the same each level of therapeutic class and preferred term will

be counted only once. Therapeutic classes will be sorted alphabetically while medication preferred terms will be sorted by decreased frequency within therapeutic class.

All prior and concomitant medications and medical/surgical procedures will be listed.

6.5 Efficacy Analysis

6.5.1 Primary Endpoint(s) Analysis

The primary efficacy endpoint is the change from baseline in the HPT-SD symptom subscale (version 1) score at Week 26. The null hypothesis for the primary efficacy endpoint is that there is no difference in symptom improvement at Week 26 from baseline between the rhPTH(1-84) and placebo treatment groups. The alternative hypothesis is that the rhPTH(1-84) treatment group shows superior symptom improvement compared with the placebo treatment group.

The HPT-SD is a 13-item subject-reported outcomes instrument that consists of a symptom subscale (items 1-7), anxiety (item 8), sadness and depression (item 9) and impact subscale (items 10-13). HPT-SD will be completed daily at a consistent time from the day after the screening visit (Week -3) until the end-of-treatment/early termination (Week 26) visit (endpoint derivation is detailed in Section 6.5.1.1).

The primary efficacy analysis will be performed on subjects from the ITT Set using a mixed-effect model for repeated measures (MMRM) analysis at post-baseline visits (Weeks 2, 4, 6, 8, 10, 12, 14, 16, 20, 24, 26), with change from baseline in the HPT-SD symptom subscale score as the outcome variable; treatment group, visit, and their interaction as fixed effect factors and subjects as random effect factors, with adjustment for baseline HPT-SD symptom subscale score. Visit will be treated as a class variable, assuming an unstructured (UN) covariance matrix to model within-subject variability. If there is a convergence problem due to the use of UN covariance matrix, a compound symmetry (CS) covariance structure will be used to model the within-subject errors. If there is a still convergence problem with CS, other covariance structure may be used. All unscheduled visits will be excluded, and all scheduled results will be included in the model. The primary analysis will be repeated on PPS.

The null hypothesis will be rejected if the statistical analysis results in a 1-sided p-value for treatment at Week 26 less than or equal to 0.025. With a mixed effects model based on restricted maximum likelihood estimation used for the primary analysis and assuming that, conditional on fixed and random effects, data are missing at random, no imputation of missing data will be performed in the primary analysis. The primary result obtained from the model will be the estimated main treatment effect at Week 26. The estimated main treatment effect of mean difference between rhPTH(1-84) and placebo, standard errors (SEs) and a 95% CI will be provided. In addition, least squares means (LSMEANs) estimated from the model for each treatment group at each post-baseline visit, and the estimated treatment effect along with standard errors (SEs) and 2-sided 95% CI, as well as p-values will also be provided. Descriptive statistics for HPT-SD symptom subscale score at each assessment including baseline, Weeks 2, 4, 6, 8, 10, 12, 14, 16, 20, 24, 26, and change from baseline at each post-baseline assessment will be reported.

Histogram, forest plot with difference of least square mean and CIs, and line-plots (mean +/- SE) by visit will also be generated for HPT-SD item scores. In addition, symptom subscale score will be graphically presented by line plots.

6.5.1.1 Derivation of Endpoint(s)

The 13 item HPT-SD consists of the HPT-SD symptoms subscale (items 1-7), anxiety (item 8), sadness and depression (item 9) and HPT-SD impact subscale (items 10-13) will be computed as follows:

For symptoms subscale (items 1-7):

Step 1: Compute item-level score from daily responses for each of the 7 symptom items. Require at least 8 days with non-missing responses during the 14-day period immediately prior to the visit date to compute item-level score. If fewer than 4 data points are available for each 7-day period, the item-level score for the period is set to missing.

$$\text{Item-level score} = \frac{\text{sum of non}-\text{missing daily responses during 14}-\text{day period}}{\text{number of days with non}-\text{missing responses during 14}-\text{day period}}$$

Step 2: Compute HPT-SD symptom scores using item-level scores calculated in Step 1. Per scoring algorithm, compute the average value of item-level score over items 1-7. At least 4 out of 7 symptom items must have non-missing item-level score to compute the HPT-SD symptom score for the period. If fewer than 4 symptom items with non-missing item-level score, the HPT-SD symptom score for the period is set to missing.

$$\text{HPT-SD symptom score} = \frac{\text{sum of non}-\text{missing item}-\text{level score over 7 symptom items}}{\text{number of symptom items with non}-\text{missing item}-\text{level score}}$$

For Items 8 (anxiety) and 9 (sadness or depression):

- Compute item-level score from daily responses for each item using equation from Step 1.
    - Require at least 8 days with non-missing responses during the 14-day period immediately prior to the visit date to compute item-level score. If fewer than 4 data points are available for each 7-day period, the item-level score for the period is set to missing.

For impact subscale (items 10-13):

Step 1: Compute item-level score from daily responses for each of the items in the impact subscale using equation (1). Require at least 8 days with non-missing responses during the 14-day period immediately prior to the visit date to compute item-level score. If fewer than 4 data points are available for each 7-day period, the item-level score for the period is set to missing.

Step 2: Compute HPT-SD impact cores using item-level scores calculated in Step 1. Per scoring algorithm, compute the average value of item-level score over items 10-13. If item-level score is missing for any of the 4 impact items, the impact score for the period is set to missing.

$$\text{HPT-SD impact score}=\frac{\text{sum of non}-\text{missing item}-\text{level score over 4 impact items}}{\text{number of impact items with non}-\text{missing item}-\text{level score}}$$

The baseline score is calculated based on the responses 2 weeks prior to Week 0.

Week 26 date is defined as last dose date plus 1.

6.5.1.2 Sensitivity Analysis

The following 7 sensitivity analyses will be performed for the primary endpoint, change from baseline in the HPT-SD symptom subscale score at week 26, to address departure from normality assumption (1), to adjust for baseline covariates (2), to address missing data (3,4) and to explore different HPT-SD symptom subscale scoring algorithms (5-7):

1) The change from baseline in HPT-SD symptom subscale score will be summarized by visits and will be compared between treatment groups using the Wilcoxon-Mann-Whitney test.
2) To explore the impact of adjustments to calcium and active vitamin D supplements on the primary endpoints, MMRM analyses will be repeated by including calcium and active vitamin D supplements as covariates in the model.
3) To explore the impact of missing data on the primary endpoint, the primary and secondary efficacy analysis will be repeated by applying last observation carried forward (LOCF) method for computed 14-day averaged HPT-SD symptom subscale score at all post-baseline visits (Weeks 2, 4, 6, 8, 10, 12, 14, 16, 20, 24, 26). For example, if the computed 14-day averaged HPT-SD symptom subscale score is missing at Week 26, the last non-missing score before week 26 will be carried forward. If all post-baseline results are missing, the baseline result will not be carried forward for the analysis.
4) Another approach to explore the impact of missing data on the primary endpoint is to use multiple imputation method. Imputations will be performed for missing 14-day averaged HPT-SD symptom subscale score at all post-baseline visits (Weeks 2, 4, 6, 8, 10, 12, 14, 16, 20, 24, 26) with the Markov chain Monte Carlo (MCMC) method. The imputation seed will be 634401 and the number of complete imputed datasets will be 20. For each of the 20 datasets the same analysis as the primary analysis will be performed and the SAS procedure MIANALYZE will be used to combine the results.
5) HPT-SD symptom subscale score at Week 26 will be calculated by replacing the last dose date with Week 26 visit date for the completed subjects. The primary analysis will be repeated using the new HPT-SD scoring algorithm.
6) HPT-SD scores will be calculated when there are at least 6 days with non-missing responses during the 14-day period immediately prior to the visit date to compute item-level score. If fewer than 3 data points are available for each 7-day period, the

item-level score for the period is set to missing. The primary analysis will be repeated using the new HPT-SD scores.

7) HPT-SD scores will be calculated when there are at least 4 days with non-missing responses during the 7-day period immediately prior to the visit date to compute item-level score. If fewer than 4 data points are available for the 7-day period, the item-level score for the period is set to missing. The primary analysis will be repeated using the new HPT-SD scores.

6.5.1.3 Supplementary Analyses

The analysis of primary endpoint (HPT-SD symptom subscale score) and key secondary endpoints (FACIT-Fatigue score and SF-36 PCS score) will be repeated for the 5 subgroups below:

- Gender: male vs. female
- Baseline calcium supplement dose: 0-2000 mg/day vs. >2000 mg/day
- Baseline active vitamin D dose: low dose 0-0.25 µg/day vs. medium dose >0.25-0.5 µg/day vs. high dose >0.5 µg/day
- History of thyroid hormone replacement: yes vs. no
- Baseline HPT-SD (symptoms subscale): moderate or below (0-2) vs. severe (>2)

Additionally, a shift table will be presented for the primary endpoint, to summarize 1-point change from baseline to each post-baseline visit. Percentages for shift tables will be based on the number of subjects with both baseline and post-baseline values at each visit.

6.5.2 Key Secondary Endpoint(s) Analysis

Secondary efficacy analyses will be performed on the ITT Set and repeated on PPS. Secondary efficacy endpoints will be analyzed similarly as the primary efficacy endpoint using an MMRM model and summarized by treatment group using descriptive statistics at each assessment visit, as described in SAP Section 6.5.1. Continuous endpoints will be summarized by mean, standard deviation, minimum, maximum, and 95% CIs. For categorical variables, statistical summaries will include number of subjects and percentages and 95% CIs for binary endpoints.

6.5.2.1 Change in FACIT-Fatigue score (version 4) at Week 26

The first key secondary endpoint is change from baseline in FACIT-Fatigue score at Week 26. FACIT-Fatigue questionnaire contains 13 fatigue-related questions. The responses to the 13 items on the FACIT-Fatigue questionnaire are each measured on a 5-point Likert scale. Thus, the total score ranges from 0 to 52. Higher fatigue subscale scores represent less fatigue and better quality of life.

The change from baseline in FACIT-Fatigue score at Week 26 will be analyzed similarly to the primary efficacy endpoint (HPT-SD symptom subscale), using a MMRM analysis over all post-baseline visits during the double-blind treatment phase.

For 2 questions, item #7 'I have energy' and item #8 'I am able to do my usual activities', no conversion is needed for the item score. For the rest of 11 questions, item score is calculated as 4- item response. Compute the fatigue total score as follows:

Fatigue Total Score = [sum of (item scores)]*13/ (#of items answered).

The observed value and change from baseline in Fatigue Total Score will be summarized by visit and graphically presented by line plots. In addition, response to each individual item will be summarized as categorical values by treatment group at each visit using descriptive statistics. By-subject listings will be provided.

6.5.2.2 Change in the PCS derived from SF-36v2 scores at Week 26

The second key secondary endpoint is change from baseline in PCS derived from SF-36v2 scores at Week 26. The SF-36 consists of 8 scaled scores, which are the weighted sums of the questions in their section. Each scale is directly transformed into a 0-100 scale on the assumption that each question carries equal weight; the lower the score the more disability. The higher the score the less disability i.e., a score of zero is equivalent to maximum disability and a score of 100 is equivalent to no disability. The eight sections included in the SF-36 assessment are physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health. The eight SF-36 scales will be aggregated using weight to calculate Physical Component Summary (PCS) and Mental (MCS) Component Summary scores. See for algorithm for deriving SF-36 scores in Appendix 9.6. The scoring of SF-36 will be computed using PRO CoRE™ Scoring Software.

The change from baseline in SF-36v2 PCS score at Week 26 will be analyzed similarly to the primary efficacy endpoint (HPT-SD), using a MMRM analysis over all post-baseline visits during the double-blind treatment phase. Line plots visit will also be generated for SF-36v2 PCS scores.

6.5.3 Other Secondary Endpoints Analysis

Secondary efficacy analyses will be performed on the ITT Set. Secondary efficacy endpoints will be analyzed similarly as the primary efficacy endpoint using an MMRM analysis where appropriate and summarized by treatment group using descriptive statistics at each assessment visit, as described in SAP Section 6.5.1. Continuous endpoints will be summarized by mean, standard deviation, minimum, maximum, and 95% CIs. For categorical variables, statistical summaries will include number of subjects and percentages and 95% CIs for binary endpoints.

6.5.3.1 Change in Other HPT-SD scores at Week 26

The change from baseline in HPT-SD impact subscale score, anxiety (item 8) score and sadness or depression (item 9) score, 7 individual HPT symptom item-level subscale (item 1-7) scores, 4 individual HPT impact item-level subscale (item 10-13) scores and most bothersome symptom score at Week 26 will be analyzed similarly to the primary efficacy endpoint (HPT-SD), using a MMRM analysis over all post-baseline visits during the double-blind treatment phase.

The most bothersome symptom score is calculated as 14-day averaged item score of the most bothersome symptom of each subject, similar to the primary analysis. The most bothersome symptom was selected by subjects at the baseline.

The number of symptom free days from study treatments start to Week 26/ET will be calculated as number of days with HPT-SD symptom subscale score equal to 0 for each subject. Complete resolution of symptoms is defined as subject with HPT-SD symptom subscale score equal to 0 at any point during the study period and remaining symptom free through the end of study. The number of symptom free days and complete resolution of symptoms will be summarized by number of subjects and percentages.

By-subject listings will be provided for questions and scores from HPT symptom questionnaire. The HTP-SD item score and impact subscale score will be graphically presented by line plots.

6.5.3.2 Response as defined as a 30% reduction in HPT symptom subscale score from Baseline to Week 26

A 30% reduction from baseline in HPT symptom subscale score at Week 26 represents benchmark of clinical significance for efficacy measures. This 30% efficacy reduction threshold for clinical meaningfulness models is established in the similar therapeutic area of clinical pain measure assessment (Farrar et al. 2001; Younger et al. 2009). The number of subjects, percentages, 95% CIs and p-values that are calculated from Chi-Square test achieving thresholds of ≥30% and ≥50% percentage of reduction in HPT-SD from baseline at Week 26 (or ET for early terminated subjects) will be reported for rhPTH(1-84) versus placebo-treated subjects.

6.5.3.3 Change in FACT-Cog score (Perceived Cognitive Impairment, Impact on Quality of Life domains, version 3) at Week 26

The FACT-Cog assessment include 24 items from 2 subscales, perceived cognitive impairment (18 items, Item #1 - #18, score range 0–72) and impact on quality of life domains life (4 items, Item #34 – #37, score range 0–16), rated on a 5-point scale ranging from 0 to 4. The item response of each question will be converted as described below to obtain an item score, except for perceived cognitive ability subscale. The subscale score is computed by multiplying the sum of the item scores by the number of items in the subscale, then dividing by the number of items answered. Higher scores indicate better quality of life.

1) Perceived cognitive impairments subscale (CogPCI):

The 18 items (item #1 - #18) from a total of 20 items in perceived cognitive impairments subscale are used, except for 2 questions, item #19 'I have trouble keeping track of what I am doing if am interrupted' and item #20 'I have trouble shifting back and forth between different activities that require thinking'. Compute the subscale score for the perceived cognitive impairments as follows:

CogPCI Subscale Score= [sum of (18-item response)]*18/ (#of items answered).

2) Impact on quality of life subscale (CogQOL):

The 4 items (Item #34 – #37) in impact on quality of life subscale are used. Compute the subscale score for the impact on quality of life as follows:

CogQOL Subscale Score = [sum of (4-item response)]*4/ (#of items answered).

The change from baseline in 2 subscale scores will be analyzed similarly to the primary efficacy endpoint (HPT-SD), using a MMRM analysis over all post-baseline visits during the double-blind treatment phase. In addition, response to each individual item will be summarized as categorical values by treatment group at each visit using descriptive statistics. By-subject listings will be provided.

6.5.3.4 Change in score of individual domains and mental component summary (MCS) of SF-36v2 at Week 26

The change from baseline in 8 scaled scores, standardized MCS will be analyzed similarly to the primary efficacy endpoint (HPT-SD), using a MMRM analysis over all post-baseline visits during the double-blind treatment phase. By-subject listings will be provided for SF-36v2 scores.

6.5.3.5 Change in Work Productivity and Activity Impairment Questionnaire: Specific Health Problem (WPAI:SHP; SHP = Hypoparathyroidism, version 2.0) score at Week 26

Subjects will be asked to complete the Work Productivity and Activity Impairment Questionnaire: Specific Health Problem (WPAI:SHP) once within the 2 days before each of the scheduled visit. The Specific Health Problem will be defined as Hypoparathyroidism in the instrument. The WPAI:SHP will be used to assess how hypoparathyroidism affects subjects' ability to work and perform regular activities. Concepts that the WPAI:SHP measures include time missed from work and impairment of work and other regular activities due to specific health problems. The following questions ask about the effect of subjects' HYPOPARATHYROIDISM (HPT) on their ability to work and perform regular activities. The questions (Q2-Q6) are about the past seven days, not including today.

Q1: Are you currently employed (working for pay)?

Q2: How many hours did you miss from work because of problems associated with HPT (hours of work missed due to HPT)?

Q3: How many hours did you miss from work because of any other reason (hours of work missed due to other reason)?

Q4: How many hours did you actually work (actual hours worked)?

Q5: HPT effect on productivity at work (HPT had no effect on work, 0-10, completely prevented from working)

Q6: HPT effect on daily activities (HPT had no effect on daily activities, 0-10, completely prevented from doing daily activities)

WPAI outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity, i.e., worse outcomes, as follows:

The scores for 4 subscales will be calculated as follows and expressed in percentages:

- Percent work time missed due to problem: Q2/(Q2+Q4)*100
- Percent impairment while working due to problem: Q5/10*100
- Percent overall work impairment due to problem: Q2/(Q2+Q4)+[(1-(Q2/(Q2+Q4)))x(Q5/10)] *100
- Percent activity impairment due to problem: Q6/10*100

The change from baseline in 4 subscale scores will be analyzed similarly to the primary efficacy endpoint (HPT-SD), using a MMRM analysis over all post-baseline visits during the double-blind treatment phase. In addition, response to each individual item will be summarized as categorical values by treatment group at each visit using descriptive statistics. By-subject listings will be provided.

6.5.3.6 Change in scores of patient's assessment of overall health status using Patient Global Impression of Severity (PGI-S, Version 1) and Patient Global Impression of Change (PGI-C, Version 1) at Week 26

Subjects will be asked to complete the PGI-S and the PGI-C assessments once within the 2 days before each of the scheduled visit.

The PGI-S is a global index that can be used to rate the severity of a specific condition. In this study, subjects will be asked to rate the severity of their hypoparathyroidism using a 5-point scale at each visit:

- No symptoms (0)
- Mild (1)
- Moderate (2)
- Severe (3)
- Very severe (4)

The PGI-C is used to rate change in hypoparathyroidism-related symptoms at specified visits compared to the beginning of this study by a 7-point scale:

- Very much improved (0)
- Much improved (1)
- Minimally improved (2)
- No change (3)
- Minimally worse (4)
- Much worse (5)
- Very much worse (6)

The change in severity assessed by PGI-S from baseline to week 26 will be analyzed similarly to the primary efficacy endpoint (HPT-SD), using an MMRM model that includes baseline PGI-S, treatment group, visit, and the interaction of treatment group and visit. The values of PGI-C scores at 26 will be summarized descriptively, by treatment group and by any other subgroup analyses that are relevant. By-subject listings will be provided.

6.5.3.7 Change in in-clinic neurocognitive assessment scores (CogState Brief Battery, CogState Groton Maze Learning Test, CogState International Shopping List Task, CogState International Shopping List Task-Delayed Recall) at Week 26, Change in at-home neurocognitive assessment scores (CogState Brief Battery) at Week 26

In-clinic neurocognitive assessments will be administered at Week -2 and Week 24 visits. The neurocognitive test battery will include tests evaluating the frontal-executive domain, which encompasses functions attributable to the prefrontal cortex and its connections to the basal ganglia (mostly striatum). The tests will include the CogState Brief Battery, CogState Groton Maze Learning Test, CogState International Shopping List Task, and CogState International Shopping List Task-Delayed Recall.

Subjects will complete the at-home CogState Brief Battery twice daily on each of the days between the Week -2 visit and the baseline (Week 0) visit, and for 14-day period prior to the scheduled visits at Week 12 and Week 26. Subjects will complete the testing once each morning after their morning routine (eg, dressing, eating, and taking medication) and once in the evening before dinner.

The analysis of in-clinic and at-home Cogstate neurocognitive assessments will be described in a separate SAP prepared by Cogstate.

6.5.3.8 Change in 24-hour urine calcium excretion at Week 26

The change from baseline in urinary calcium (mg/24hr) will be analyzed similarly to the primary efficacy endpoint (HPT-SD), using a MMRM analysis over all post-baseline visits during the double-blind treatment phase, with change from baseline in corresponding score as the outcome variable; treatment group, visit, and their interaction as fixed effect factors and subjects as a random effect with adjustment for corresponding baseline score. Least squares means (LSMEANs) and standard errors (SEs) with corresponding 95% CI will be calculated for each treatment group for each visit. The difference in LSMEANS with 95% CI between rhPTH(1-84) and placebo will be estimated, with the corresponding 2-sided 95% confidence interval constructed for each visit. The observed value and change from baseline will also be summarized by visit. By-subject listings will be provided.

6.5.3.9 Change in serum phosphate level (mmol/L) at Week 26

The change from baseline in serum phosphate (mmol/L) will be analyzed similarly to the primary efficacy endpoint (HPT-SD), using a mixed-effect model for repeated measures (MMRM) analysis over all post-baseline visits during the double-blind treatment phase. By-subject listings will be provided.

6.5.3.10 Changes in doses of active vitamin D (µg/day) and calcium supplements (mg/day) at Week 26

Efficacy endpoints based on the elemental calcium and active vitamin D (calcitriol or alfacalcidol) supplement will be derived using the investigator prescribed data.

The daily dose of elemental calcium and active vitamin D (calcitriol or alfacalcidol) supplement based on investigator prescription will be determined by using the latest prescribed dose prior to the assessment date of albumin corrected calcium for each analysis visit. The prescribed supplemental medication dose at the EOT Visit (Month 26) for subjects who complete the treatment or Early Termination will be excluded from the derivation of the post-baseline prescribed dose.

The change from baseline in elemental calcium and active vitamin D (calcitriol or alfacalcidol) supplement will be analyzed similarly to the primary efficacy endpoint (HPT-SD), using a MMRM analysis over all post-baseline visits during the double-blind treatment phase. In addition, the frequency of dose change (0, 1-2, 4-5…) in prescribed elemental calcium and active vitamin D (calcitriol or alfacalcidol) supplement during double-blind treatment period will be presented by treatment group. By-subject listings will be provided.

6.5.3.11 Proportion of subjects achieving composite endpoint at Week 26

- Albumin-corrected serum calcium between 1.87 mmol/L (7.5 mg/dL) and the ULN for the central laboratory normal range
- Dose of prescribed active vitamin D decreased by 50% at Week 26
- At least a 50% reduction from the baseline prescribed oral elemental calcium supplement dose (this criterion will be considered met if the subject's baseline calcium dose is <1000 mg and it does not increase during the study).

The number of subjects and percentages achieving each criterion and all the 3 criteria above at Week 26 will be presented by treatment group. The 95% CIs of proportion of subjects with achieving all the 3 criteria above at Week 26 and p-values for treatment will be reported using chi-square test. In addition, the observed value, change and percent change from baseline in albumin-corrected serum calcium (mmol/L) will be summarized by visit. Albumin-corrected serum calcium (mmol/L) will be calculated as total calcium (mmol/L) + 0.02 * [40 g/L - albumin (g/L)].

6.5.3.12 Markers of bone turnover biomarkers at Week 26

The change from baseline in bone turnover biomarkers, including serum bone-specific alkaline phosphatase, procollagen amino-terminal peptide, C-terminal telopeptide of type 1 collagen, and osteocalcin will be analyzed similarly to the primary efficacy endpoint (HPT- SD), using a MMRM analysis over all post-baseline visits during the double-blind treatment phase. By-subject listings will be provided.

6.5.3.13 Sensitivity analysis for subjects with duplicated ERT results

Sensitivity analysis will be conducted for secondary endpoints. Subjects with duplicated ERT result for FACT, FACIT, SF36, PGI-C, PGI-S, WPAI will be removed and secondary main analysis will be repeated.

6.5.4 Exploratory Efficacy Endpoint(s) and Analyses

6.5.4.1 Plasma FGF-23 levels

Plasma FGF-23 levels will be measured at Week 0 and Week 26 (EOT). The observed value, change and percent change from baseline will be summarized by treatment group. By-subject listings will be provided.

6.5.4.2 24-hour urine phosphate, sodium, citrate, magnesium excretion

The observed value, change and percent change from baseline in urinary phosphate, sodium, citrate and magnesium will be summarized by visit and at EOT. By-subject listings will be provided.

6.5.4.3 Creatinine clearance

The observed value, change and percent change from baseline in creatinine clearance will be summarized by visit and at EOT. By-subject listings will be provided.

Creatinine clearance (mL/min) = [Urine creatinine concentration (mg/dl) * Urine flow rate (L/day)] / Serum creatinine concentration (mg/dl)*1000/1440 (min/day). Then creatinine clearance will be normalized to a body surface area as CrCl * 1.73/BSA. The subject's BSA can be calculated with Dubois formula: 0.20247 * height (m)^0.725 * weight (kg)^0.425.

6.5.4.4 Estimated glomerular filtration rate (eGFR)

The observed value, change and percent change from baseline in eGFR will be summarized by visit and at EOT. By-subject listings will be provided.

eGFR (mL/min/1.73m^2) will be calculated from serum creatinine using the equation that the CKD-Epidemiology Collaboration group (CKD-EPI) developed and validated in 2009 which is called the CKD-EPI equation. The equation is as follows (where Scr is serum creatinine (mg/dL), k is 0.7 for females and 0.9 for males, a is -0.329 for females and -0.411 for males, min indicates the minimum of Scr/k or 1, and max indicates the maximum of Scr/k or 1):

$$eGFR = 141 \times \min(Scr/k, 1)^{a} \times \max(Scr/k, 1)^{-1.209} \times 0.993^{Age} \times 1.018 (\text{if female}) \times 1.159 (\text{if black})$$

6.5.4.5 EuroQol five dimensions questionnaire 5-level version (EQ-5D-5L): EQ-5D-5L index value, descriptive system, and EQ VAS measure

The EQ-5D is a generic, multi-attribute, HRQoL instrument composed of a descriptive system and a visual analog scale (VAS). The EQ-5D descriptive system has the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has five levels of severity (states consisting of 5 dimensions no problems (1), slight problems (2), moderate problems (3), severe problems (4), and extreme problems (5). The EQ-5D VAS records

the subject's self-rated health state on a 100-point vertical VAS (0=worst imaginable health state; 100=best imaginable health state).

EQ-5D-5L index value (utility index scores) will be calculated using Crosswalk Index Value Calculator. For the countries not in the calculator, the value set of United Kingdom will be used.

The number and percentage of subjects endorsing each dimension will be summarized categorically at each visit and EOT (Week 26) by treatment group. The observed value and change from baseline in analogue scale Score (VAS) and utility index (UI) scores will be presented at each visit and EOT by treatment. By-subject listings will be provided.

6.5.4.6 Measures of healthcare resource utilization: frequency of encounters (outpatient visits, laboratory tests, and procedures not scheduled in the protocol; emergency room visits; hospitalization), length of stay (hospitalization), and reasons/diagnoses for the encounters.

Measures of healthcare resource utilization will include encounters such as outpatient visits, laboratory tests, and procedures that are not scheduled in the protocol, emergency department visits, and hospitalizations. The responses to each question of health care services utilization questionnaire will be summarized descriptively over time by visit. For subjects who were in a hospital overnight, the total duration of hospitalization (days) will be calculated as the sum of all hospitalization durations at each visit. The average duration of hospitalizations (days) will be calculated as the mean duration of hospitalization per subject. The total and average duration of hospitalization will be summarized by visit. The response of health care services utilization questionnaire will be listed.

6.5.4.7 Dose of native vitamin D supplements (recorded as concomitant medications)

Dose of prescribed native vitamin D supplements will be summarized descriptively by treatment group similar prescribed active vitamin D calcitriol supplement based on investigator prescription each analysis visit, as described in SAP Section 6.5.3.10.

6.6 Safety Analysis

The safety analysis will be performed using the Safety Analysis Set. All values presented in summaries will be observed values.

6.6.1 Adverse Events

Classification of Adverse Events will be accomplished per documented MedDRA Dictionary version 20.1 in the Data Validation Manual.

Treatment-emergent adverse events (TEAEs) are defined as AEs whose onset occurs, severity worsens or intensity increases after receiving the first dose of double-blind investigational product. Events which occur more than 30 days after the last dose of study drug will not be considered treatment-emergent. If any AE records contain only partial dates, these will be handled by imputation as described in SAP Section 9.5.3.

An overall summary of the number of subjects with TEAEs will be presented, including the number of events (except for summaries by highest category), number and percentage of subjects with any TEAEs, hypocalcemic TEAE, treatment-related TEAEs, TEAEs leading to withdrawal, severity of TEAEs (any and highest category), serious TEAE, relationship of serious TEAE to treatment, severity of serious TEAE, and TEAEs leading to death.

The number of events (except for summaries by highest category), number and percentage of subjects reporting TEAEs in each treatment group will be tabulated by system organ class (SOC) and preferred term; by SOC, preferred term, and maximum severity; by SOC, preferred term, and relationship. Hypocalcemic TEAE, TEAE related to investigational product, TEAE leading to withdrawal, SAEs, and deaths will be similarly summarized and listed by SOC and preferred term. If more than 1 AE occurs with the same preferred term for the same subject, then the subject will be counted only once for that preferred term using the most severe and most related occurrence for the summarization by severity and by relationship to investigational product.

Listings will be provided for all AE, AEs related to investigational product, AEs leading to withdrawal, SAEs, and deaths. Adverse event listings will provide the verbatim term as well as the SOC and preferred term for each recorded event.

6.6.2 Other Safety Analysis

6.6.2.1 Clinical Laboratory Variables

Descriptive statistics for clinical laboratory values (in SI and conventional units) and changes from baseline will be reported by treatment group at each visit for quantitative parameters. Change will only be reported if both baseline and the corresponding post-baseline time point are available.

Additionally, shift tables will be presented, summarizing cross tabulations of low, normal, and high based on the parameter normal range, from baseline to each post-baseline visit and EOT. Percentages for shift tables will be based on the number of subjects with both baseline and post-baseline values at each visit. All the laboratory data below will be listed by parameter, subject and visit. Only scheduled lab parameters will be included in the lab summaries.

The following clinical laboratory parameters will be summarized, except for the parameters that are separately summarized in efficacy analysis:

Hematology	Hemoglobin, Platelet count, Hematocrit, White blood cell (WBC) count, Red blood cells (RBC), Automatic differential
Serum chemistry	Sodium, Potassium, Blood urea nitrogen, Creatinine, Chloride, Total carbon dioxide (bicarbonate), Aspartate aminotransferase, Alanine aminotransferase, Alkaline phosphatase, Uric acid, PTH, TSH, FSH, Calcium, Albumin, Phosphate, Magnesium, 25-hydroxyvitamin D, 1,25-dihydoxyvitamin D levels
Urinalysis	pH, Blood, Glucose, Ketones, Leukocyte esterase, Protein, Bilirubin, Specific gravity

Urine chemistry Calcium, Phosphate, Creatinine, Sodium, Magnesium, Potassium, Chloride, Citrate, pH, Urea nitrogen, Oxalate, Ammonium, Uric acid, Sulfate**,** Calcium per Creatinine, Calcium per Kg Body Weight, Creatinine per Kg Body Weight, Protein Catabolic Rate, Calcium oxalate, Calcium phosphate (brushite), Uric acid

Clinical laboratory test values are potentially clinically important (PCI) if they meet either the low or high PCI criteria listed in Table 2. PCI values will be flagged in the by-subject laboratory listings.

Table 2 Criteria for Potentially Clinically Important Laboratory Tests

Parameter	SI Unit	Lower Limit	Higher Limit
Biochemistry			
Alkaline Phosphatase	U/L	NA	>2*ULN
Albumin	g/L	<=20	>=90
Alanine aminotransferase	U/L	NA	>3*ULN
Aspartate aminotransferase	U/L	NA	>3*ULN
Blood urea nitrogen	mmol/L	NA	>=10.7
Calcium (total)	mmol/L	<=2.1	>=3.0
Chloride	mmol/L	<=80	>=125
Creatinine	µmol/L	NA	>=177
Creatinine clearance	ml/min	<=60	NA
Potassium	mmol/L	<=2.5	>=6.5
Phosphate	mmol/L	NA	>=2
Sodium	mmol/L	<=120	>=165
Hematology			
Hematocrit	L/L	<=0.37 (males) <=0.32 (females)	>0.54 (males) NA (females)
Hemoglobin	g/L	<=115 (males) <=95 (females)	NA
Platelets	10^9/L	<=75	>=700
Red blood cells (RBC)	10^12/L	<=2.5 (males) <=2.0 (females)	NA
White blood cell (WBC)	10^9/L	<=2.8	>=16.0
24-Hour urine			
Urine calcium	mg	NA	>300 for men >250 for women

6.6.2.2 Vital Signs

Descriptive statistics for vital signs (blood pressure, pulse, body temperature, and respiration rate) and their changes from baseline by treatment group at each post-baseline visit will be presented.

Vital sign values will be considered PCI if they meet both the observed value criteria and the change from baseline criteria listed in Table 3. PCI values will be flagged in the by – subject vital sign listing.

Table 3 Criteria for Potentially Clinically Significant Vital Signs

Vital Sign Parameter	Flag	Criteria[a]	
		Observed Value	Change from Baseline
Systolic blood pressure (mmHg)	High	≥180	Increase of ≥20
	Low	≤90	Decrease of ≥20
Diastolic blood pressure (mmHg)	High	≥105	Increase of ≥15
	Low	≤50	Decrease of ≥15
Pulse (bpm)	High	≥120	Increase of ≥15
	Low	≤50	Decrease of ≥15
Temperature (°C/°F)	High	≥38.3	Increase of ≥0.8
	Low	-	-

[a] post-baseline value is considered as a PCI value if its meets both criteria for observed value and change from baseline.

6.6.2.3 Electrocardiogram (ECG)

Descriptive statistics for ECG variables (eg, heart rate, PR interval, QRS interval, QT interval, and QTc interval) and their changes from baseline at each assessment time point will be presented by treatment group. QTc interval will be calculated using both Bazett ($QTcB=QT/(RR)^{1/2}$) and Fridericia ($QTcF=QT/(RR)^{1/3}$) corrections; and if RR is not available, it will be replaced with 60/hr in the correction formula. A shift table from baseline to each visit for ECG interpretation will be presented.

Electrocardiogram variable values will be considered PCI if they meet or exceed the upper limit values listed in Table 4. PCI values will be flagged in the by – subject ECG listing.

Table 4 Criteria for Potentially Clinically Important ECG Values

ECG Parameter	Unit	Higher Limit
QRS Interval	msec	≥150
PR Interval	msec	≥250
QTc Interval	msec	≥500

6.6.2.4 Physical Examination

Physical examination results will be listed only.

6.6.2.5 Serum and Urine Pregnancy

Categorical urinalysis findings and urine pregnancy results will be listed only.

6.6.2.6 Anti-PTH Antibodies

The number and percentage of subjects classified as having antibodies to PTH at each visit will be tabulated and listed.

6.6.3 Extent of Exposure and Compliance

6.6.3.1 Exposure to Investigational product

Exposure to double-blind investigational product for the Safety Analysis Set will be summarized in terms of treatment duration in weeks by treatment period (dose-titration Period and maintenance-dosing Period) and overall double-blind treatment period, which is calculated as [(date of last dose - date of first dose in each treatment period) +1]/ 7. The extent of exposure will also be categorized into weeks (<=1, >1-8, >8-16, >16-24, >24-26 and >26) and tabulated. The average dose of study drug is calculated as total dose amount/ total number of injections for each treatment period and overall double-blind treatment period. Descriptive statistics will be presented for exposure and average dose of study drug.

6.6.3.2 Measurement of Treatment Compliance

Treatment compliance is defined as [100*doses administered / days on treatment] for each treatment period and overall double-blind treatment period, where days on treatment is calculated as (date of last dose in the period – date of first dose of the period) + 1. Descriptive statistics for investigational product compliance will be presented as continues parameter and by categories of <=80% and >80%. Subjects will be considered compliant overall for study drug if the calculated compliance is >80%.

Exposure, treatment compliance and dispense/account information will be listed for each subject.

6.7 Pharmacokinetic, Pharmacodynamic, and Biomarker Analyses

Not applicable.

6.8 Patient Reported Outcomes (PROs) and Health Care Utilization Endpoints Analysis

PRO Analysis and Health Care Utilization Analysis will be summarized as described in SAP Section 6.5.

6.9 Other Analyses

The COVID-19 analysis will be performed to assess the impact of COVID-19 to the response of primary and secondary endpoints where questionnaires are used in response. The repeated measures (MMRM) analysis will be repeated for the following endpoints by the subject who

completed or discontinued prior to the COVID-19 pandemic vs subjects who are still in the study during the COVID-19 pandemic:

- Change in the HPT-SD symptom subscale score from baseline to Week 26
- Change in FACIT-Fatigue score score at Week 26
- Change in the PCS derived from SF-36v2 scores at Week 26
- Change in the HPT-SD impact subscale score
- Change in score of individual HPT-SD impact items
- Change in the HPT-SD symptom item anxiety and symptom item sadness or depression individual item score
- Change in score of individual HPT-SD symptom items
- Response as defined as a 30% reduction in HPT symptom subscale score from Baseline to Week 26
- Change in the most bothersome symptom score
- Change in FACT-Cog score (Perceived Cognitive Impairment, Impact on Quality of Life domains)
- Change in score of individual domains of SF-36v2
- Change in score of MCS of SF-36v2
- Change in WPAI:Hypoparathyroidism score
- Change in scores of patient's assessment of overall health status using:
    - Patient PGI-S for Hypoparathyroidism
    - Patient PGI-C for Hypoparathyroidism

Subject who completed or discontinued prior to the COVID-19 pandemic is defined as subject with treatment end date prior to local pandemic start date, and subjects who are still in the study during the COVID-19 pandemic is defined as subject with treatment end date on or after local pandemic start date. The start date of COVID pandemic is defined as the date of first case that was reported in a specific country:

Table 5 Start Date of COVID Pandemic

Country	Start Date of COVID Pandemic
Spain	31st Jan 2020 (Working group for the surveillance and control of COVID-19 in Spain, et al, 2020)
United Kingdom	30th Jan 2020 (Lillie PJ, et al, 2020)
Israel	21st Feb 2020 (Last M, 2020)
United States of America	20th Jan 2020 (Holshue, et al, 2020)
Belgian	3rd Feb 2020 (Luyten and Schokkaret, 2020)

Table 5 Start Date of COVID Pandemic

Country	Start Date of COVID Pandemic
Canada	27th Jan 2020 (Zhao, et al, 2020)
Denmark	27th Feb 2020 (Holler, et al, 2021)
France	24th Jan 2020 (Or, et al, 2021)
Italy	23rd Jan 2020 (Giovanetti, et al, 2020)
Netherlands	27th Feb 2020 (Hoekman, et al, 2020)
Norway	26th Feb 2020 (Ursin, et al, 2020)
Portugal	2nd Mar 2020 (Milhinho, et al, 2020)
Sweden	31st Jan 2020 (Pashakhanlou, et al, 2020)

6.10 Interim Analyses

An unblinded, interim analysis for the primary endpoint will be performed after 74% of all randomized subjects (N= approximately 68) have either completed 26 weeks of treatment (completed the study) or prematurely withdrawn from the study, whichever comes first. The purpose of this unblinded interim analysis is to reassess the appropriateness of assumptions used for the sample size calculation of the primary efficacy endpoint when the study was designed and to assess for futility. The reassessment of the sample size will utilize the conditional power approach under certain conditions that do not inflate the type I error (Mehta and Pocock, 2011). The planned interim analysis will be conducted by an external independent statistical group; the individuals involved in the day-to-day conduct of the trial will not be involved in the interim analysis or have access to the results of the interim analysis.

The sponsor will only be notified by the external independent statistical group of their recommendations to (1) stop the trial early for futility, (2) maintain the sample size as outlined in the current study design, or (3) update the sample size from the conditional power calculation; this will be detailed in the pre-specified interim SAP. Details of the interim analysis has been documented separately in the interim analysis SAP.

6.11 Data Monitoring Committee/Internal Review Committee/ [Other Data Review Committees]

Not applicable.

6.12 Missing Data Analysis

Handling of missing primary efficacy endpoint is described in SAP Section 6.5.1.2. See Appendix 9.5 for missing date imputation method of safety variables.

7.0 REFERENCES

Giovanetti M., Benvenuto D., Angeletti S., Ciccozzi M. The first two cases of 2019-nCoV in Italy: Where they come from? J. Med. Virol. 2020;92:518–521. doi: 10.1002/jmv.25699.

Hoekman LM, Smits MMV, Koolman X. The Dutch COVID-19 approach: Regional differences in a small country. Health Policy Technol 2020;9(4):613–622. doi: 10.1016/j.hlpt.2020.08.008.

Holler J.G., Eriksson R., Jensen T.O., van Wijhe M., Fischer T.K., Sogaard O.S. First wave of COVID-19 hospital admissions in Denmark: a Nationwide population-based cohort study. BMC Infect Dis. 2021;21:39.

Holshue ML, DeBolt C, Lindquist S, et al. First Case of 2019 Novel Coronavirus in the United States. N Engl J Med. 2020 March; 382:929-936.

Last M. The first wave of COVID-19 in Israel—Initial analysis of publicly available data. Plos One. 2020 Oct.; https://doi.org/10.1371/journal.pone.0240393.

Lillie PJ, Samson A, Li A, et al. Novel coronavirus disease (Covid-19): The first two patients in the UK with person to person transmission. J Infect. 2020 May; 80(5):578–606.

Luyten J and Schokkaert E. Belgium's response to the COVID-19 pandemic. Health Econ Policy Law. 2021 Jul 5 : 1–11.

Mehta, C. R. and Pocock, S. J. 2011. Adaptive increase in sample size when interim results are promising: a practical guide with examples. Statistics in medicine, 30, 3267-84.

Milhinho A., Costa P.R. On the Progression of COVID-19 in Portugal: A Comparative Analysis of Active Cases Using Non-Linear Regression. Front. Public Health. 2020;8:1–6.

Or Z, Gandre C, Durand Zaleski I, Steffen M. France's response to the Covid-19 pandemic: between a rock and a hard place. Health Econ Policy Law. (2021) 2021:1–13. 10.1017/S1744133121000165.

Pashakhanlou AH. Sweden's coronavirus strategy: The Public Health Agency and the sites of controversy. World Medical & Health Policy. 2021. 10.1002/wmh3.449.

Ursin G, Skjesol I, Tritter J. The COVID-19 pandemic in Norway: the dominance of social implications in framing the policy response. Health Policy and Technology. 2020;9(4):663–672.

Ware J.E., Jr., Kosinski M., Bjorner J.B., Turner-Bowker, D.M., Gandek, B. and Maruish, M.E. (2007) User's Manual for the SF- 36v2TM Health Survey (2nd Ed). Lincoln, RI, QualityMetric Incorporated.

Working group for the surveillance and control of COVID-19 in Spain. The first wave of the COVID-19 pandemic in Spain: characterisation of cases and risk factors for severe outcomes, as at 27 April 2020. Euro Surveill. 2020;25(50): 2001431. 10.2807/1560-7917.ES.2020.25.50.2001431d.

Zhao N, Liu Y, Smaragiassi A, Bernatsky S. Tracking the origin of early COVID-19 cases in Canada. Int J Infect Dis. 2020.

8.0 CHANGES TO PROTOCOL PLANNED ANALYSES

Not applicable.

9.0 APPENDIX

9.1 Changes From the Previous Version of the SAP

SAP Section	Impacted Text (shown in bold)	Change	Rationale for Change
Not applicable.			

9.2 Data Handling Conventions

9.2.1 General Data Reporting Conventions

Continuous variables will be summarized using the following descriptive statistics: n, mean, median, standard deviation, minimum and maximum. Categorical and count variables will be summarized by the number of subjects (n) and the percent of subjects in each category. Percentages will be presented as whole numbers.

For rules on rounding and decimal presentation of these statistics, along with the rules for presenting certain derived values, see below:

- For measures of median and mean, use 1 decimal place beyond those used for the measurement.
- For measures of standard deviation and standard error, use 2 decimal places beyond those used for the measurement.
- For measures of minimum and maximum values, use the same number of decimal places as those used for the measurement.
- >=5 is rounded up away from zero, whereas <5 is rounded down toward zero to account for rounding of negative numbers.
- For p-values use 3 decimal places.
- Presentation of p-values, display p-values that would round to 0.000 as <0.001.
- BMI should be rounded to 1 decimal place for reporting.
- Derived questionnaire scores, and other similar efficacy parameters recorded as integers, should be rounded to 1 decimal place for reporting.
- Averaged lab/vital/ECG results (when taken in triplicate) should be rounded to 1 decimal place for reporting.

9.3 Definition of Visit Windows

Refer to SAP Section 6.1.3.

9.4 Repeated or Unscheduled Assessments of Safety Parameters

- If a subject has repeated assessments before the start of investigational product, then the results from the final assessment made prior to the start of investigational product will be used as baseline.
- If end of study assessments is repeated or unscheduled, the last post-baseline assessment will be used as the end of study assessment for all analyses.
- If a subject has repeated assessments in the same post-baseline scheduled visit, the last non-missing assessment in the visit will be used for all analyses.
- All post-baseline assessments will be used for PCS value determination and all assessments will be presented in the data listings.

9.5 Handling of Missing, Unused, and Spurious Data

9.5.1 Missing Date of Investigational Product

When the date of the last dose of investigational product is missing for a subject in the Safety Population, all efforts should be made to obtain the date from the investigator. If it is still missing after all efforts, then the last visit date when investigational product was returned will be used in the calculation of treatment duration.

9.5.2 Missing Date Information for Prior or Concomitant Medications (Therapies/Procedures)

For prior or concomitant medications (and/or therapies/procedures), including rescue medications, incomplete (i.e., partially missing) start date and/or stop date will be imputed. When the start date and the stop date are both incomplete for a subject, impute the start date first.

9.5.2.1 Incomplete Start Date

The following rules will be applied to impute the missing numerical fields. If the stop date is complete and the imputed start date is after the stop date, then the start date will be imputed using the stop date.

9.5.2.1.1 Missing Day and Month

- If the year of the incomplete start date is the same as the year of the date of the first dose of investigational product, then the day and month of the date of the first dose of investigational product will be assigned to the missing fields.
- If the year of the incomplete start date is before the year of the date of the first dose of investigational product, then December 31 will be assigned to the missing fields.
- If the year of the incomplete start date is after the year of the date of the first dose of investigational product, then 01 January will be assigned to the missing fields.

9.5.2.1.2 Missing Month Only

- The day will be treated as missing and both month and day will be replaced according to the above procedure.

9.5.2.1.3 Missing Day Only

- If the month and year of the incomplete start date are the same as the month and year of the date of the first dose of investigational product, then the day of the date of the first dose of investigational product will be assigned to the missing day.
- If either the year is before the year of the date of the first dose of investigational product or if both years are the same but the month is before the month of the date of the first dose of investigational product, then the last day of the month will be assigned to the missing day.
- If either the year is after the year of the date of the first dose of investigational product or if both years are the same but the month is after the month of the date of the first dose of investigational product, then the first day of the month will be assigned to the missing day.

9.5.2.2 Incomplete Stop Date

The following rules will be applied to impute the missing numerical fields. If the date of the last dose of investigational product is missing, then replace it with the last visit date. If the imputed stop date is before the start date (imputed or non-imputed start date), then the imputed stop date will be equal to the start date.

9.5.2.2.1 Missing Day and Month

- If the year of the incomplete stop date is the same as the year as of the date of the last dose of investigational product, then the day and month of the date of the last dose of investigational product will be assigned to the missing fields.
- If the year of the incomplete stop date is before the year of the date of the last dose of investigational product, then 31 December will be assigned to the missing fields.
- If the year of the incomplete stop date is after the year of the date of the last dose of investigational product, then 31 December will be assigned to the missing fields. If the imputed date is after the last date of the study, the last date of the study will be assigned.

9.5.2.2.2 Missing Month Only

- The day will be treated as missing and both month and day will be replaced according to the above procedure.

9.5.2.2.3 Missing Day Only

- If the month and year of the incomplete stop date are the same as the month and year of the date of the last dose of investigational product, then the day of the date of the last dose of investigational product will be assigned to the missing day.

- If either the year is before the year of the date of the last dose of investigational product or if both years are the same but the month is before the month of the date of the last dose of investigational product, then the last day of the month will be assigned to the missing day.
- If either the year is after the year of the last dose of investigational product or if both years are the same but the month is after the month of the date of the last dose of investigational product, then the first day of the month will be assigned to the missing day.

9.5.3 Missing Date Information for Adverse Events

For AEs, the start dates will be imputed following the same rules as in SAP Section 9.5.2. Adverse events with completely missing onset dates and a stop date after the date of first dose (or unknown stop date) will also be considered treatment-emergent.

9.5.4 Missing Severity Assessment for Adverse Events

If severity is missing for an AE starting prior to the date of the first dose of investigational product, then a severity of "Mild" will be assigned. If the severity is missing for an AE starting on or after the date of the first dose of investigational product, then a severity of "Severe" will be assigned. The imputed values for severity assessment will be used for incidence summaries, while the actual values will be used in data listings.

9.5.5 Missing Relationship to Investigational Product for Adverse Events

If the relationship to investigational product is missing for an AE starting on or after the date of the first dose of investigational product, a causality of "Related" will be assigned. The imputed values for relationship to investigational product will be used for incidence summaries, while the actual values will be presented in data listings.

9.6 Algorithm For Deriving Sf36 Aggregate Scores (Sf-36®, Version 2, Acute Version)

To derive SF36 SCORE the following algorithm was applied as in Ware, J.E. Jr et al (2007): User's manual for the SF-36v2® Health Survey (2nd ed.).

Item recoding:

Some of the items in the SF-36 were recoded as follows:

Precoded and final values for items 3a-3j:

Response choices	Precoded item value	Final item value
Yes, limited a lot	1	1
Yes, limited a little	2	2
No, not limited at all	3	3

Precoded and final values for items 4a-4d:

Response choices	Precoded item value	Final item value
All of the time	1	1
Most of the time	2	2
Some of the time	3	3
A little of the time	4	4
None of the time	5	5

Precoded and final values for item 7:

Response choices	Precoded item value	Final item value
None	1	6.0
Very mild	2	5.4
Mild	3	4.2
Moderate	4	3.1
Severe	5	2.2
Very severe	6	1.0

Scoring for item 8--if both items 7 and 8 are answered:

Response choices	If item 8 precoded item value	And	Item 7 precoded item value	then	Item 8 final item value
Not at all	1		1		6
Not at all	1		2 through 6		5
A little bit	2		1 through 6		4
Moderately	3		1 through 6		3
Quite a bit	4		1 through 6		2
Extremely	5		1 through 6		1

Scoring for item 8--if item 7 is not answered:

Response choices	Precoded item value	Final item value
Not at all	1	6.0
A little bit	2	4.75
Moderately	3	3.5
Quite a bit	4	2.25
Extremely	5	1.0

Precoded and final values for items 1 and 11a-11d:

Item 1	Response choices	Precoded item value	Final item value
	Excellent	1	5.0
	Very good	2	4.4
	Good	3	3.4
	Fair	4	2.0
	Poor	5	1.0

Item 11a & 11c	Response choices	Precoded item value	Final item value
	Definitely true	1	1
	Mostly true	2	2
	Don't know	3	3
	Mostly false	4	4
	Definitely false	5	5

Item 11b & 11d	Response choices	Precoded item value	Final item value
	Definitely true	1	5
	Mostly true	2	4
	Don't know	3	3
	Mostly false	4	2
	Definitely false	5	1

Precoded and final values for items 9a, 9e, 9g, & 9i:

Item 9a & 9e	Response choices	Precoded item value	Final item value
	All of the time	1	5
	Most of the time	2	4
	Some of the time	3	3
	A little of the time	4	2
	None of the time	5	1

Item 9g & 9i	Response choices	Precoded item value	Final item value

All of the time	1	1
Most of the time	2	2
Some of the time	3	3
A little of the time	4	4
None of the time	5	5

Precoded and final values for items 6 & 10:

Item 6	Response choices	Precoded item value	Final item value
	Not at all	1	5
	Slightly	2	4
	Moderately	3	3
	Quite a bit	4	2
	Extremely	5	1

Item 10	Response choices	Precoded item value	Final item value
	All of the time	1	1
	Most of the time	2	2
	Some of the time	3	3
	A little of the time	4	4
	None of the time	5	5

Precoded and final values for items 5a - 5c:

Response choices	Precoded item value	Final item value
All of the time	1	1
Most of the time	2	2
Some of the time	3	3
A little of the time	4	4
None of the time	5	5

Precoded and final values for items 9b, 9c, 9d, 9f, & 9h:

Items 9b,9c, & 9f	Response choices	Precoded item value	Final item value
	All of the time	1	1
	Most of the time	2	2
	Some of the time	3	3
	A little of the time	4	4
	None of the time	5	5

Item 9d & 9h	Response choices	Precoded item value	Final item value
	All of the time	1	5
	Most of the time	2	4
	Some of the time	3	3
	A little of the time	4	2
	None of the time	5	1

Transformation of scale scores:

After item recoding, a raw score is computed for each scale and transformed to a 0 to 100 scale as shown below:

Scale	Actual raw score (sum final item values after recoding)	Lowest and highest possible raw scores	Possible raw score range
Physical functioning (PF)	3a+3b+3c+3d+3e+3f+ 3g+3h+3i+3j	10, 30	20
Role-physical (RP)	4a+4b+4c+4d	4, 20	16
Bodily pain (BP)	7+8	2, 12	10
General health (GH)	1+11a+11b+11c+11d	5, 25	20
Vitality (VT)	9a+9e+9g+9I	4, 20	16
Social functioning (SF)	6+10	2, 10	8
Role-emotional (RE)	5a+5b+5c	3, 15	12
Mental health (MH)	9b+9c+9d+9f+9h	5, 25	20

Formula and example for transformation of raw scale scores:

$$\text{Transformed scale} = \left[\frac{(\text{Actual raw score} - \text{lowest } \textit{possible} \text{ raw score})}{\textit{possible} \text{ raw score range}} \right] \times 100$$

Example: A Physical Functioning raw score of 21 would be transformed as follows:

$$\left[\frac{(21 - 10)}{20} \right] \times 100 = 55$$

Where lowest possible score = 10 and possible raw score range = 20

Scoring of the PCS and MCS measures

Scoring of the Physical (PCS) and Mental (MCS) Component Summary measures involves three steps. First, the eight SF-36 scales are standardized using means and standard deviations from the general US population. Second, they are aggregated using weights (factor score coefficients) from the general US population. Finally aggregate PCS and MCS scores are standardized using a linear T-score transformation to have a mean of 50 and a standard deviation of 10, in the general US Population.

General US population statistics used in the standardization and in the aggregation of SF-36 scale scores are presented in the table below. Detailed information including formulas for scale aggregation and transformation of scores are presented below:

General US Population Means, Standard Deviations and Factor Score Coefficients Used to Derive PCS and MCS Scale Scores

			Factor Score Coefficients	
SF-36 Scale	**Mean**	**SD**	**PCS**	**MCS**
PF	82.62455	24.43176	0.42402	-0.22999
RP	82.65109	26.19282	0.35119	-0.12329
BP	73.86999	24.00884	0.31754	-0.09731
GH	70.78372	21.28902	0.24954	-0.01571
VT	58.41968	20.87823	0.02877	0.23534
SF	85.11568	23.24464	-0.00753	0.26876
RE	87.50009	22.01216	-0.19206.	0.43407
MH	75.76034	18.04746	-0.22069	0.48581

The PCS and MCS scales are scored using norm-based methods. The means, standard deviations, and factor score coefficients used in scoring come from the general US population. A liner

T-score transformation method was used so that both the PCS and MCS have a mean of 50 and a standard deviation of 10 in the general US population.

Step 1. Formulas for z-score standardizations of SF-36 scales:

PF_Z = (PF - 82.62455) / 24.43176

RP_Z = (RP -82.65109) / 26.19282

BP_Z = (BP - 73.86999) / 24.00884

GH_Z = (GH - 70.78372) / 21.28902

VT_ Z = (VT - 58.41968) / 20.87823

SF_Z = (SF - 85.11568) / 23.24464

RE_Z = (RE - 87.50009) / 22.01216

MH_Z = (MH - 75.76034) / 18.04746

Step 2. Formulas for aggregating standardized scales in estimating aggregate physical and mental components scores:

AGG_PHYS = (PF_Z* .42402) + (RP_Z* .35119) + (BP_Z* .31754)
+ (GH_Z* .24954) + (VT_Z* .02877) + (SF_Z* -.00753) +
(RE_Z* -.19206) + (MH_Z* -.22069)

AGG_MENT = (PF_Z* -.22999) + (RP_Z* -.12329) +
(BP_Z* -.09731) + (GH_Z* -.01571) + (VT_Z* .23534) +
(SF_Z* .26876) + (RE_Z* .43407) + (MH_Z* .48581)

Step 3. Formulas for T-score transformation of component scores:

Transformed Physical (PCS) = 50+ (AGG_PHYS* 10)

Transformed Mental (MCS) = 50+ (AGG_MENT* 10)

Common problems in scoring procedures:

1.1 Items with out-of-range response values: out-of-range values are those that are lower than an item's precoded minimum value or higher than an item's precoded maximum value. Any out-of-range values should be treated as missing data if they could not be corrected in the raw data.

1.2 Missing item responses: the Half-Scale Rule for missing data imputation is applied for calculating the domain scores. The scores should be missing if the respondent answers less than 50% of the items in a multi-item scale; however, if the respondent answers at least 50% of the items, the value for any missing item data should be substituted by the respondent's average final item responses value across the other completed items.

1.3 PCS and MCS missing data estimation: Under Half-Scale Rule, PCS and MCS measures are not scored if any one of the eight health domain scale scores is missing.

9.7 Association of Early Termination Assessments to Scheduled Visits

For purposes of reporting early termination assessments during the study, each early termination visit will be assigned the next nominal visit number after the last completed visit. This rule applies to both efficacy variables (e.g., HPT-SD) and safety variables (e.g., vital signs) that are analyzed and/or summarized by visit.

9.8 Analysis Software

Statistical analyses will be performed using Version 9.4 (or newer) of SAS® on a suitably qualified environment.