STATISTICAL ANALYSIS PLAN
PHASE 3

Version: 2.0

Date: *18 September, 2023*

Based on: Protocol Amendment No. 3: 10 May 2022
Protocol Amendment No. 2: 06 August 2021
Protocol Amendment No. 1: 22 December 2020
Original Protocol Final Version: 22 September 2020

Study Treatment: Crinecerfont (NBI-74788)

Study Number: NBI-74788-CAH2006

Study Title: A Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Efficacy of Crinecerfont (NBI-74788) in Pediatric Subjects with Classic Congenital Adrenal Hyperplasia, Followed by Open-Label Treatment

Study Sponsor: Neurocrine Biosciences, Inc.
12780 El Camino Real
San Diego, CA 92130
Telephone: +1-858-617-7600

This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

CONFIDENTIAL

NCT04806451

DOCUMENT HISTORY

Version Number	Date	Author(s)	Page Number	Modification(s)
1.0	08 August 2023	[redacted]	N/A	N/A
2.0	18 September 2023	[redacted]	24	The definition of the baseline and Week 4 serum androstenedione values have been updated from the average of all serial sampling values to the average of the pre-morning glucocorticoid dose values.
2.0	18 September 2023	[redacted]	24	The definition of the baseline and Week 4 serum 17-OHP values have been updated from the average of all serial sampling values to the average of the pre-morning glucocorticoid dose values.
2.0	18 September 2023	[redacted]	43	The supplemental analysis using the average of the pre-morning glucocorticoid dose values of serum androstenedione has been updated to use the average of all serial sampling values.
2.0	18 September 2023	[redacted]	45	The supplemental analysis using the average of the pre-morning glucocorticoid dose values of serum 17-OHP has been updated to use the average of all serial sampling values.

SIGNATURE PAGE

This document has been prepared by:

[redacted]	eSigned and Dated	
Senior Biostatistician Neurocrine Biosciences, Inc.	Signature	Date
[redacted]	eSigned and Dated	
Principal Biostatistician Neurocrine Biosciences, Inc.	Signature	Date

This document has been reviewed by:

[redacted]	eSigned and Dated	
Executive Director, Biostatistics Neurocrine Biosciences, Inc.	Signature	Date
[redacted] MD	eSigned and Dated	
Executive Medical Director, Clinical Development Neurocrine Biosciences, Inc.	Signature	Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Term
17-OHP	17-hydroxyprogesterone
ACTH	Adrenocorticotropic hormone
AE	Adverse event
ANCOVA	Analysis of covariance
ATC	Anatomic Therapeutic Chemical
BID	Twice daily
BLQ	Below the lower limit of quantification
BMI	Body mass index
BPRS-c	Brief psychiatric rating scale for children
BSA	Body surface area
CAH	Congenital adrenal hyperplasia
CI	Confidence interval
CMH	Cochran-Mantel-Haenszel
CSR	Clinical study report
C-SSRS	Columbia-Suicide Severity Rating Scale
DB	Double-blind
DMC	Data Monitoring Committee
ECG	Electrocardiogram
eCRF	Electronic case report form
EDC	Electronic data capture
EQ-5D-5L	EuroQoL 5 Dimensions 5 Levels
EQ-5D-Y	EuroQoL 5 Dimensions Youth
FAS	Full analysis set
FCS	Fully conditional specifications
FDA	Food and Drug Administration
FSH	Follicle stimulating hormone
GC	Glucocorticoid
GTT	Glucose tolerance test
HDL	High-density lipoprotein

Abbreviation	Term
HOMA-IR	Homeostatic model assessment of insulin resistance
HRQL	Health-related quality of life
HTN	Hypertension
ICH	International Council for Harmonization
IPD	Important protocol deviation
IRT	Interactive response technology
IUD	Intrauterine device
LDL	Low-density lipoprotein
LH	Luteinizing hormone
LLN	Lower limit of normal
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
NBI	Neurocrine Biosciences, Inc.
OL	Open-label
OLE	Open-label extension
PDHP	Protocol Deviation Handling Plan
PedsQL	Pediatric Quality of Life Instrument
PKAS	PK analysis set
PRA	Plasma renin activity
PT	Preferred Term
QD	Daily
QID	Four times per day
SAE	Serious adverse event
SAP	Statistical analysis plan
SAS	Safety analysis set
SD	Standard deviation
SDS	Standard deviation score
SE	Standard error
SEM	Standard error of the mean
SMQ	Standardized MedDRA Queries

Abbreviation	Term
SOC	System organ class
TART	Testicular adrenal rest tumor
TEAE	Treatment-emergent adverse event
TID	Three times per day
ULN	Upper limit of normal
ULQ	Upper limit of quantification
VAS	Visual analog scale
WHO	World Health Organization

1. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the planned analyses and data displays that will be prepared to summarize the data from the double-blind, placebo-controlled treatment period (Day 1 to Week 28; "DB period") and open-label treatment period (Week 28 to Week 52; "OL period") of the Phase 3 Study NBI-74788-CAH2006. Summaries of data from the open-label extension (OLE) for continued access to crinecerfont (starting after Week 52) will be described in a separate SAP.

This SAP was developed in accordance with International Council for Harmonization (ICH) E9 guidance. Data analyses described in this SAP will occur at 2 milestones in the study: after all subjects have had the opportunity to complete Week 28 and after all subjects have had the opportunity to complete Week 52. The first analysis will include the final analysis of data through Week 28 and an analysis of all available safety and efficacy data up to Week 52. The second analysis will be the final analysis of data through Week 52. Decisions regarding each set of analyses will be made prior to the corresponding database lock and treatment unblinding and will be documented. Changes to the planned analyses described in this SAP will be statistically justified and described in the clinical study report (CSR). Further information related to study design and methodology can be found in the study protocol.

2. STUDY OBJECTIVES

The objectives of this study are:

Primary:

- To evaluate the efficacy of crinecerfont, compared with placebo, in reducing adrenal steroid levels during a glucocorticoid-stable period.

Secondary:

- To evaluate the efficacy of crinecerfont, compared with placebo, in reducing daily glucocorticoid dosage while maintaining adrenal androgen control.
- To evaluate the effect of crinecerfont, compared with placebo, on clinical endpoints associated with supraphysiologic glucocorticoid dosing and androgen excess.
- To evaluate plasma concentrations of crinecerfont and metabolites.
- To assess the safety and tolerability of crinecerfont.

3. STUDY DESIGN

This is a Phase 3, randomized, double-blind, placebo-controlled study to evaluate the efficacy, safety, and tolerability of crinecerfont versus placebo administered bid with breakfast and evening meals for 28 weeks in approximately 81 pediatric subjects with classic CAH due to 21-hydroxylase deficiency. Eligible subjects will be randomly assigned in a 2:1 ratio (active:placebo) to either crinecerfont (25 mg bid via oral solution for subjects 10 to <20 kg, 50 mg bid via oral solution for subjects 20 to <55 kg, or 100 mg bid via oral capsules for subjects ≥55 kg) or matching placebo (oral solution placebo for subjects <55 kg and oral capsule placebo for subjects ≥55 kg). Dose assignment from Day 1 to Week 28 will be based on the subject's weight at Day 1. After the 28-week placebo-controlled treatment period, there will be a 24-week, open-label treatment period, during which all subjects will receive crinecerfont at doses based on their Week 28 body weight.

At Month 12 (Week 52), subjects will have the option to participate in an OLE. During the OLE, subjects will continue to receive crinecerfont at their weight-based dose, unless the subject has inadequate efficacy (in the opinion of the investigator), in which case their crinecerfont dose can be increased by doubling the evening dose. At Month 12, subjects can also elect to switch the crinecerfont formulation (oral solution or capsule) based on preference. Subjects have the option to remain in the OLE until crinecerfont becomes commercially available, the Sponsor elects to discontinue development of crinecerfont for CAH, or the Sponsor elects to discontinue the study.

3.1. Randomization

Eligible subjects will be randomized 2:1 to either crinecerfont or placebo on Day 1 using an interactive response technology (IRT). Randomization will be stratified by pubertal stage (Tanner breast or genital stage 1 or 2 versus 3, 4 or 5) and sex.

3.2. Blinding

Blinding will be maintained unless unblinding is necessary for subject safety. All subjects will be encouraged to complete follow-up even if an unblinding event has occurred. The subject, investigator, and all study center personnel will remain blinded to the subject's randomized treatment assignment(s) through database lock and unblinding of the Week 52 data. The Sponsor will remain blinded until all subjects complete the Week 28 visit and the database has been locked, at which time a limited number of Sponsor personnel will be unblinded to individual treatment assignments. Following this analysis, the subject, investigator, all study center personnel, and Sponsor personnel with direct contact with the site will continue to be blinded to the subject's blinded treatment assignment through database lock and unblinding of the Week 52 data.

An independent DMC will periodically review ongoing unblinded clinical and safety data to ensure the safety and well-being of the study subjects.

3.3. Sample Size Considerations

The sample size of 81 subjects (54 in the crinecerfont treatment group and 27 in the placebo group) is based on a power calculation for the primary endpoint (change from baseline in serum androstenedione at Week 4). In the Phase 2 Study NBI-74788-CAH2001, the mean change from

baseline to Day 14 in androstenedione was -286.2 ng/dL with a standard deviation of 345.02 (pooled across dosing cohorts). Based on a 2-sample t-test with an effect size of 0.83, a sample size of 81 subjects will have greater than 90% power to detect a treatment difference at a 0.05 level of significance. With an effect size as small as 0.70, a sample size of 81 subjects will have greater than 80% power to detect a treatment difference at the same level of significance.

4. ENDPOINTS

The efficacy, safety, pharmacokinetic, patient and caregiver reported outcome, and other endpoints are listed below.

4.1. Primary Efficacy – DB Period

The primary efficacy endpoint in the DB period is the change from baseline in serum androstenedione at Week 4.

4.2. Key Secondary Efficacy – DB Period

Key secondary efficacy endpoints in the DB period are:

- Change from baseline in serum 17-hydroxyprogesterone (17-OHP) at Week 4
- Percent change from baseline in glucocorticoid daily dose (in hydrocortisone equivalents adjusted for BSA [mg/m^2/day]) at Week 28, while Week 28 serum androstenedione is ≤120% of the baseline value or ≤ ULN, according to sex and either age (for Tanner stage 1) or pubertal stage (for Tanner stages 2 to 5)

4.3. Secondary Efficacy – DB Period

Secondary efficacy endpoints in the DB period are:

- The achievement of a reduction in glucocorticoid daily dose to physiologic levels (≤11 mg/m^2/day in hydrocortisone dose equivalents adjusted for BSA) at Week 28 while serum androstenedione is ≤120% of the baseline value or ≤ULN for sex and either age (for Tanner stage 1) or pubertal stage (for Tanner stages 2 to 5)
- Change from baseline in body mass index SDS (standard deviation score) at Week 28
- Change from baseline in mean 24-hour salivary 17-OHP at Week 28

In the subgroup of subjects not at adult height:

- Change from baseline in the ratio of bone age to chronological age (BA:CA) at Week 28

4.4. Exploratory Efficacy – DB Period

Exploratory efficacy endpoints in the DB period include:

- Change from baseline in body weight SDS at Week 28
- Change from baseline in all other serum hormone parameters (adrenocorticotropic hormone [ACTH; plasma], testosterone, follicle stimulating hormone [FSH], luteinizing hormone [LH], and cortisol) at Week 28
- Change from baseline in all other salivary hormone parameters (androstenedione, testosterone, cortisol, progesterone, pregnenolone) at Week 28
- Change from baseline in metabolic parameters (total cholesterol, low-density lipoprotein [LDL], high-density lipoprotein [HDL], triglycerides, and HOMA-IR) at Week 28

- Change from baseline in testicular adrenal rest tumor (TART) volume (expressed as a percentage of the total testicular volume) at Week 28 (male subjects with TARTs at baseline only)
- Change from baseline in Bayley-Pinneau predicted adult height SDS at Week 28 (in the subgroup of subjects not at adult height)
- Change from baseline in height SDS at Week 28 (in subjects not at adult height)
- Change from baseline in hirsutism at Week 28 (female subjects only)
- Change from baseline in acne at Week 28
- The achievement of menstrual regularity (in female subjects who have undergone menarche and are not on hormonal or IUD contraceptives) at Week 28
- Achievement of a reduction in glucocorticoid dosing regimen instances per day at Week 28 (subjects on a hydrocortisone-alone regimen at baseline)
- Achievement of switching to a hydrocortisone-alone glucocorticoid dose regimen at Week 28 (subjects on prednisone [or equivalent] or a hydrocortisone plus prednisone [or equivalent] regimen at baseline)

4.5. Efficacy – OL Period

A secondary efficacy endpoint in the OL period is:

- Change from baseline in Bayley-Pinneau predicted adult height SDS at Week 52 (in the subgroup of subjects not at adult height)

Exploratory efficacy endpoints in the OL period include:

- Change from baseline and OL baseline in serum hormone parameters (ACTH [plasma], 17-OHP, androstenedione, testosterone, FSH, LH, and cortisol) at Week 52
- Change from baseline and OL baseline in salivary hormone parameters (17-OHP, androstenedione, testosterone, progesterone, pregnenolone, and cortisol) at Week 52
- Percent change from baseline and OL baseline in glucocorticoid daily dose (in hydrocortisone equivalents adjusted for BSA [mg/m^2/day]) at Week 52, while Week 52 serum androstenedione is ≤120% of the baseline value or ≤ ULN, according to sex and either age (for Tanner stage 1) or pubertal stage (for Tanner stages 2 to 5)
- Achievement of a reduction in glucocorticoid daily dose to physiologic levels at Week 52
- Change from baseline and OL baseline in metabolic parameters (total cholesterol, LDL, HDL, triglycerides, and HOMA-IR) at Week 52
- Change from baseline and OL baseline in body weight SDS at Week 52
- Change from baseline and OL baseline in BMI SDS at Week 52
- Change from baseline and OL baseline in height SDS at Week 52 (in the subgroup of subjects not at adult height)

- Change from baseline and OL baseline in BA:CA at Week 52 (in the subgroup of subjects not at adult height)
- Change from OL baseline in height velocity at Week 52 (in the subgroup of subjects not at adult height)
- Change from baseline and OL baseline in hirsutism at Week 52 (female subjects only)
- Change from baseline and OL baseline in acne at Week 52
- Change from baseline in testicular adrenal rest tumor (TART) volume (expressed as a percentage of the total testicular volume) at Week 52 (male subjects with TARTs at baseline only)
- The achievement of menstrual regularity at Week 52 (in female subjects who have undergone menarche and are not on hormonal or IUD contraceptives)
- Change in menstrual regularity at Week 28 to Week 52, compared to Day 1 to Week 28 (in female subjects who have undergone menarche and are not on hormonal or IUD contraceptives and are assigned to the placebo group for the DB period)

4.6. Safety

Safety endpoints in the DB and OL periods include:

- Occurrence of treatment-emergent adverse events (TEAEs), exposure-adjusted subject incidence rate, serious adverse events (SAEs), severe TEAEs, adrenal insufficiency TEAEs, TEAEs leading to glucocorticoid stress dosing, and TEAEs leading to discontinuation from study drug
- Observed and changes from baseline in clinical laboratory test values (including hematology, clinical chemistry, thyroid stimulating hormone [TSH], free T4, plasma renin activity [PRA], and coagulation)
- Observed and changes from baseline in vital sign values
- Observed and changes from baseline in 6- or 12-lead electrocardiogram (ECG) intervals
- Observed and changes from baseline in total scores from the Brief Psychiatric Rating Scale for Children (BPRS-c)
- "Yes" responses to suicidal behavior or ideation at postbaseline assessments from the Columbia-Suicide Severity Rating Scale (C-SSRS; only for subjects ≥6 years of age)

4.7. Pharmacokinetics

The pharmacokinetic endpoint is observed plasma concentrations of crinecerfont and its metabolites

4.8. Patient and Caregiver Reported Outcomes

Patient and caregiver reported outcome endpoints in the DB and OL periods include:

- The 5 levels for each health dimension from the EQ-5D-Y (subjects 8 to 15 years of age) and EQ-5D-5L (subjects ≥ 16 years of age) over time
- Observed and changes from baseline in EQ-Visual Analog Scale (VAS) score
- Observed and changes from baseline in the Pediatric Quality of Life Instrument (PedsQL) psychosocial health summary score, the physical health summary score, and the total score
- Observed and changes from baseline in the PedsQL (family impact module) parent health-related quality of life (HRQL) summary score, the family functioning summary score, and the total score

4.9. Other

A secondary endpoint in the DB period is:

- Acceptability and palatability of the study drug at Week 4

Other endpoints in the DB and OL periods include:

- Acceptability and palatability of the study drug at Week 28
- Acceptability and palatability of the study drug at Week 52

5. DATA ANALYSES

5.1. Week 28 (DB Period) Final Analysis

The final unblinded analysis of the double-blind, placebo-controlled period will be conducted once all subjects have had the opportunity to complete the Week 28 visit. Data through Week 28, including the primary, key secondary, secondary, and exploratory endpoints occurring on or before Week 28, will be analyzed. At this time, an analysis of all available safety and efficacy data through Week 52 will also be performed. Following this analysis, the subject, investigator, all study center personnel, and Sponsor personnel with direct contact with the site will continue to be blinded to the subject's blinded treatment assignment. All planned analyses for the final Week 28 analysis are included in this SAP.

5.2. Week 52 (OL Period) Final Analysis

The final analysis of the open-label period will be conducted once all subjects have had the opportunity to complete the Week 52 visit. At this time, all data collected up to Week 52 will be analyzed. All planned analyses for the final Week 52 analysis are included in this SAP.

6. GENERAL CONSIDERATIONS FOR DATA ANALYSES AND HANDLING

6.1. General Statistical Procedures

All analyses described in this plan are considered a priori analyses in that they have been defined prior to locking the study database for the Week 28 analysis. Analyses defined subsequent to locking the database will be considered post hoc analyses and will be applied as exploratory methodology. Any post hoc analyses will be statistically justified and described as appropriate in the CSR. Statistical analysis will be conducted, and all tables, figures, and listings generated using SAS® software (version 9.4 or later), unless stated otherwise.

Unless otherwise noted, data from the double-blind, placebo-controlled period (Day 1 to Week 28; "DB period") will be summarized by randomized treatment group, defined as crinecerfont or placebo (DB treatment group). Data from the open-label period (Week 28 to Week 52; "OL period") will be summarized by OL treatment group, defined as placebo/crinecerfont or crinecerfont/crinecerfont to describe the subject's randomized treatment assignment and their OL treatment, as well as all subjects combined. Note that while crinecerfont (and matching placebo) dose levels (25 mg bid via oral solution for subjects 10 to <20 kg, 50 mg bid via oral solution for subjects 20 to <55 kg, or 100 mg bid via oral capsules for subjects ≥55 kg) are assigned based on weight at Day 1 and again at Week 28, summaries of the DB and OL period data will not be broken out by dose level unless otherwise specified.

Descriptive and inferential statistical methods will be used to evaluate and summarize the data from this study. The term "descriptive statistics" refers to the number of subjects (n), mean, median, standard deviation (SD) or standard error (SE), minimum, and maximum for continuous variables. Ordinal categorical data will be summarized using median, minimum and maximum values. Number and percentage of subjects will be summarized for categorical variables. The term "inferential statistics" refers to hypothesis tests that will be performed to assess differences between the treatment group and the control group. All hypothesis tests will be tests of the null hypothesis of no difference between the groups being compared versus the two-sided alternative hypothesis that there is a difference. The level of significance (type I error) for declaring statistical significance will be 0.05.

Summary statistics will be presented using the following decimal precision (ie, number of digits to the right of the decimal point): the minimum and maximum will have the same number of decimal places as the data; the mean, median, SD and SE will have one more decimal place than the data being summarized; the sample size (N) will be reported as an integer; with the exception of 100 percent, percentages will be reported to one decimal place (percentages for zero counts are omitted); and p-values will be displayed using four decimal places. Confidence intervals for means will be reported to the same number of decimal places as mean values; and confidence intervals for percentages will be reported to one decimal place. These rules may be modified if warranted, based on practical considerations.

All available study data will be included in relevant data displays, including data for subjects with incomplete or missing values. Replacement of missing data values with imputed values will generally not be performed unless specified otherwise in relevant endpoint subsections.

6.2. Analysis Sets

Table 1: Definitions of Analysis Sets

Analysis Set	Description
Full Analysis Set	The full analysis set (FAS) will include all randomized subjects. Subjects will be analyzed according to their randomized treatment group, regardless of adherence to study drug administration.
DB Period Safety Analysis Set	The DB period safety analysis set (SAS) will include all randomized subjects who take at least 1 dose of study drug in the double-blind period. Subjects will be analyzed according to their randomized treatment group, unless they receive the incorrect study drug for the entire double-blind treatment duration.
OL Period Safety Analysis Set	The OL period safety analysis set (SAS) will include all subjects who take at least 1 dose of active study drug (crinecerfont) during the open-label period.
PK Analysis Set	The PK analysis set (PKAS) will include all subjects randomized to crinecerfont who take at least 1 dose of study drug and have at least one post-dose PK concentration.

6.2.1. Application of Analysis Sets

Summaries of subject disposition, analysis set inclusion/exclusion status, and protocol deviations will include all randomized subjects for the DB period. Summaries of subject disposition and protocol deviations in the OL period will use the OL period safety analysis set. The summary of analysis set inclusion/exclusion status in the OL period will include all randomized subjects. Analysis sets used for all other summaries are defined at the beginning of each major section.

6.3. Baseline Definitions

The following sections describe the baseline definitions used in the DB and OL periods, as well as baseline definitions specifically defined for the primary and key secondary efficacy endpoints. For each subject, study day is calculated relative to Day 1, where Day 1 is defined as the date of first dose of study drug. For subjects who were randomized but never received study drug, Day 1 is defined as the date of randomization.

6.3.1. Baseline Definition for Analyses of DB Period Data

Other than the definitions described in Sections 6.3.3 through 6.3.5, the assessments collected on Day 1 prior to the first dose of study drug in the DB period will serve as the baseline value for all assessments through Week 28 of the study. If a Day 1 baseline value is not available, then the last measurement collected prior to study drug will serve as baseline. Bone age and testicular ultrasound assessments obtained within 1 month of Day 1 may serve as the baseline value if needed, as these parameters are not expected to change to a degree that could be clinically

detected over this time frame. This is considered the study baseline and will be referred to as "baseline" throughout this document.

Note that for serum hormone parameters that are collected both pre- and post-glucocorticoid (GC) dose, two baselines will be derived for all postbaseline pre-GC dose and visit average values where a change from baseline calculation is necessary, unless otherwise specified. Baseline pre-GC dose will be defined as the average of the pre-GC dose values (when collected more than once) or the single pre-GC dose value collected prior to study drug dosing. Baseline average will be defined as the average of serial sampling values collected prior to study drug dosing. If the Day 1 values are missing, the last measurement collected prior to study drug dosing will only be used as baseline average only.

For salivary hormone parameters with serial sampling, three baselines will be derived for all pre-dose, mean morning, and mean 24-hour values where a change from baseline calculation is necessary, unless otherwise specified. Baseline pre-dose will be defined as the average of the 0600 and 0800 timepoint values at the baseline visit. Baseline mean morning will be defined as the average of the 0600, 0800, +3 hours post-dose, and +6 hours post-dose timepoint values at the baseline visit. Baseline mean 24-hour will be defined as the average of all timepoints at the baseline visit.

6.3.2. Baseline Definition for Analyses of OL Period Data

Two baselines will be defined for analyses of data from the OL period. For all efficacy summaries in the OL period, the Week 28 assessments collected prior to the first dose of study drug in the OL period will serve as one of the baseline values for assessments through Week 52 of the study. This baseline will be referred to as the "OL baseline" throughout this document. All efficacy summaries will be repeated using the study baseline (as defined in Section 6.3.1) unless otherwise specified. For both the OL baseline and study baseline, if the respective Week 28 or Day 1 value is not available, then the last measurement collected prior to study drug in each period will serve as baseline. Specifically, for the OL period, if the Week 28 value is not available, then the last measurement collected prior to study drug in the OL period will serve as baseline as long as the last non-missing measurement is not the study baseline value. For assessments that are performed less frequently (ie, assessments that are only collected at Day 1 and Week 28 within the DB period), if the Week 28 assessment is not available then the OL baseline and postbaseline values in the OL period will not be included in the summary.

Bone age and testicular ultrasound assessments obtained within 1 month of Week 28 may serve as the OL baseline value if needed, as these parameters are not expected to change to a degree that could be clinically detected over this time frame.

Note that for serum hormone parameters that are collected both pre- and post-GC dose, two OL baselines will be derived for all postbaseline pre-GC dose and visit average values (after Week 28) where a change from baseline calculation is necessary, unless otherwise specified. OL baseline pre-GC dose will be defined as the pre-GC dose value collected prior to study drug dosing at Week 28. OL baseline average will be defined as the average of the Week 28 values collected prior to study drug dosing. If the Week 28 values are missing, the last measurement collected prior to study drug dosing will only be used as baseline average only.

For salivary hormone parameters with serial sampling, two OL baselines will be derived for all pre-dose and mean 24-hour values (after Week 28) where a change from OL baseline calculation is necessary, unless otherwise specified. OL baseline pre-dose will be defined as the average of the 0600 and 0800 timepoint values at the OL baseline visit. OL baseline mean 24-hour will be defined as the average of all timepoints at the OL baseline visit.

6.3.3. Baseline Definition of Androstenedione for Primary Endpoint

For subjects ≥6 years of age with a body weight ≥20 kg, serum androstenedione at Day 1 and Week 4 was collected 15 minutes before and just prior to dosing with the morning glucocorticoid dose and study drug (Week 4) and 2, 3, 4, and 6 hours after dosing. For subjects who are <6 years of age or with a body weight <20 kg, serum androstenedione at Day 1 and Week 4 was collected prior to the morning dose of glucocorticoid and study drug (at Week 4). Both baseline and Week 4 serum androstenedione values for subjects ≥6 years of age with a body weight ≥20 kg will be calculated as the average of all pre-morning glucocorticoid dose serum androstenedione values on Day 1 and Week 4, respectively. Subjects <6 years of age or <20 kg will have only a single pre-morning glucocorticoid dose serum androstenedione value for both baseline and Week 4. If a subject is missing all pre-morning glucocorticoid dose serum androstenedione values at Day 1, the serum androstenedione value collected at the screening visit will be used instead.

6.3.4. Baseline Definition of Serum 17-OHP for First Key Secondary Endpoint

For subjects ≥6 years of age with a body weight ≥20 kg, serum 17-OHP at Day 1 and Week 4 was collected 15 minutes before and just prior to dosing with the morning glucocorticoid dose and study drug (Week 4) and 2, 3, 4, and 6 hours after dosing. For subjects who are <6 years of age or with a body weight <20 kg, serum 17-OHP at Day 1 and Week 4 was collected prior to the morning dose of glucocorticoid and study drug (at Week 4). Both baseline and Week 4 serum 17-OHP values for subjects ≥6 years of age with a body weight ≥20 kg will be calculated as the average of all pre-morning glucocorticoid dose serum 17-OHP values on Day 1 and Week 4, respectively. Subjects <6 years of age or <20 kg will have only a single pre-morning glucocorticoid dose serum 17-OHP value for both baseline and Week 4. If a subject is missing all pre-morning glucocorticoid dose serum 17-OHP values at Day 1, the serum 17-OHP value collected at the screening visit will be used instead.

6.3.5. Baseline Definition of Serum Androstenedione "Control" for Second Key Secondary Endpoint

The second key secondary endpoint is the percent change from baseline in glucocorticoid daily dose (in hydrocortisone dose equivalents adjusted for BSA [mg/m^2/day]) at Week 28, while androstenedione is in "control". Androstenedione is considered in "control" if the Week 28 serum androstenedione is ≤120% of the baseline value or ≤ ULN, according to sex and either age (for Tanner stage 1) or pubertal stage (for Tanner stages 2 to 5). For all subjects, baseline serum androstenedione used to define "control" will be defined as the average of all pre-morning glucocorticoid serum androstenedione values on Day 1. Subjects who are <6 years of age or <20 kg will have only a single pre-morning glucocorticoid serum androstenedione value for baseline and Week 28. If values on Day 1 are missing, the last pre-morning glucocorticoid serum androstenedione value prior to Day 1 will serve as the baseline. The pre-morning glucocorticoid

dose androstenedione value at Week 28 will be used for the comparison to the baseline androstenedione in assessing control (as defined above). If the pre-morning glucocorticoid dose androstenedione value at Week 28 is not available, the post-morning glucocorticoid dose value will be used. The missing data handling approach is described in Section 9.4.2.

6.4. Completers, Retrieved Dropouts, Non-Retrieved Dropouts, and On Drug-Missing Endpoint Subjects

The methods of multiply imputing missing data require classifying subjects based on having evaluable measurements and their treatment status at the time of the endpoint (namely, Week 4 or Week 28). For the purpose of missing data imputation, the following types of subjects are defined in the table below.

Table 2: Evaluability of Subjects at Endpoint

Subject Classification	Description
Completers	Subjects in the FAS who completed through the time of the endpoint (Week 4 or Week 28, as applicable) on study drug and have an evaluable measurement for that endpoint. For endpoints with more than one data component, the subject must have complete data for all components of the endpoint.
Retrieved Dropouts	Subjects in the FAS who discontinue study drug prior to the time of the endpoint collection but remain in the study and have an evaluable measurement for that endpoint. For endpoints with more than one data component, the subject must have complete data for all components of the endpoint.
Non-retrieved Dropouts	Non-retrieved dropouts are subjects in the FAS who discontinue from the study and from study drug and do not have an evaluable measurement for that endpoint.
On Drug-Missing Endpoint	Subjects in the FAS who are on study drug at the time of the endpoint collection but do not have an evaluable measurement for that endpoint.

6.5. Hormone Reference Ranges

The appropriate reference ranges for androstenedione will be assigned based on reference ranges provided in the lab manual for either sex and age (for Tanner stage 1) or sex and pubertal stage (for Tanner stages 2 to 5). These ranges are shown in Table 3. Reference ranges are also given by age for 17-OHP and by sex and age for testosterone.

The assignment of these hormone reference ranges will apply to all summaries and analyses that use reference ranges. DB period analyses will use the Tanner stage documented in the eCRF at Day 1. OL period analyses will use the Tanner stage documented in the eCRF at Week 28 or, if unavailable, the last tanner stage available prior to dosing in the OL period.

Table 3: Hormone Reference Ranges

Lab Analyte	**Reference Ranges:** conventional units; SI units
Serum 17-OHP (conversion factor from ng/dL to nmol/L: x 0.0303)	1-11 months: ≤ 147 ng/dL; ≤ 4.4541 nmol/L 1 year: ≤ 139 ng/dL; ≤ 4.2117 nmol/L 2 years: ≤ 134 ng/dL; ≤ 4.0602 nmol/L 3 years: ≤ 131 ng/dL; ≤ 3.9693 nmol/L 4 years: ≤ 131 ng/dL; ≤ 3.9693 nmol/L 5 years: ≤ 133 ng/dL; ≤ 4.0299 nmol/L 6 years: ≤ 137 ng/dL; ≤ 4.1511 nmol/L 7 years: ≤ 145 ng/dL; ≤ 4.3935 nmol/L 8 years: ≤ 154 ng/dL; ≤ 4.6662 nmol/L 9 years: ≤ 166 ng/dL; ≤ 5.0298 nmol/L 10 years: ≤ 180 ng/dL; ≤ 5.454 nmol/L 11 years: ≤ 196 ng/dL; ≤ 5.9388 nmol/L 12 years: ≤ 213 ng/dL; ≤ 6.4539 nmol/L 13 years: ≤ 233 ng/dL; ≤ 7.0599 nmol/L 14 years: ≤ 254 ng/dL; ≤ 7.6962 nmol/L 15 years: 19-276 ng/dL; 0.5757-8.3628 nmol/L 16 years: 23-300 ng/dL; 0.6969-9.09 nmol/L 17 years: 26-325 ng/dL; 0.7878-9.8475 nmol/L
Serum Androstenedione (conversion factor from ng/dL to nmol/L: x 0.0349)	Females: 1-11 months: ≤ 41 ng/dL; ≤ 1.4309 nmol/L 1 year: ≤ 35 ng/dL; ≤ 1.2215 nmol/L 2 years: ≤ 34 ng/dL; ≤ 1.1866 nmol/L 3 years: ≤ 38 ng/dL; ≤ 1.3262 nmol/L 4 years: ≤ 42 ng/dL; ≤ 1.4658 nmol/L 5 years: ≤ 45 ng/dL; ≤ 1.5705 nmol/L 6 years: ≤ 45 ng/dL; ≤ 1.5705 nmol/L 7 years: ≤ 48 ng/dL; ≤ 1.6752 nmol/L 8 years: ≤ 57 ng/dL; ≤ 1.9893 nmol/L 9 years: ≤ 77 ng/dL; ≤ 2.6873 nmol/L

Lab Analyte	**Reference Ranges:** conventional units; SI units
	10 years: 15-111 ng/dL; 0.5235-3.8739 nmol/L 11 years: 24-149 ng/dL; 0.8376-5.2001 nmol/L 12 years: 32-182 ng/dL; 1.1168-6.3518 nmol/L 13 years: 37-205 ng/dL; 1.2913-7.1545 nmol/L 14 years: 42-221 ng/dL; 1.4658-7.7129 nmol/L 15 years: 46-238 ng/dL; 1.605-8.3062 nmol/L 16 years: 50-252 ng/dL; 1.745-8.7948 nmol/L 17 years: 53-265 ng/dL; 1.8497-9.2485 nmol/L Tanner Stages II-III: 43-180 ng/dL; 1.5007-6.282 nmol/L Tanner Stages IV-V: 73-220 ng/dL; 2.5477-7.6678 nmol/L Males: 1-11 months: ≤ 41 ng/dL; ≤ 1.4309 nmol/L 1 year: ≤ 31 ng/dL; ≤ 1.0819 nmol/L 2 years: ≤ 27 ng/dL; ≤ 0.9423 nmol/L 3 years: ≤ 27 ng/dL; ≤ 0.9423 nmol/L 4 years: ≤ 29 ng/dL; ≤ 1.0121 nmol/L 5 years: ≤ 31 ng/dL; ≤ 1.0819 nmol/L 6 years: ≤ 36 ng/dL; ≤ 1.2564 nmol/L 7 years: ≤ 43 ng/dL; ≤ 1.5007 nmol/L 8 years: ≤ 54 ng/dL; ≤ 1.8846 nmol/L 9 years: ≤ 65 ng/dL; ≤ 2.2685 nmol/L 10 years: 10-77 ng/dL; 0.349-2.6873 nmol/L 11 years: 12-90 ng/dL; 0.4188-3.141 nmol/L 12 years: 14-106 ng/dL; 0.4886-3.6994 nmol/L 13 years: 17-124 ng/dL; 0.5933-4.3276 nmol/L 14 years: 19-139 ng/dL; 0.6631-4.8511 nmol/L 15 years: 21-154 ng/dL; 0.7329-5.3746 nmol/L 16 years: 24-172 ng/dL; 0.8376-6.0028 nmol/L 17 years: 27-192 ng/dL; 0.9423-6.7008 nmol/L Tanner Stages II-III: 17-82 ng/dL; 0.5933-2.8618 nmol/L Tanner Stages IV-V: 57-150 ng/dL; 1.9893-5.235 nmol/L

Lab Analyte	**Reference Ranges:** conventional units; SI units
Serum Testosterone (conversion factor from ng/dL to nmol/L: x 0.0347)	Females: 1-<6 years: ≤ 8 ng/dL; ≤ 0.2776 nmol/L 6-<8 years: ≤ 20 ng/dL; ≤ 0.694 nmol/L 8-<11 years: ≤ 35 ng/dL; ≤ 1.2145 nmol/L 11-<12 years: ≤ 40 ng/dL; ≤ 1.388 nmol/L 12- <14 years: ≤ 40 ng/dL; ≤ 1.388 nmol/L 14-17 years: ≤ 40 ng/dL; ≤ 1.388 nmol/L Males: 1-<6 years: ≤ 5 ng/dL; ≤ 0.1735 nmol/L 6-<8 years: ≤ 25 ng/dL; ≤ 0.8675 nmol/L 8-<11 years: ≤42 ng/dL; ≤ 1.4574 nmol/L 11-<12 years: ≤ 260 ng/dL; ≤ 9.022 nmol/L 12- <14 years: ≤ 420 ng/dL; ≤ 14.574 nmol/L 14-17 years: ≤ 1000 ng/dL; ≤ 34.7 nmol/L

6.6. Derived and Transformed Data

Change from baseline is calculated as the postbaseline value minus the baseline value; a negative value will represent a decrease at the postbaseline visit. Percent change from baseline is calculated as: change from baseline/baseline value × 100. If either the baseline or postbaseline value is missing, the change from baseline and/or percent change from baseline will also be missing. The percent change from baseline will also be missing if the baseline value is equal to zero.

For the purpose of analysis, glucocorticoids other than hydrocortisone (ie, methylprednisolone, prednisolone, and prednisone) will be converted to hydrocortisone equivalents using an equivalency ratio of 4 (eg, 1 mg prednisone = 4 mg hydrocortisone). While dexamethasone is excluded and thus subjects are not expected to be taking dexamethasone at the time of randomization, an equivalency ratio of 60 will be used to convert dexamethasone to hydrocortisone equivalents (eg, 1 mg dexamethasone = 60 mg hydrocortisone) if dexamethasone is added to the subject's glucocorticoid regimen during the study.

With the exception of pharmacokinetic concentration data, all below the lower limit of quantification (BLQ) values will be set equal to the BLQ value in all summaries and all above the upper limit of quantification (ULQ) values will be set equal to the ULQ value in all summaries. For pharmacokinetic concentration data, all BLQ values (crinecerfont < 5.00 ng/mL; metabolites < 0.500 ng/mL) will be set equal to zero (0) in the plasma concentration summaries.

6.7. Study Day

For each subject, study day is calculated relative to the date of Day 1, where Day 1 is defined as the date of the first dose of study drug (or for subjects who were randomized but never received study drug, the date of randomization). If the date of interest occurs on or after Day 1, then the study day will be calculated as: date of interest – date of Day 1 + 1. If the date of interest occurs prior to Day 1, then the study day will be calculated as: date of interest – date of Day 1.

The nominal visit number for each visit, including scheduled, unscheduled, repeat, and early termination/end of study visits, will be re-mapped to an analysis visit according to Table 4 or Table 5. If multiple measurements occur within the same visit window after mapping, the measurement that is closest to the target study day will be used for the summary tables where one observation per visit is needed, unless otherwise specified. Where there are ties between the earlier and later observation within the visit window, the earlier observation will be used.

The second key secondary endpoint of percent change from baseline in glucocorticoid daily dose at Week 28 is based on the glucocorticoid concomitant medications eCRF. The method of collecting concomitant medication data involves entering a start date and stop date for each medication, unless it is ongoing, in which case the "Ongoing?" field is marked as "yes." This differs from assessments that are collected discretely at visits. Additionally, the androstenedione values collected at the Week 28 visit are a critical component to the assessment of this endpoint as subjects who do not maintain androstenedione control and have a decrease from baseline in glucocorticoid total daily dose will have their percent change from baseline value in glucocorticoid total daily dose set to 0. For the purposes of determining the glucocorticoid total daily dose from the eCRF at Week 28, the Week 28 visit date at which the androstenedione values were collected (based on the analysis visit mapping described below) will be used to determine the glucocorticoid dose (as of the same date) to be used for the second key secondary endpoint. If a subject has a Week 28 visit performed based on the analysis visit window but androstenedione was not collected at this visit (ie, missed hormone labs), then the analysis visit date from the collection of the safety labs will be used instead for determining the Week 28 glucocorticoid total daily dose. If both hormone and safety labs were missed at the visit, then the vital signs assessment and its respective Week 28 analysis visit date will be used for determining the glucocorticoid total daily dose. The same methodology described above will be used to determine the glucocorticoid total daily dose from the eCRF at each of the other intermediate visits.

Table 4 includes the analysis visit windows for all assessments, with the exception of those assessments that have a broader window due to more infrequent assessments (ie, every 6-7 months). These more-infrequent assessments include measurements of bone age (via x-ray) and TARTs (via testicular ultrasound). These assessments will use the analysis visit windows defined in Table 6. Table 5 includes the analysis visit windows for subjects who discontinue from study drug prior to OL period or OLE and therefore do not have a first dose date in the OL period or OLE.

Table 4: Primary Analysis Visit Windows

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
Week 4	28	2 to 43
Week 8	56	44 to 70
Week 12	84	71 to 98
Week 16	112	99 to 126
Week 20	140	127 to 168
Week 28	196	169 to (first dose date in OL period)
Week 32	224	(First dose date in OL +1) to 238
Week 36	252	239 to 266
Week 40	280	267 to 294
Week 44	308	295 to 336
Week 52	364	337 to (first dose date in OLE or Week 56)

Table 5: Primary Analysis Visit Windows for Subjects Who Discontinue Study Drug Prior to Week 28 or Prior to Week 52

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
Week 4	28	2 to 43
Week 8	56	44 to 70
Week 12	84	71 to 98
Week 16	112	99 to 126
Week 20	140	127 to 168
Week 28	196	169 to 210
Week 32	224	211 to 238
Week 36	252	239 to 266
Week 40	280	267 to 294
Week 44	308	295 to 336
Week 52	364	337 to 378

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
DB Safety Follow-up*[a]	Study Day of last dose of study drug in DB period + 28 Days	Study Day of last dose of study drug in DB period + 14 days to date of last dose of study drug in DB period + 42 days
OL Safety Follow-up*[b]	Study Day of last dose of study drug in OL period + 28 Days	Study Day of last dose of study drug in OL period + 14 days to date of last dose of study drug in OL period + 42 days

*Note that only safety labs, vital signs, and ECG assessments will be mapped to the "DB Safety Follow-up" and "OL Safety Follow-up" visits. Visits that qualify for DB Safety Follow-up or OL Safety Follow-up will take precedence over any other visit window that the assessment qualifies for.

[a]Subjects who discontinue study and study drug in the DB period and have a visit in the study day range specified will have their safety data mapped to the "DB Safety Follow-up."

[b]Subjects who discontinue study and study drug in the OL period and have a visit in the study day range specified will have their safety data mapped to the "OL Safety Follow-up."

Bone age and TART analysis visit windows are presented in Table 6. These assessments have a wider visit window (including a qualifying baseline assessment) as these parameters are not expected to change to a degree that could be clinically detected over this time frame.

Table 6: Analysis Visit Windows for Bone Age and TART

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
Week 28	196	169 to (first dose date in OL period + 28)
Week 52	364	337 to (first dose date in OLE period + 28)

6.8. Handling of Missing Data

Missing data will be minimized by implementing programming checks to identify and query missing values prior to database lock.

6.8.1. Start Dates for Adverse Events

For the purpose of estimating the time of the event in relationship to study drug (ie, study day), missing and incomplete ("partial") dates for AEs will be imputed using the following algorithm. For start dates that are missing day and month: impute 1st January, unless the year is the same as the first dose of study drug, in which case the start date will be imputed to the date of first dose of study drug. For start dates that are missing day alone: impute to the 1st of the month unless the

month and year are the same as the first dose of study drug, in which case the start date will be imputed to the date of first dose of study drug. There will be no imputation for AE stop dates.

6.8.2. Start and Stop Dates for Concomitant Medications

To handle missing/partial dates for concomitant medications, the following algorithm will be employed to derive the time of medication usage relative to study drug (ie, study day). For start dates that are missing day and month: impute 1st January, unless the year is the same as the first dose of study drug, in which case the start date will be imputed to the date of first dose of study drug. For start dates that are missing day alone: impute 1st of the month unless the month and year are the same as the first dose of study drug, in which case the start date will be imputed to the date of first dose of study drug. For stop dates that are missing day and month: impute 31st December, unless the year is the same as the last dose of study drug, in which case impute to the date of last dose of study drug. For stop dates that are missing day alone: impute to the last day of the month unless the month and year is the same as the last dose of study drug, in which case impute to the date of last dose of study drug. If any of the above imputations result in a start date that is later than an observed (not imputed) medication stop date, the start date will be imputed as the stop date. Likewise, if an imputed stop date is earlier than an observed (not imputed) start date, then the stop date will be imputed as the start date. If a medication is marked as "ongoing" at the time of the subject's discontinuation from study, or if the subject is lost to follow-up, the stop date for the medication will be imputed to the subject's end of study date.

6.8.3. First and Last Study Drug Dose Dates

Missing and incomplete ("partial") dates for first and last study drug dose dates will be imputed for the purpose of estimating exposure and defining treatment periods. Missing dates will not be imputed for subjects when the subject is known to have not taken at least one dose of study drug, as documented by the site in the study drug dosing electronic case report form (eCRF).

The imputation rules for first dose date are as follows:

- If the date is completely missing or if both the day and month are missing, the date will be imputed as the randomization date
- If only the day is missing, the date will be imputed as the randomization date if the month and year match the month and year of the randomization date; if the month or year occur after the randomization date, the missing day will be imputed as the first day of the documented month.

If the date of the last dose of study drug is missing, then the last dose date will be imputed as the earliest of:

- The end of treatment date
- The last visit prior to study discontinuation.

6.9. Coding Dictionaries

Medical history and AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA, version 26.0). Prior and concomitant medications will be coded using the World Health Organization (WHO) Drug Dictionary (March 2022 B3 Global version).

6.10. Impact of COVID-19 Pandemic

This section describes analyses and summaries that will be produced to help determine the potential impact of the COVID-19 pandemic on the study conduct/data and additional details regarding how data that is potentially impacted by the COVID-19 pandemic will be handled in the analysis plan. It is in alignment with the guidance put forth by the US Food and Drug Administration (FDA; Conduct of Clinical Trials of Medical Products During the COVID-19 Public Health Emergency [March 2020, updated January 2021]) and European Medicines Agency (EMA; Points to consider on implications of Coronavirus disease [COVID-19] on methodological aspects of ongoing clinical trials [March 2020]).

To help understand the impact of the COVID-19 pandemic on the clinical trial data, the following listings will be generated:

- A listing of all subjects affected by the COVID-19 pandemic. The listing will identify subjects who experience at least one of the following situations due to the COVID-19 pandemic (additional situations may be included):
    - Discontinued study drug or withdrew from study
    - Presumed or confirmed diagnosis of COVID-19
    - Had at least one COVID-19 pandemic-related major protocol deviation
    - Missed at least one study visit or assessment
    - Required at least one assessment to be collected using a method other than that defined in the protocol (eg, remotely)
    - Had at least one known study drug interruption
- A listing of subjects who discontinued study drug and/or withdrew from study due to the COVID-19 pandemic, which will include the specific reasons.
- A listing by subject of visits and assessments affected by the COVID-19 pandemic (eg, missing, partial, collected remotely).

The MedDRA standardized MedDRA queries (SMQ) of COVID-19 will be utilized to retrieve preferred terms (PTs) pertaining to COVID-19 to identify subjects with a presumed or confirmed diagnosis of COVID-19.

Further classification and summaries of protocol deviations related to the COVID-19 pandemic are detailed in Section 7.2.

Additional summaries may be generated to address the potential impact of the COVID-19 pandemic on the efficacy and safety data, as needed, based on ongoing data review or during the final analysis of study data in each of study periods.

7. STUDY POPULATION

7.1. Disposition

The summary of subject enrollment and disposition in the DB period will be displayed by DB treatment group (as defined in Section 6.1) as well as all subjects combined.

The table will include:

- The total number of subjects who provided informed consent/assent and were screened (provided in table footnote)
- Number randomized (within column header)
- Number received study drug
- Number completed planned study drug through Week 28
- Number completed study through Week 28
- Number who did not complete study drug through Week 28, including reasons for early discontinuation of study drug
- Number who did not complete the study through Week 28, including reasons for early termination from the study

The summary of subject disposition in the OL period will be displayed by OL treatment group (as defined in Section 6.1) based on the number of subjects in the OL period safety analysis set (within column header).

The table will include:

- Number who received study drug in the OL period
- Number who completed planned study drug through Week 52
- Number completed study through Week 52
- Number who did not complete study drug through Week 52, including reasons for early discontinuation of study drug
- Number who did not complete the study through Week 52, including reasons for early termination from the study

A summary of the number and percentage of subjects in the FAS classified as Completers, Retrieved Dropouts, Non-retrieved Dropouts, and On Drug-Missing Endpoint Subjects (as defined in Section 6.4) for each of the primary and key secondary endpoints will be presented by treatment group for the DB period.

A listing of randomized subjects will also be provided and will include subject ID, informed consent/assent date, randomization date, stratification factors, and randomized treatment group (including dose level).

A summary of randomization by study site will be presented. This summary will display the number of subjects randomized to each treatment group by site.

7.2. Protocol Deviations

Protocol deviations described in the study-specific Protocol Deviation Handling Plan (PDHP) will be entered into the clinical trial management system. Prior to database lock, all major

protocol deviations that have been entered into the clinical trial management system will be exported to a file and integrated into the study data.

A protocol deviation that is classified as "major" is a deviation that has impacted or may significantly affect a subject's rights, safety, or well-being, or is likely to have a significant impact on the primary or key secondary efficacy endpoint(s) for at least one subject. An assessment of protocol deviations will be performed by a committee composed of NBI Clinical Development study team members on an ongoing basis and prior to database lock and unblinding at the end of the DB period. At the time of NBI protocol deviation review, a major protocol deviation may be further classified as an IPD (Important Protocol Deviation) based on the magnitude of the deviation. An IPD is defined as a protocol deviation that may significantly impact the completeness, accuracy, and/or reliability of key study data or that may significantly affect a subject's rights, safety, or well-being. Examples of what constitutes an IPD are listed in the PDHP. The above review will also be performed prior to database lock at the end of the OL period. The summaries described below will be produced for both treatment periods.

A summary of the number and percentage of subjects with IPDs by deviation category will be presented by treatment group and overall. The summary will be repeated for the subset of IPDs that are related to the COVID-19 pandemic.

All major protocol deviations will also be presented in a data listing. Any major PDs that are classified as IPDs or are related to the COVID-19 pandemic will also be flagged in the listing.

7.3. Demographic and Baseline Characteristics

Demographics, baseline characteristics, and stratification factors (as entered into IRT) at study entry will be summarized descriptively by DB treatment group and overall using the FAS. The following variables will be summarized:

Demographics:

- age (years)
- age categories (2 - < 12 years; ≥ 12 to < 18 years)
- sex
- ethnicity
- race
- region (OUS vs US)

Baseline characteristics (based on eCRF):

- height: percentile and SDS (based on CDC [Centers for Disease Control and Prevention] stature-for-age charts, by sex and age)
- weight: percentile and SDS (based on CDC weight-for-age charts [2-20 years], by sex and age)
- body mass index (BMI): percentile and SDS (based on CDC BMI-for-age charts, by sex and age)
- BMI percentile category (< 85^{th} percentile vs ≥ 85^{th} percentile)
- BSA (m^2; DuBois formula where BSA [m^2] = Weight [kg]$^{0.425}$ × height (cm)$^{0.725}$ × 0.007184])

- Tanner stage (breast or testicular)
- Tanner stage category (breast or testicular) (Stage 1 or 2 vs Stage 3, 4, or 5)
- Tanner stage (pubic hair)
- Undergone menarche in the subset of female subjects (Y/N)
- Subjects of childbearing potential (Y/N)
- Serum androstenedione (average of Day 1 pre-morning glucocorticoid dose values)
- Serum androstenedione (average of all serial sampling Day 1 values)
- Serum androstenedione (midpoint of normal range to < ULN vs ≥ ULN)
- Serum 17-OHP (average of Day 1 pre-morning glucocorticoid dose values)
- Serum 17-OHP (average of all serial sampling Day 1 values)
- total daily glucocorticoid dose adjusted for BSA (mg/m^2/day; in hydrocortisone equivalents [as defined in Section 6.6])
- total daily glucocorticoid dose (mg/day; in hydrocortisone equivalents [as defined in Section 6.6])
- glucocorticoid type (hydrocortisone alone; prednisone, prednisolone, methylprednisolone, with or without hydrocortisone)
- glucocorticoid type (hydrocortisone, prednisone, prednisolone, methylprednisolone; note that subjects can contribute to more than one category if taking more than one type of glucocorticoid)
- total daily fludrocortisone dose (µg/day; for subjects taking fludrocortisone)
- Taking gonadotropin-releasing hormone agonist therapy (Y/N)
- Taking growth hormone therapy (Y/N)
- Taking aromatase inhibitors (Y/N)

Stratification Factors as entered in IRT

- pubertal stage (Tanner breast or genital stage 1 or 2 vs. 3, 4, or 5)
- sex

7.4. CAH History

7.4.1. CAH Medical Conditions of Interest

A summary of the number and percentage of subjects with each medical condition of interest ("Yes – Treatment Required"; "Yes – Treatment Not Required"; "Yes – Unknown Treatment History"; "No") as specified on the "Conditions of Interest" eCRF) at study entry will be summarized by DB treatment group and overall. For conditions that are sex-specific, the denominator will be based on the number of subjects of that sex within the treatment group.

CAH medical conditions of interest include:

- Advanced bone age
- Early puberty
- Central precocious puberty
- Short stature
- Hypertension
- Obesity or overweight

- Insulin resistance
- Type 2 diabetes mellitus
- Attention deficit hyperactivity disorder
- Depression
- Anxiety
- Hirsutism (female subjects only)
- Acne (presented by sex as well as sexes combined)
- TARTs (male subjects only)
- Polycystic ovary syndrome (female subjects only)
- Irregular menstrual cycles (female subjects only)

7.4.2. CAH Medical History

CAH medical history at study entry will be summarized by DB treatment group and overall using descriptive statistics. For conditions that are sex-specific, the denominator will be based on the number of subjects of that sex within the treatment group.

CAH history assessments include:

- Age at CAH diagnosis
- How the CAH diagnosis was made (elevated 17-OHP; CYP21A2 genetic testing; cosyntropin stimulation)
- Lifetime total number of adrenal crises (requiring hospitalization and parenteral glucocorticoid administration)
- Number of (lifetime) adrenal crises related to stopped/decreased glucocorticoid medication
- Causes of adrenal crises (infection with fever; respiratory tract infection; gastrointestinal infection; urinary tract infection; stopped or decreased glucocorticoid medication; unknown; other)
- Number of adrenal crises within the past 10 years (requiring hospitalization and parenteral glucocorticoid administration)
- Number of adrenal crises within the past 10 years related to stopped/decreased glucocorticoid medication
- Number of adrenal crises within the past 5 years (requiring hospitalization and parenteral glucocorticoid administration)
- Number of adrenal crises within the past 5 years related to stopped/decreased glucocorticoid medication
- Number of adrenal crises within the past year (requiring hospitalization and parenteral glucocorticoid administration)
- Number of adrenal crises within the past year related to stopped/decreased glucocorticoid medication
- Age at menarche (years; female subjects only)
- Age at Tanner Stage 2 (breast or testicular)
- Age at Tanner Stage 2 (pubic hair)

7.5. Summary of Analysis Sets

A summary of the number and percentage of subjects included in the full analysis set, the DB period safety analysis set, the OL period safety analysis set, and the PK analysis set will be provided by DB period treatment group and overall. The number and percentage of subjects excluded from each analysis set by reason for exclusion will also be provided.

7.6. Medical History

Medical history data will be summarized descriptively for the full analysis set, by treatment group and overall. Medical history data will be coded using MedDRA. The medical history data will be summarized with frequencies and percentages of subjects with at least one medical history item, and subject frequencies and percentages according to the System Organ Class (SOC) and Preferred Term (PT) levels. The table will be sorted alphabetically by SOC and then, within a SOC, by PT.

8. PHARMACOKINETICS

The plasma concentrations of crinecerfont and its metabolites will be summarized with descriptive statistics by dose level, nominal visit, and nominal time postdose in the DB period. This summary will be repeated for all plasma concentration values collected in the OL period. All outputs for the summaries will be based on the PK Analysis Set for the DB period and on the OL Period Safety Analysis set for the OL period.

All BLQ values (crinecerfont < 5.00 ng/mL; metabolites < 0.500 ng/mL) will be set equal to zero (0) in the plasma concentration summaries. If a subject receives incorrect study drug at any point during the study, the subject's plasma concentration values will be excluded from the summaries that are directly impacted by the incorrect administration and will be documented in the CSR.

Plasma concentration data from this study will be used for population PK analysis including data from other studies. The details of the population PK analysis and the results will be described separately from the clinical study report.

9. EFFICACY – DB PERIOD

The efficacy endpoints and planned analysis methods for the DB period are described below. Unless otherwise specified, the FAS will be used for all efficacy analyses.

9.1. Multiple Comparisons and Multiplicity

A fixed-sequence testing procedure will be followed for the primary and key secondary efficacy endpoint analyses to control for the treatment group comparisons of multiple endpoints. The fixed-sequence testing procedure will consist of performing the hypothesis tests in the following prespecified order:

1. Primary endpoint: change from baseline in serum androstenedione at Week 4
2. Key secondary endpoint: change from baseline in serum 17-OHP at Week 4
3. Key secondary endpoint: percent change from baseline in daily glucocorticoid dose (in hydrocortisone dose equivalents adjusted for BSA) at Week 28, while Week 28 serum androstenedione is ≤120% of the baseline value or ≤ ULN, according to sex and either age (for Tanner stage 1) or pubertal stage (for Tanner stages 2 to 5)

Each step in the sequential testing procedure will use a 2-sided 0.05 level of significance for the null hypothesis being tested. Testing of hypotheses at each step of the procedure commences only if all null hypotheses in prior steps were rejected.

All other p-values will not be adjusted for multiplicity and should be considered nominal p-values.

9.2. Statistical Models

9.2.1. ANCOVA – continuous endpoints

The continuous efficacy endpoints will be analyzed using an analysis of covariance (ANCOVA) model, which includes the change from baseline to the respective postbaseline visit where the endpoint is being evaluated. The model will include the relevant baseline value as a covariate, treatment group (crinecerfont, placebo), and stratification factors used in the randomization. Values of the stratification factors entered during randomization in IRT will be used in the primary analysis methods. Any corrected values as reported in the eCRFs for the stratification factors may be used in any supplementary or exploratory analyses and will be stated as such.

An example of the SAS software code is provided below. Note that the PROC MIXED statements shown in this section will be accompanied by additional statements to create and output the required statistics.

```
PROC MIXED;

CLASS SEX PUBSTAGE TRTP;

MODEL CHG = TRTP BASE SEX PUBSTAGE;

RUN;
```

9.2.2. CMH – categorical endpoints

Categorical efficacy endpoints will be analyzed using the Cochran-Mantel-Haenszel (CMH) test. The CMH test will compare treatment groups (crinecerfont vs. placebo) and will include the stratification factors used in the randomization.

An example of the SAS software code is provided below. Note that the PROC FREQ statements shown in this section will be accompanied by additional statements to create and output the required statistics.

```
PROC FREQ;
TABLES SEX*PUBSTAGE*TRTP*RESPONSE / CMH;
RUN;
```

9.3. Analysis of the Primary Efficacy Endpoint

9.3.1. Primary Estimand

The Primary Estimand is defined by the following:

Population: Pediatric subjects in the FAS with classic CAH due to 21-hydroxylase deficiency on a stable, supraphysiologic glucocorticoid dose regimen (> 12 mg/m^2/day) at study entry and have an androstenedione level (prior to the morning glucocorticoid dose) greater than the midpoint of the reference range. Subjects must have met the other inclusion and exclusion criteria defined in the protocol.

Variable (or endpoint) to be obtained for each subject: change from baseline in serum androstenedione at Week 4

Intercurrent events will be handled using the following strategy:

1) If a subject discontinues study drug but stays on study through Week 4 and has serum androstenedione measured at Week 4, the subject's observed serum androstenedione at Week 4 will be used in the analysis.
2) If a subject is on study drug through Week 4 but is missing Week 4 serum androstenedione, the subject's endpoint will be multiply imputed from subjects within the same treatment group with non-missing data at Week 4 (Week 4 Completers) (Section 17.1).
3) If a subject in the crinecerfont treatment group discontinues study drug prior to Week 4 and is missing serum androstenedione at Week 4, the subject's endpoint will be multiply imputed using data from Retrieved Dropouts (Section 17.1). If there is an insufficient number of Retrieved Dropouts, the subject's endpoint will be imputed using observed data from subjects in the placebo treatment group (Section 17.1). For subjects in the placebo treatment group who discontinue study drug (prior to Week 4) and are missing serum androstenedione at Week 4, missing data will be imputed using non-missing data from subjects in the placebo treatment group (Section 17.1).
4) If a subject's Week 4 androstenedione value is obtained within a timeframe that could be confounded by stress dosing with glucocorticoids, the subject's androstenedione value

will be imputed using non-missing data from subjects in the placebo treatment group (Section 17.1).

Population-level summary: The least-squares mean treatment difference for this endpoint will be estimated using an ANCOVA model, as described in Section 9.2.1. Missing data will be imputed using the multiple imputation procedure defined in Section 17.1.

9.3.2. Primary Efficacy Analysis

The primary efficacy endpoint is the change from baseline in serum androstenedione at Week 4. The definitions of the baseline and Week 4 serum androstenedione values that will be used in this analysis are described in Section 6.3.3. The primary analysis of the primary endpoint will be performed using an ANCOVA model, as described in Section 9.2.1. The model will include treatment group, baseline serum androstenedione (as defined in Section 6.3.3), and stratification factors used in the randomization.

Subjects who are missing serum androstenedione at all timepoints at the Week 4 visit will have their data imputed through the multiple imputation procedure described in Section 17.1.

Statistical significance for the treatment comparison will be determined using the multiple comparison procedure described in Section 9.1. Descriptive statistics and LS means will be presented by treatment group along with the LS mean treatment difference, 95% confidence intervals, and p-value.

Descriptive statistics will be presented for serum androstenedione by treatment group and visit from baseline through Week 28. For visits where androstenedione is collected more than once (eg, Day 1; Week 4), the average of the visit values will be analyzed. For visits where more than one pre-morning GC dose sample is collected, the average of the pre-morning GC dose will be analyzed; otherwise, the pre-morning GC dose is the single value collected. Observed and change from timepoint-matched baseline values will be summarized for visit average values and pre-morning GC dose values. The androstenedione observed values from the screening visit will also be summarized.

Mean (±SEM) values of serum androstenedione at each visit will be summarized in line graphs by treatment group. Pre-morning GC dose values and visit average values will be presented on separate plots. Similar graphs will be presented for the changes from baseline.

9.3.3. Sensitivity Analyses of the Primary Efficacy Results

The assumptions relative to the ANCOVA model (eg, linearity, normality) specified for the primary analysis of the primary endpoint will be checked and any violations of these assumptions will be handled through non-parametric statistical methods.

A tipping point sensitivity analysis will be performed to assess the robustness of the missingness assumptions. This sensitivity analysis is based on “delta adjustments” which is a commonly used approach to assess the impact of missing data in clinical trials (O’Kelly and Ratitch, 2014). The imputed Week 4 androstenedione values (from the MI procedure described in Section 17.1) for subjects in both the crinecerfont treatment group and the placebo treatment group will be delta-adjusted (with a range of plausible penalties for each treatment group) until the treatment difference at Week 4 is no longer statistically significant. The implementation procedure for the

tipping point analysis is described in Appendix 17.3. The combinations of delta adjustments that tip the result from statistically significant to insignificant, along with the corresponding insignificant p-values, will be presented in a table and may be presented in a figure.

A complete-case sensitivity analysis will also be performed to test the robustness of the multiple imputation procedure. The ANCOVA analysis for the primary endpoint will be repeated using only subjects with observed data at Week 4. Subjects who are missing the Week 4 primary endpoint will be excluded from the analysis.

Descriptive statistics and LS means for the complete-case sensitivity analyses will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and nominal p-values.

9.3.4. Supplementary Analyses of the Primary Efficacy Results

The following supplementary analyses of the primary efficacy results will also be performed:

- The percent change from baseline in serum androstenedione at Week 4 will be analyzed using the same method as described in the primary analysis of this endpoint. Missing data will be imputed using the same method as described in the primary analysis of this endpoint.
- The change from baseline in serum androstenedione at Week 4 will be analyzed using the average of all serial sampling androstenedione values at Day 1 and Week 4. If all Day 1 values are missing, then the screening androstenedione value will be used instead. Note that subjects <6 years of age or <20 kg will have only a single serum androstenedione value for both baseline and Week 4. This analysis will be performed using the same method as described in the primary analysis of this endpoint. Missing data will also be imputed using the same method as described in the primary analysis of this endpoint.
- The change from baseline to Week 4 in serum androstenedione, derived as a multiple of the ULN (calculated as observed value divided by the ULN value) [based on age and sex], will be analyzed using the same method as described in the primary analysis of this endpoint. Missing data will also be imputed using the same method as described in the primary analysis of this endpoint.

Descriptive statistics and LS means for the supplementary analyses will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and nominal p-values.

9.4. Analysis of Key Secondary Efficacy Endpoints

9.4.1. Serum 17-OHP

The first key secondary endpoint is the change from baseline in serum 17-OHP at Week 4. The definition of the baseline and Week 4 serum 17-OHP values that will be used in this analysis are described in Section 6.3.4. The primary analysis of this endpoint will be performed using an ANCOVA model, as described in Section 9.2.1. The model will include treatment group, baseline serum 17-OHP (as defined in Section 6.3.4), and stratification factors used in the randomization. Subjects who are missing serum 17-OHP levels at all timepoints at the Week 4

visit will have their data imputed through the multiple imputation procedure described in Section 17.1.

Statistical significance for the treatment comparison will be determined using the multiple comparison procedure described in Section 9.1. Descriptive statistics and LS means will be presented by treatment group along with the LS mean treatment difference, 95% confidence intervals, and p-value.

Descriptive statistics will be presented for serum 17-OHP by treatment group and visit from baseline through Week 28. For visits where 17-OHP is collected more than once, the average of the visit values will be analyzed. For visits where more than one pre-morning GC dose sample is collected, the average of the pre-morning GC dose will be analyzed; otherwise, the pre-morning GC dose is the single value collected. Observed and change from timepoint-matched baseline values will be summarized for visit average values and pre-morning GC dose values. The serum 17-OHP observed values from the screening visit will also be summarized.

Mean (±SEM) values of serum 17-OHP at each visit will be summarized in line graphs by treatment group. Pre-morning GC dose values and average values will be presented on separate plots. Similar graphs will be presented for the changes from baseline.

9.4.1.1. Sensitivity Analyses

The assumptions relative to the ANCOVA model (eg, linearity, normality) for the first key secondary endpoint will be checked and any violations of these assumptions will be handled through non-parametric statistical methods.

A tipping point sensitivity analysis will be performed to assess the robustness of the missingness assumptions. This sensitivity analysis is based on “delta adjustments” which is a commonly used approach to assess the impact of missing data in clinical trials (O’Kelly and Ratitch, 2014). The imputed Week 4 serum 17-OHP values (from the MI procedure described in Section 17.1) for subjects in both the crinecerfont treatment group and the placebo treatment group will be delta-adjusted (with a range of plausible penalties for each treatment group) until the treatment difference at Week 4 is no longer statistically significant. The implementation procedure for the tipping point analysis is described in Appendix 17.4. The combinations of delta adjustments that tip the result from statistically significant to insignificant, along with the corresponding insignificant p-values, will be presented in a table and may be presented in a figure.

A complete-case sensitivity analysis will also be performed to test the robustness of the multiple imputation procedure. The ANCOVA analysis for the first key secondary endpoint will be repeated using only subjects with observed data at Week 4. Subjects who are missing the Week 4 17-OHP measurement will be excluded from the analysis.

Descriptive statistics and LS means for the complete-case sensitivity analyses will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and nominal p-values.

9.4.1.2. Supplementary Analyses

As a supplementary analysis, the percent change from baseline in serum 17-OHP at Week 4 will be analyzed using the same method as described in the primary analysis of this endpoint. Missing

data will also be imputed using the same methods as described in the primary analysis of this endpoint.

As a supplementary analysis, the change from baseline in serum 17-OHP at Week 4 will be analyzed using the average of all serial sampling 17-OHP values at Day 1 and Week 4. If all Day 1 17-OHP values are missing, then the screening 17-OHP value will be used instead. Note that subjects <6 years of age or <20 kg will have only a single serum 17-OHP value for both baseline and Week 4. This analysis will be performed using the same method as described in the primary analysis of this endpoint. Missing data will also be imputed using the same methods as described in the primary analysis of this endpoint.

As a supplementary analysis, the change from baseline to Week 4 in serum 17-OHP, derived as a multiple of the ULN (calculated as the observed value divided by the ULN value), will be analyzed using the same method as described in the primary analysis of this endpoint. Missing data will be imputed using the same methods as described in the primary analysis of this endpoint.

9.4.2. Glucocorticoid Daily Dose

The second key secondary endpoint is the percent change from baseline in glucocorticoid daily dose (in hydrocortisone dose equivalents adjusted for BSA [mg/m^2/day]) at Week 28, while Week 28 serum androstenedione is ≤120% of the baseline value or ≤ ULN, according to sex and either age (for Tanner stage 1) or pubertal stage (for Tanner stages 2 to 5). Glucocorticoid daily dose will be calculated at baseline (as defined in Section 6.3.1) and at Week 28 using the doses reported on the glucocorticoid medications eCRF (see details in Section 6.7 regarding mapping of the glucocorticoid medication to Week 28). Glucocorticoids other than hydrocortisone (ie, methylprednisolone, prednisolone, and prednisone) will be converted to hydrocortisone equivalents using an equivalency ratio as described in Section 6.6. BSA will be calculated using the DuBois formula (Section 7.3). The calculation of glucocorticoid dose in hydrocortisone equivalents adjusted for BSA at Day 1 and Week 28 will be based on height and weight measurements at Day 1 and Week 28, respectively. If glucocorticoid dosing alternates over the course of two consecutive days, then the average of the glucocorticoid dose on those days will be used to calculate the glucocorticoid daily dose. If glucocorticoid dosing alternates between weekdays and weekends (or another weekly regimen that varies by day of the week), then the average daily dose will be calculated as the sum of the glucocorticoid doses over the 7-day period divided by 7.

The serum androstenedione values at baseline and Week 28 to be used in the assessment of androstenedione “control” is defined in Section 6.3.5.

Subjects who have a decrease in glucocorticoid daily dose at Week 28 and who are not able to maintain their serum androstenedione at Week 28 at ≤120% of their baseline value or ≤ ULN for sex and either age (for Tanner stage 1) or pubertal stage (for Tanner stages 2 to 5) will be considered to have a 0 percent change from baseline in the glucocorticoid daily dose at Week 28. Subjects with an increase in glucocorticoid daily dose at Week 28 will have their observed glucocorticoid daily dose value used regardless of androstenedione control. Subjects who are missing glucocorticoid dose at Week 28 and serum androstenedione at all timepoints at the Week 28 visit will have their data imputed through the multiple imputation procedure described in Section 17.2.

The primary analysis of this endpoint will be performed using an ANCOVA model, as described in Section 9.2.1. The model will include treatment group, baseline glucocorticoid daily dose (mg/m^2/day), and stratification factors used in the randomization.

Descriptive statistics will be presented by treatment group for the glucocorticoid total daily dose (mg/m^2/day) observed, change from baseline, and percent change from baseline values at baseline (observed data) and each postbaseline visit. Mean (±SEM) values of the glucocorticoid total daily dose at each visit will be summarized in line graphs by treatment group. Similar graphs will be presented for the change from baseline and percent change from baseline. These summaries will be presented for both the glucocorticoid total daily doses *without* the derivation of 0 for loss of androstenedione control as well as the glucocorticoid total daily doses *with* the derivation of 0 for loss of androstenedione control (for change from baseline and percent change from baseline summaries only).

9.4.2.1. Sensitivity Analyses

The assumptions relative to the ANCOVA model (eg, linearity, normality) for the second key secondary endpoint will be checked and any violations of these assumptions will be handled through non-parametric statistical methods.

A tipping point sensitivity analysis will be performed to assess the robustness of the missingness assumptions. This sensitivity analysis is based on "delta adjustments" which is a commonly used approach to assess the impact of missing data in clinical trials (O'Kelly and Ratitch, 2014). The imputed Week 28 serum glucocorticoid daily dose values (from the MI procedure described in Section 17.2) for subjects in both the crinecerfont treatment group and the placebo treatment group will be delta-adjusted (with a range of plausible penalties for each treatment group) until the treatment difference at Week 28 is no longer statistically significant. The implementation procedure for the tipping point analysis is described in Appendix 17.5. The combinations of delta adjustments that tip the result from statistically significant to insignificant, along with the corresponding insignificant p-values, will be presented in a table and may be presented in a figure.

A complete-case sensitivity analysis will also be performed to test the robustness of the multiple imputation procedure. The ANCOVA analysis for the first key secondary endpoint will be repeated using only subjects with observed data at Week 28. Subjects who are missing any component of the Week 28 second key secondary endpoint will be excluded from the analysis.

Descriptive statistics and LS means for the complete-case sensitivity analyses will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and nominal p-values.

9.4.2.2. Supplementary Analyses

The ANCOVA analysis as described for this endpoint will be repeated using the absolute (rather than the percent) change from baseline in the glucocorticoid total daily doses (in mg/m^2/day) captured in the glucocorticoid concomitant medication eCRF at baseline and Week 28. All conventions used in the primary analysis of this endpoint will be used in this supplementary analysis.

The ANCOVA analysis (as described above) for the primary analysis of the second key secondary endpoint will be repeated using the Week 28 data without setting glucocorticoid dose reductions to 0 due to loss of androstenedione control. That is, all observed and multiply imputed GC dose data at Week 28 will be used regardless of the subject's androstenedione values. Missing data at Week 28 will be imputed using the same methods as in the primary analysis of this endpoint.

The ANCOVA analysis for the primary analysis of the second key secondary endpoint will be repeated with a modification to the androstenedione control criteria, with androstenedione control being defined as ≤150% of the baseline value or ≤ ULN, according to sex and either age (for Tanner stage 1) or pubertal stage (for Tanner stages 2 to 5). Missing data at Week 28 will be imputed using the same methods as in the primary analysis of this endpoint.

The ANCOVA analysis for the primary analysis of the second key secondary endpoint will be repeated with a modification to the androstenedione control criteria, with androstenedione control being defined as ≤180% of the baseline value or ≤ ULN, according to sex and either age (for Tanner stage 1) or pubertal stage (for Tanner stages 2 to 5). Missing data at Week 28 will be imputed using the same methods as in the primary analysis of this endpoint.

The ANCOVA analysis for the primary analysis of this endpoint will be repeated using the latest observed glucocorticoid total daily dose at Week 4, Week 8, Week 12, Week 16, or Week 20 where androstenedione is controlled if control is not met at Week 28. For subjects who do not meet the androstenedione control definition at any visit between Week 4 and Week 28, inclusive, their percent change from baseline in glucocorticoid total daily dose will be set to 0. If control is not met at Week 28 and glucocorticoid total daily dose is missing at all visits between Week 4 and Week 20 (inclusive) where control is met, then the Week 28 glucocorticoid total daily dose will be used in analysis (ie, percent change from baseline will be set to 0). Missing data at Week 28 will be imputed using the same methods as in the primary analysis of this endpoint. Missing data at all other visits will not be imputed.

The ANCOVA analysis for the primary analysis of this endpoint will be repeated using the highest observed glucocorticoid dose at Week 4, Week 8, Week 12, Week 16, or Week 20 where androstenedione is controlled if control is not met at Week 28. For subjects who do not meet the androstenedione control definition at any visit between Week 4 and Week 28, inclusive, their percent change from baseline in glucocorticoid total daily dose will be set to 0. If control is not met at Week 28 and glucocorticoid total daily dose is missing at all visits between Week 4 and Week 20 (inclusive) where control is met, then the Week 28 glucocorticoid total daily dose will be used in analysis (ie, percent change from baseline will be set to 0). Missing data at Week 28 will be imputed using the same methods as in the primary analysis of this endpoint. Missing data at all other visits will not be imputed.

9.5. Analysis of Secondary Efficacy Endpoints

The following sections describe the secondary efficacy endpoints and methods of summarization and analysis. With the exception of the physiologic glucocorticoid daily dose responder analysis (see Section 9.5.1), missing data will not be imputed and only observed cases will be used for analyses of secondary efficacy endpoints. All p-values produced from the analyses in this section will be considered nominal p-values and will not be adjusted for multiplicity.

9.5.1. Physiologic Glucocorticoid Daily Dose Responder Analysis

The achievement of a reduction in glucocorticoid daily dose to physiologic levels (≤11 mg/m^2/day in hydrocortisone dose equivalent adjusted for BSA) at Week 28 while serum androstenedione is ≤120% of the baseline value or ≤ULN for sex and either age (for Tanner stage 1) or pubertal stage (for Tanner stages 2 to 5) is a secondary endpoint. The glucocorticoid daily dose and androstenedione values used in the derivation of this endpoint will follow the same conventions as described in Section 9.4.2. The primary analysis of this endpoint will be performed using a CMH test as described in Section 9.2.2. The CMH test will compare treatment groups and will include the stratification factors used in the randomization. Subjects whose androstenedione levels at Week 28 are not controlled will be considered non-responders. For subjects who are missing glucocorticoid dose, androstenedione, or both at Week 28, the multiply-imputed datasets from the second key secondary endpoint will be used to determine response.

Descriptive statistics will be presented by treatment group for the number and percentage of subjects classified as responders and non-responders at each postbaseline visit, using the same methods of determining androstenedione control as described above for Week 28. No data will be imputed at visits other than Week 28. For the Week 28 visit, the p-value from the CMH analysis will also be displayed.

9.5.2. Body Mass Index

An additional secondary endpoint is the change from baseline in BMI SDS at Week 28 which will be analyzed using an ANCOVA model, as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 28 in BMI SDS will be presented by treatment group along with the LS-mean treatment difference and p-value. Note that BMI SDS and BMI percentiles will be calculated based on the published CDC Growth Charts (ages 2 to 20 years) (Kuczmarski et al., 2002).

As a supplementary analysis, the change from baseline in BMI percentile at Week 28 will be analyzed using an ANCOVA model, as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 28 in BMI percentile will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

Descriptive statistics will be presented by treatment group and visit for BMI SDS and BMI percentile observed and change from baseline values. Mean (±SEM) observed values at each visit will be summarized in line graphs by treatment group. Similar graphs will be presented for the change from baseline.

Shifts from baseline to Week 28 in BMI percentile categories will be presented in tables. The categories are defined as follows: underweight (BMI < 5^{th} percentile), normal weight (BMI ≥ 5^{th} percentile and < 85^{th} percentile), overweight (BMI ≥ 85^{th} percentile and < 95^{th} percentile), and obese (BMI ≥ 95^{th} percentile). Each shift table will have five rows and five columns, with rows reflecting the categories at baseline as well as missing values and columns reflecting the categories at Week 28 as well as missing values. A “Total” row and “Total” column will also be included. The number and percentage of subjects in each shift category will be displayed in the table; percentages will be based on the number of subjects included in the table.

In the subset of subjects classified as overweight or obese (where BMI ≥85th percentile is overweight and BMI ≥ 95th percentile is obese), the number and percentage of subjects achieving normal BMI (defined as BMI < 85th percentile) from baseline to Week 28 will also be presented by treatment group.

9.5.3. Salivary 17-OHP

An additional secondary endpoint is the change from baseline in mean 24-hour salivary 17-OHP at Week 28. The salivary 17-OHP values collected at baseline and Week 28 will be averaged at each respective visit and analyzed using an ANCOVA model, as described in Section 9.2.1. Descriptive statistics and LS means for the change from baseline to Week 28 in mean 24-hour salivary 17-OHP will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

As a supplementary analysis, the percent change from baseline in mean 24-hour salivary 17-OHP at Week 28 will be analyzed using the same method as described in the primary analysis of this endpoint.

As a supplementary analysis, the change from baseline in mean morning salivary 17-OHP (average of 0600, 0800, +3 hours post-GC dose, and +6 hours post-GC dose values) at Week 28 will be analyzed using an ANCOVA model as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 28 in mean morning salivary 17-OHP will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

Descriptive statistics will be presented for salivary 17-OHP by treatment group, visit, and timepoint from baseline through Week 28. For visits where pre-dose salivary 17-OHP is collected more than once (eg, Day 1; Week 4), the average of the pre-dose values will be analyzed. Otherwise, the pre-dose result is the single value collected. Observed and change from timepoint-matched baseline will be summarized.

Mean (±SEM) values of salivary 17-OHP at each visit will be summarized in line graphs by treatment group. Pre-dose values and visit average values will be presented on separate plots. Similar graphs will be presented for the changes from baseline.

The relationship between salivary and serum 17-OHP values will be depicted graphically. Scatterplots of salivary vs. timepoint-matched serum 17-OHP values will be generated by treatment group. The correlation coefficient, R, will be calculated and displayed in the scatterplots. A regression line will also be displayed on the scatterplots.

9.5.4. Bone Age

The change from baseline to Week 28 in the ratio of bone age to chronological age (BA:CA) is a secondary endpoint in the subgroup of subjects not at adult height (female subjects with the most recent prior bone age <14 years and males with the most recent prior bone age <16 years at study entry) and not on gonadotropin-releasing hormone agonist therapy, growth hormone therapy, or aromatase inhibitors at study entry. BA:CA will be calculated at baseline and Week 28 by taking the bone age of the subject and dividing it by the chronological age of the subject at each respective visit. The change from baseline to Week 28 in BA:CA will be analyzed using an ANCOVA model, as described in Section 9.2.1. Descriptive statistics and LS means for change

from baseline to Week 28 in BA:CA will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

The change from baseline to Week 28 in the difference in bone age and chronological age (BA-CA) will also be analyzed using an ANCOVA model, as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 28 in BA-CA will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

9.6. Exploratory Efficacy Endpoints

The following sections describe the exploratory efficacy endpoints and methods of summarization and analysis. For all analyses of the exploratory efficacy endpoints, missing data will not be imputed and only observed cases will be used. All p-values produced from the analyses in this section will be considered nominal p-values and will not be adjusted for multiplicity.

9.6.1. Serum or Plasma Hormone Measurements

Serum or plasma hormone measurements including ACTH, testosterone, FSH, LH, and cortisol will be summarized with descriptive statistics by visit and treatment group. For hormones and visits where more than one sample is collected, the average of the visit values will be analyzed. For hormones or visits where more than one pre-morning GC dose sample is collected, the average of the pre-morning GC dose will be analyzed; otherwise, the pre-morning GC dose result is the single value collected. Observed and change from timepoint-matched baseline values will be summarized for visit average values and pre-morning GC dose values. Summaries of testosterone, LH, and FSH will be stratified by sex and Tanner stage (breast or testicular Tanner stages 1, 2 vs. Tanner stages 3-5).

The change from baseline to Week 28 for each of these parameters will be analyzed using an ANCOVA model as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 28 in each of these hormone parameters will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

Mean (±SEM) values of serum hormone parameters at each visit will be summarized in line graphs by treatment group. Pre-morning GC dose values and visit average values will be presented on separate plots. Similar graphs will be presented for the changes from baseline.

A responder definition for the normalization of testosterone in female subjects will be defined as female subjects with testosterone levels that are greater than or equal to 1.2x ULN (based on age) at baseline and decrease to within the normal range at Week 28. The number and percentage of female subjects meeting this responder definition at Week 28 will be summarized by treatment group.

A responder definition for the ratio of androstenedione to testosterone (A4/T) in male subjects (Tanner stage 2 [testicular] and above) will be defined as male subjects who have A4/T greater than or equal to 0.5 (based on the pre-glucocorticoid dose values for androstenedione and testosterone) at baseline and who achieve A4/T less than 0.5 at Week 28. The number and percentage of male subjects meeting this responder definition at Week 28 will be summarized by treatment group.

The achievement of response for testosterone in female subjects and A4/T in male subjects at Week 28 (as defined for each of the responder analyses described above) will be analyzed using a CMH test as described in Section 9.2.2. The CMH test will compare treatment groups and will include stratification factors used in the randomization (with the exception of sex). Subjects who are missing Week 28 data will be omitted from the analysis.

9.6.2. Salivary Hormone Measurements

Salivary hormone measurements, including androstenedione, testosterone, cortisol, progesterone, and pregnenolone will be summarized with descriptive statistics by timepoint and treatment group. Summaries of testosterone, progesterone, and pregnenolone will be stratified by sex and Tanner stage (breast or testicular Tanner stages 1, 2 vs. Tanner stages 3-5). For visits where more than one pre-dose sample is collected, the average of the pre-dose values will be analyzed. Otherwise, the pre-dose result is the single value collected. For visits where more than one sample is collected, the average of all timepoints at that visit will be averaged and analyzed as the mean 24-hour value. Observed and change from timepoint-matched baseline will be summarized for pre-dose values and mean 24-hour values.

The change from baseline to Week 28 in mean 24-hour values for each of the parameters will be analyzed using an ANCOVA model as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 28 in each of these hormone parameters will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

As an additional analysis, the change from baseline to Week 28 in mean morning salivary samples (average of 0600, 0800, +3 hours post-GC dose, and +6 hours post-GC dose) for each of the parameters will be analyzed using an ANCOVA model as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 28 in each of these parameters will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

Mean (±SEM) values of salivary hormone parameters at each visit will be summarized in line graphs by treatment group. Pre-dose values and mean 24-hour values will be presented on separate plots. Similar graphs will be presented for the changes from baseline.

The relationship between salivary and serum values will be depicted graphically for androstenedione, testosterone, and cortisol. Scatterplots of salivary vs. timepoint-matched serum values will be generated by treatment group. The correlation coefficient, R, will be calculated and displayed in the scatterplots. A regression line will also be displayed on the scatterplots.

9.6.3. Metabolic Assessments (Fasting Lipids and HOMA-IR)

Metabolic assessments collected on this study include HOMA-IR and fasting lipids (total cholesterol, LDL, HDL, and triglycerides). The change from baseline to Week 28 for each of these parameters will be analyzed using an ANCOVA model as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 28 for each of these parameters will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

9.6.4. Testicular Adrenal Rest Tumors (TARTs)

For male subjects with TARTs, the length, width, and depth of each lesion will be measured and recorded on the eCRF. For the purpose of summarization and analysis, the volume of each lesion will be used to assess size (in cubic millimeters) at visits where the testicular ultrasound is performed. Volume will be calculated using the following formula: volume = length x width x depth x 0.5236. The volume of each testicle will be calculated using the same formula. Descriptive statistics will be used to summarize the total TART volume (calculated as the sum of the volume of each lesion across both testes) for each subject by treatment group and visit. Observed and change from baseline values will be included in the summary table. Note that only male subjects with TARTs at baseline will have a testicular ultrasound performed at Week 28.

Additionally, the total TART volume will be expressed as a percentage of the total testicular volume and calculated as the sum of the volume of each lesion across both testes divided by the total testicular volume across both testes for each subject. The total TART volume as a percentage of the total testicular volume will be summarized with descriptive statistics by treatment group and visit. Observed and change from baseline values will be included in the summary table.

Change from baseline to Week 28 in total TART volume as a percentage of total testicular volume in the subset of male subjects with TARTs at baseline will be analyzed using an ANCOVA model, as described in Section 9.2.1 (with the exception of sex as a stratification factor in the model). Descriptive statistics and LS means for change from baseline to Week 28 in total TART volume as a percentage of total testicular volume will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

The number and percentage of subjects with the presence of TARTs (Y/N), whether they are unilateral or bilateral, and the average number of TARTs per subject will be summarized with descriptive statistics by visit and treatment group. The number and percentage of TARTs with each shape (round; lobular; irregular), boundary (clear; indistinct), hypervascularity (Y/N), calcification (Y/N), and the echogenicity will be summarized with descriptive statistics by visit and treatment group.

A response assessment will be conducted at Week 28 for subjects with the presence of TARTs at baseline. Response categories and the criteria for each category are detailed in Table 7. The frequency and percentage of subjects meeting each classification at Week 28 (as compared to baseline) will be summarized by treatment group. Complete responders will also be analyzed using a CMH test as described in Section 9.2.2. The CMH test will compare treatment groups and will include the stratification factors used in the randomization (with the exception of sex). The p-value from the CMH analysis will also be displayed in a table. This analysis will be repeated with complete and partial responders pooled together.

Table 7: Response Categories

Response Category	Definition
Complete Response (CR)	Disappearance of all TARTs.

Response Category	Definition
Partial Response (PR)	≥30% decrease in total TART volume compared with baseline
Progressive Disease (PD)	≥30% increase in total TART volume compared with baseline
Stable Disease (SD)	Neither CR, PR nor PD

9.6.5. Menstrual Regularity

Female subjects who have undergone menarche and are not on hormonal or intrauterine device contraceptives were asked to input the dates for which they were assessing their menstrual cycle as well as the amount of flow (light [spotting], medium [normal flow], heavy [heavy flow with flooding or clotting]). For the purposes of analysis, a menstrual cycle will be defined as two consecutive calendar days with any amount of flow. Menstrual cycle regularity is defined as a menstrual cycle every 21-35 days, translating to approximately 5 to 10 menstrual cycles from baseline to Week 28. As menstrual cycles can be irregular in the first year of menses, the following analysis will be restricted to female subjects in Tanner stages 4 and 5 only. The number and percentage of female subjects meeting the definition for regular menstrual cycle at Week 28 will be summarized by treatment group. This summary will be stratified based on menstrual cycle regularity at study entry (as collected on the "conditions of interest" eCRF, "irregular menstrual cycles" [yes or no]). This endpoint will be analyzed using a CMH test as described in Section 9.2.2. The CMH test will compare treatment groups and will include the stratification factors used in the randomization (with the exception of sex) as well as menstrual cycle regularity at study entry. The p-value from the CMH analysis will also be displayed in a table.

9.6.6. Growth

In subjects not at adult height (female subjects with the most recent prior bone age <14 years and males with the most recent prior bone age <16 years at study entry) and not on gonadotropin-releasing hormone agonist therapy, growth hormone therapy, or aromatase inhibitors at study entry, height SDS will be summarized with descriptive statistics by visit and treatment group.

The change from baseline to Week 28 in height SDS is an exploratory endpoint and will be analyzed using ANCOVA model as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 28 in height SDS will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

9.6.7. Bayley-Pinneau Predicted Adult Height SDS

The change from baseline in Bayley-Pinneau predicted adult height SDS at Week 28 will be analyzed using an ANCOVA model, as described in Section 9.2.1. Bayley-Pinneau predicted adult height SDS will be determined at baseline and Week 28 using the height and bone age of the subject at these respective visits and compared against the boys and girls tables of predicted heights (Bayley and Pinneau, 1952). The Bayley-Pinneau predicted height for all subjects will be

based on the "average" boys/girls tables with the exception of subjects who have an advanced or delayed bone age in excess of 1 year from the chronological age, in which case the "accelerated" or "retarded" predicted height tables will be used, respectively. Descriptive statistics and LS means for change from baseline to Week 28 in Bayley-Pinneau predicted adult height SDS will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

9.6.8. Hirsutism

Visual analog scales (VAS) were used to assess the subject's perception of severity of hirsutism in all female subjects, scored on a 100 mm visual analog scale, from 0 mm (no symptoms) to 100 mm (very severe symptoms). Observed and change from baseline values in VAS scores for hirsutism will be summarized descriptively by visit and treatment group.

The change from baseline to Week 28 in VAS score for hirsutism (female subjects only) will be analyzed using an ANCOVA model as described in Section 9.2.1 (with the exception of sex as a stratification factor in the model). Descriptive statistics and LS means for change from baseline to Week 28 in the VAS score for hirsutism will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

9.6.9. Acne

Visual analog scales (VAS) were used to assess the subject's perception of severity of acne in all subjects, scored on a 100 mm visual analog scale, from 0 mm (no symptoms) to 100 mm (very severe symptoms). Observed and change from baseline values in VAS scores for acne will be summarized descriptively by visit and treatment group for each sex as well as both sexes combined.

The change from baseline to Week 28 in VAS score for acne is an exploratory endpoint and will be analyzed using an ANCOVA model as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 28 in the VAS score for acne will be presented by sex and treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

9.6.10. Weight

Weight SDS will be summarized with descriptive statistics by visit and treatment group.

The change from baseline to Week 28 in weight SDS is an exploratory endpoint and will be analyzed using ANCOVA model as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 28 in weight SDS will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

9.6.11. Glucocorticoid Dose Regimen

In subjects on hydrocortisone alone at baseline with a dosing frequency of more than once per day (eg, 5x/day, QID, TID, or BID), a responder is defined as a subject whose dose is reduced by at least one dosing instance per day at Week 28. For example, responders include subjects who reduce from QID to TID (or BID or QD), TID to BID (or QD), or BID to QD, from baseline to Week 28. The number and percentage of subjects meeting this responder definition at Week 28

will be summarized by treatment group. A shift table of hydrocortisone dosing frequency from baseline to Week 28 will also be presented. The rows of the shift table will reflect the dosing frequency at baseline and the columns will reflect the dosing frequency at Week 28. The number and percentage of subjects in each shift category will be displayed in the table; percentages will be based on the number of subjects included in the table.

In subjects on a prednisone or equivalent (eg, methylprednisolone or prednisolone) or a hydrocortisone + prednisone (or equivalent) regimen at baseline, a responder is defined as a subject who switches to a hydrocortisone alone regimen at Week 28 (ie, a subject who stops taking prednisone or equivalent prior to the Week 28 visit and is on a hydrocortisone-only regimen at Week 28). The number and percentage of subjects meeting this responder definition at Week 28 will be summarized by treatment group.

The achievement of response for hydrocortisone alone dosing frequency reduction and switching from prednisone regimen to hydrocortisone alone at Week 28 (as defined for each of the responder analyses described above) will be analyzed using a CMH test as described in section 9.2.2. The CMH test will compare treatment groups and will include stratification factors used in the randomization. Subjects who are missing Week 28 data will be omitted from the analysis.

9.7. Examination of Subgroups

The following subgroups have been pre-planned and will be used to examine consistency of effect for the primary efficacy endpoint:

- Region (OUS versus US)
- Sex (male versus female; based on eCRF)
- BMI ($< 85^{th}$ percentile versus $\geq 85^{th}$ percentile)
- Race (White, Black or African American, Asian, Native Hawaiian or Other Pacific Islander, American Indian or Alaska Native)
- Pubertal stage (Tanner stages [breast or testicular] 1, 2 versus Tanner stages 3, 4, 5; based on eCRF)
- Age (< 12 years versus 12 – 17 years)
- Baseline androstenedione (midpoint of normal range to < ULN versus ≥ ULN)

If any of these categories are sparse in number of subjects (eg, <15), the categories with the fewest number of subjects may be collapsed until there are at least 15 subjects or the subgroup analysis may be omitted entirely. The decision to omit a subgroup analysis will be justified in the CSR. Each subgroup will be analyzed separately using an ANCOVA model that is similar to the model used in the primary analysis of the primary endpoint (as described in Section 9.2.1), with the addition of the subgroup variable added as a main effect term and an interaction term between treatment group and the subgroup variable. Note that for the sex and pubertal stage subgroups, the values reported on the eCRF will be used in the model instead of the values reported as stratification factors in the IRT. Nominal two-sided p-values for comparing treatment groups and the associated 95% confidence intervals will be reported in summary tables and a

forest plot, with p-values for the interaction term between treatment groups and subgroup variables presented.

10. EFFICACY – OL PERIOD

The efficacy endpoints and planned analysis methods for the OL period are described below. The OL period safety analysis set will be used for all efficacy analyses and descriptive statistics in the OL period. Unless otherwise noted, the summaries described below will follow the treatment group and baseline definitions defined for the OL period (as described in Section 6). Missing data will not be imputed for any of the following analyses and all summary displays will start with Week 32 as the first postbaseline visit, unless otherwise specified.

10.1. Serum or Plasma Hormone Measurements

Descriptive statistics will be presented by treatment group and visit for the observed and change from OL baseline values for the following serum or plasma hormone measurements in the OL period: ACTH, 17-OHP, androstenedione, testosterone, cortisol, LH, and FSH. Summaries of testosterone, LH, and FSH will be stratified by sex and Tanner stage (breast or testicular Tanner stages 1, 2 vs. Tanner stages 3-5). Visits where the hormone measurement is collected more than once will be handled as described in Section 9.6.1. Observed and change from timepoint-matched OL baseline sample will be summarized for pre-morning GC dose values and visit average values.

Mean (±SEM) values of the above hormone measurements at each visit in the OL period will be summarized in line graphs by treatment group. Pre-morning GC dose values and average values will be presented on separate plots. Similar graphs will be presented for the changes from OL baseline.

The above summaries will be repeated for the change from baseline to Week 52 in each of the hormone parameters.

A responder definition for the normalization of testosterone in female subjects will be defined as female subjects with testosterone levels that are greater than or equal to 1.2x ULN (based on age) at OL baseline and decrease to within the normal range at any post-OL baseline visit. Descriptive statistics will be presented by treatment group for the number and percentage of female subjects classified as responders at each visit in the OL period.

A responder definition for the ratio of androstenedione to testosterone (A4/T) in male subjects (Tanner stage 2 [testicular] and above) will be defined as male subjects who have A4/T greater than or equal to 0.5 (based on the pre-glucocorticoid dose values for androstenedione and testosterone) at OL baseline and who achieve A4/T less than 0.5 at any post-OL baseline visit thereafter. Descriptive statistics will be presented by treatment group for the number and percentage of male subjects classified as responders at each visit thereafter in the OL period.

10.2. Salivary Hormone Measurements

Descriptive statistics will be presented by treatment group and visit for the observed and change from OL baseline values for the following salivary hormone measurements in the OL period: 17-OHP, androstenedione, testosterone, progesterone, pregnenolone, and cortisol. Visits where the hormone measurement is collected more than once will be handled as described in Section 9.6.2. Observed and change from timepoint-matched OL baseline will be summarized for pre-dose values and mean 24-hour values. Summaries of testosterone, progesterone, and pregnenolone

will be stratified by sex and Tanner stage (breast or testicular Tanner stages 1, 2 vs. Tanner stages 3-5).

Mean (±SEM) values of the above salivary hormone measurements at each visit in the OL period will be summarized in line graphs by treatment group. Pre-dose values and visit average values will be presented on separate plots. Similar graphs will be presented for the changes from OL baseline.

The above summaries will be repeated for the change from baseline to Week 52 in each of the hormone parameters.

10.3. Glucocorticoid Total Daily Dose

Descriptive statistics will be presented by treatment group and visit for the glucocorticoid total daily dose (mg/m^2/day) observed, change from OL baseline, and percent change from OL baseline values to Week 52. Mean (±SEM) values of the glucocorticoid total daily dose (mg/m^2/day) at each visit will be summarized in line graphs by treatment group. Similar graphs will be presented for the change from OL baseline and percent change from OL baseline. These summaries will be presented for both the glucocorticoid total daily doses *without* the derivation of 0 for loss of androstenedione control as well as the glucocorticoid total daily doses *with* the derivation of 0 for loss of androstenedione control (only for subjects with a decrease in glucocorticoid total daily dose).

The above summaries will be repeated for the change from baseline to Week 52 in glucocorticoid total daily dose.

Similar to the responder analysis defined in Section 9.5.1, a physiologic glucocorticoid dose responder will be defined as the achievement of a reduction in glucocorticoid total daily dose to physiologic levels (≤11 mg/m^2/day hydrocortisone equivalent adjusted for BSA) at OL period visits through Week 52, while androstenedione is adequately controlled at ≤120% of baseline or ≤ULN for age and sex, respectively, at each given visit. Androstenedione control will be assessed based on the androstenedione values at study baseline and each respective OL period visit through Week 52 following the same method as described in Section 6.3.3. Descriptive statistics will be presented by treatment group for the number and percentage of subjects classified as responders and non-responders at each visit in the OL period.

10.4. Metabolic Assessments

Descriptive statistics will be presented by treatment group and visit for the observed and change from OL baseline values for the following metabolic assessments: HOMA-IR, total cholesterol, LDL, triglycerides, and HDL.

The above summaries will be repeated for the change from baseline to Week 52 in each of the metabolic assessments.

10.5. Bone Age

The change from baseline to Week 52 in the ratio of bone age to chronological age (BA:CA) will be summarized with descriptive statistics by treatment group in the subgroup of subjects not at adult height (female subjects with the most recent prior bone age <14 years and males with the

most recent prior bone age <16 years) and not on gonadotropin-releasing hormone agonist therapy, growth hormone therapy, or aromatase inhibitors at study entry. BA:CA will be calculated at baseline and Week 52 by taking the bone age and dividing it by the chronological age of the subject at each respective timepoint.

The change from baseline to Week 52 in the difference in bone age and chronological age (BA-CA) will also be summarized with descriptive statistics by treatment group.

The above summaries will be repeated for the change from OL baseline to Week 52.

10.6. Bayley-Pinneau Predicted Adult Height SDS

The change from baseline in Bayley-Pinneau predicted adult height SDS at Week 52 is a secondary endpoint, which will be analyzed using an ANCOVA model, as described in Section 9.2.1. Bayley-Pinneau predicted adult height SDS will be determined at baseline and Week 52 using the height and bone age of the subject at these respective visits as described in Section 9.6.7.

Descriptive statistics and LS means for change from baseline to Week 52 in Bayley-Pinneau predicted adult height SDS will be presented by treatment group (from the double-blind period) along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

10.7. Growth

In subjects not at adult height (female subjects with the most recent prior bone age <14 years and males with the most recent prior bone age <16 years at study entry) and not on gonadotropin-releasing hormone agonist therapy, growth hormone therapy, or aromatase inhibitors at study entry, height SDS and height velocity (cm/year) will be summarized with descriptive statistics by visit and treatment group.

Height velocity will be calculated in subjects not at adult height. OL baseline will be calculated as the difference in height at Week 28 and height at study baseline, divided by the number of days between both measurements. Height velocity at Week 52 will be calculated as the difference in height at Week 52 and Week 28, divided by the number of days between both measurements. If height is missing at Week 28, the last measured height prior to Week 28 will be used. All height velocity values will be converted to cm/year. Change from OL baseline to Week 52 in height velocity will be summarized with descriptive statistics by treatment group.

10.8. Body Weight and BMI

Descriptive statistics will be presented by treatment group and visit for the observed and change from OL baseline values for body weight SDS and BMI SDS.

The above summaries will be repeated for the change from baseline to Week 52.

In the subset of subjects classified as overweight or obese (BMI $\geq 85^{th}$ percentile at OL baseline), the number and percentage of subjects achieving normal weight (BMI $< 85^{th}$ percentile) at Week 52 will also be presented by treatment group.

10.9. Menstrual Regularity

Menstrual regularity (in female subjects who have undergone menarche, are not on hormonal or intrauterine device contraceptives) will also be summarized for the OL period. As menstrual cycles can be irregular in the first year of menses, the following summaries will be restricted to female subjects in Tanner stages 4 and 5 only. The number and percentage of subjects meeting the definition for having regular menstrual cycles from OL baseline to Week 52 will be summarized by treatment group (where a menstrual cycle is defined as two consecutive calendar days with any amount of flow and menstrual cycle regularity is defined as a menstrual cycle every 21-35 days, translating to approximately 4 to 8 menstrual cycles from OL baseline to Week 52). This summary will be stratified based on menstrual cycle regularity at study entry.

The rate of menstrual cycles will be calculated in female subjects (Tanner stages 4 and 5 at baseline only) assigned to the placebo group for the DB period, who have undergone menarche and are not on hormonal or intrauterine device contraceptives. OL baseline will be calculated as the number of menstrual cycles from study baseline to Week 28, divided by the number of days in the DB period. The rate of menstrual cycles at Week 52 will be calculated as the number of menstrual cycles from Week 28 to Week 52, divided by the number of days in the OL period. All rates of menstrual cycles will be converted to menstrual cycles/year. Change from OL Baseline to Week 52 in rate of menstrual cycles will be summarized with descriptive statistics by treatment group.

10.10. TARTs

Change from baseline to Week 52 in TARTs will also be summarized for the OL period in male subjects with TARTs at baseline. For the purpose of summarization, the volume of each lesion will be used to assess size (in cubic millimeters) at visits where the testicular ultrasound is performed. Volume will be calculated using the following formula: volume = length x width x depth x 0.5236. The volume of each testicle will be calculated using the same formula. Descriptive statistics will be used to summarize the total TART volume (calculated as the sum of the volume of each lesion across both testes) for each subject by treatment group and visit. Observed and change from baseline values will be included in the summary table.

Additionally, the total TART volume will be expressed as a percentage of the total testicular volume and calculated as the sum of the volume of each lesion across both testes divided by the total testicular volume across both testes for each subject. The total TART volume as a percentage of the total testicular volume will be summarized with descriptive statistics by treatment group and visit. Observed and change from baseline values will be included in the summary table.

The number and percentage of subjects with the presence of TARTs, whether they are unilateral or bilateral, and the average number of TARTs per subject will be summarized with descriptive statistics by visit and treatment group. The number and percentage of TARTs with each shape, boundary, hypervascularity, calcification, and echogenicity will be summarized with descriptive statistics by treatment group.

A response assessment will be conducted at Week 52 for subjects with the presence of TARTs at baseline. Response categories and the criteria for each category are detailed in Table 7 within

Section 9.6.4. The frequency and percentage of subjects meeting each classification at Week 52 (as compared to baseline) will be summarized by treatment group.

10.11. Hirsutism and Acne

Visual analog scales (VAS) were used to assess the subject's perception of severity of hirsutism (female subjects only) and acne (all subjects), scored on a 100 mm visual analog scale from 0 mm (no symptoms) to 100 mm (very severe symptoms). Observed and change from OL baseline values in VAS scores for hirsutism and acne will be summarized descriptively by visit and treatment group. Summaries of acne VAS scores will be presented by sex as well as both sexes combined.

The above summaries will be repeated for the change from baseline to Week 52 in VAS scores for hirsutism and acne.

11. SAFETY

The safety objective of the study is to characterize the safety and tolerability profile of crinecerfont as measured by TEAEs (including SAEs), clinical laboratory tests, vital signs, ECG, and neuropsychiatric assessments. All outputs for safety endpoints will be based on the DB Period Safety Analysis Set during the DB period and on the OL Period Safety Analysis Set during the OL period. Safety data will be summarized descriptively by treatment group. Safety data will not be subject to any imputation and will be summarized on an observed-case basis. No formal hypothesis-testing analysis of safety data will be performed. Unless otherwise noted, the below summaries will be produced for the DB period and the OL period separately and will follow the treatment-period specific treatment group and baseline definitions (as described in Section 6). All by-visit safety tables will also include a safety follow-up visit summary (as defined in Table 5) and subject data collected after the safety follow-up visit will not be summarized.

11.1. Study Drug Dosing and Compliance

The duration of exposure to crinecerfont will be calculated as: last dose date – first dose date +1.

Duration of exposure to crinecerfont will be summarized with descriptive statistics over the DB and OL periods (ie, table will include cumulative exposure from both treatment periods). The number and percentage of subjects with the following exposure categories will also be presented:

- > 0 to < 4 weeks
- ≥ 4 weeks to < 16 weeks
- ≥ 16 weeks to < 28 weeks
- ≥ 28 weeks to < 40 weeks
- ≥ 40 weeks to < 52 weeks
- ≥ 52 weeks

At each visit where study drug is dispensed, the site will enter the kit number(s) dispensed and returned as well as the total number of capsules dispensed and returned for the kits combined. The subject's compliance will be calculated at each visit (when kits are returned) as the total number of capsules dispensed – total capsules returned divided by the expected number of capsules taken by the subject during the visit interval. For subjects receiving the oral solution, compliance will be determined based on the weight of the dispensed and returned study drug. The number and percentage of subjects in each treatment group who are dosing compliant (defined as $< 80\%$, $\geq 80\%$ to $\leq 120\%$, $> 120\%$) will be presented for each postbaseline visit when study drug is returned. Compliance will be summarized separately for the DB and OL periods.

11.2. Adverse Events

A treatment-emergent adverse event (TEAE) is an AE not present prior to the initiation of study drug dosing, or is an already present event that worsens either in intensity or frequency following the initiation of study drug dosing with an onset on or before the last dose of study drug + 28 days (to account for the study drug washout period).

TEAEs will be further defined for the two treatment periods. For the DB period, Investigators will be asked to respond "Yes" or "No" on the eCRF as to whether the AE started after the subject took the first dose of study drug. An AE with a response of "Yes" and that occurs on or before the last dose of study drug in the DB period will be classified as a TEAE for the DB period. If the subject's last dose on study occurs in the DB period, then any AE occurring prior to or on the last dose of study drug in the DB period + 28 days will be considered a TEAE. If the investigator's response is missing, then the treatment emergent status will be derived based on the AE onset date relative to the date of the subject's first dose of study drug in the DB period or first dose of study drug in the OL period. If the AE onset date is unknown, it will be assumed that the AE is a TEAE and will be reported in the DB period TEAE summaries.

For the OL period, TEAEs will be defined as AEs starting after the first dose of study drug in the OL period and prior to the last dose of study drug in the OL period. If the subject's last dose on study occurs in the OL period, then any AE occurring prior to or on the last dose of study drug in the OL period + 28 days will be considered a TEAE.

TEAEs that are ongoing at the time of a datacut will be summarized once according to the treatment period in which the TEAE started.

Adverse events with an onset date after the last dose of study drug + 28 days will not be considered treatment emergent (non-TEAEs).

The following TEAE summaries will be produced for the DB period (Day 1 up to Week 28) and for the 6-month OL period (Week 28 to Week 52) separately:

- TEAEs categorized by MedDRA system organ class (SOC) and/or preferred term (PT) will be summarized in frequency tables. The frequency tables will include the number and percentage of unique subjects experiencing each event one or more times by treatment group.
- Two similar frequency tables will be generated for TEAEs—one including both SOC and PT (sorted in alphabetical order), and one including PT only, with PTs sorted in order of decreasing frequency of subjects reporting the PT in the DB crinecerfont treatment group. Non-TEAEs will be summarized by SOC and PT only.
- An overall summary table will be provided that summarizes the number and percentage of unique subjects with any TEAE, any SAE, any TEAE leading to study drug discontinuation, any TEAE leading to study discontinuation, and any TEAE resulting in death. The summary table will also include the frequency distribution of the maximum TEAE intensity (mild, moderate, severe) reported for each subject.

The above summaries (TEAEs by SOC / PT, TEAEs by PT) will be generated for the following subgroups as well:

- age group (2 - < 12 years; 12 – 17 years)
- Tanner stage (Tanner stage 1, 2; Tanner stages 3-5).

11.2.1. Adverse Events Resulting in Discontinuation from Study Drug

The number and percentage of subjects with a TEAE resulting in study drug discontinuation will be presented by PT within SOC (presented in the same method as the analogous primary TEAE

table) by treatment group. More than one AE can contribute to study drug discontinuation per subject. The first line of each table will display the number and percentage of subjects with at least one TEAE leading to study drug discontinuation. This summary will be produced for both the DB period and OL period separately.

A listing of TEAEs resulting in premature study drug discontinuation will be provided, which includes subject ID, treatment period, treatment group, study day of the study drug discontinuation, and other relevant information from the AE eCRF.

11.2.2. Deaths and Other Serious Adverse Events

Summary tables of treatment-emergent serious adverse events (SAEs) will be presented. The number and percentage of subjects with an SAE will be presented by PT within SOC (each sorted in alphabetical order) by treatment group. The first line of the table will display the number and percentage of subjects with at least one SAE. This table will be produced for both the DB period and the OL period separately.

Separate listings of SAEs and fatal TEAEs will also be provided. Each listing will include subject ID, treatment period, treatment group, study day of the SAE or fatal TEAE, and any additional relevant information from the AE eCRF.

11.2.3. Adverse Events of Acute Adrenal Insufficiency

In addition to the TEAE summaries described above, adverse events of acute adrenal insufficiency will be presented in a table by PT and treatment group. This table will be produced for the DB period and the OL period separately.

The following PT (MedDRA, version 26.0) will be used to identify AEs of acute adrenal insufficiency/adrenal crisis:

- adrenocortical insufficiency acute

11.2.4. Adverse Events Leading to Glucocorticoid Stress Dosing

The number and percentage of subjects with a TEAE leading to glucocorticoid stress dosing will be presented by PT within SOC (presented in the same method as the analogous primary TEAE table) by treatment group. The first line of each table will display the number and percentage of subjects with at least one TEAE leading to glucocorticoid stress dosing. This summary will be produced for both the DB period and OL period separately.

An overall summary table will be provided that summarizes the maximum severity of glucocorticoid stress dosing based on the following categories (in order of least to most severe):

- Oral dosing (alone, without parenteral dosing/ER visit/SAE)
- Parenteral dosing (without ER visit/SAE)
- ER visit without SAE (requiring glucocorticoid stress dosing)
- SAE (requiring glucocorticoid stress dosing)

Each subject will be counted once by their maximum severity of illness requiring glucocorticoid stress dosing. This table will be produced for the DB period and the OL period separately.

11.2.5. Exposure-Adjusted Subject Incidence Rate

Exposure-adjusted subject incidence rate will be defined as the number of unique subjects experiencing the TEAE one or more times (numerator) divided by the sum of the time at risk for the first TEAE across all subjects (denominator). For subjects experiencing the TEAE, the time at risk (exposure time) is the time from the first dose to the date of the first occurrence of the TEAE. For subjects not experiencing the TEAE, exposure time is the time from the first dose to the last dose + 28 days (or data cutoff). Note that exposure time will vary across subjects for the same TEAE. The exposure-adjusted subject incidence rate in the crinecerfont and placebo groups, over the DB and OL periods combined, will be summarized by descending preferred term in the crinecerfont group.

11.3. Clinical Laboratory Data

The hematology, clinical chemistry, coagulation, PRA, TSH, and free T4 data will be summarized with descriptive statistics by treatment group and visit. Both observed values and changes from baseline will be summarized for the DB period and OL period separately.

Shift tables will be presented for selected clinical laboratory variables based on the reference range-based categories of "Low," "Normal," or "High." A clinical laboratory variable value will be assigned to one of these three categories according to the reference ranges provided by the central clinical laboratory.

A table of shifts from baseline to Week 28 will be presented within the double-blind treatment period. Each shift table will have four rows and four columns, with rows reflecting the reference range category at baseline (or missing if applicable), and columns reflecting the reference range category at the specified postbaseline visit (or missing if applicable). A "Total" row and "Total" column will also be included. The number and percentage of subjects in each shift category will be displayed in the table; percentages will be based on the number of subjects included in the table. Two additional shift tables will be presented within the OL period: shifts from baseline to Week 52 and shifts from OL baseline to Week 52.

Shift tables will be presented for the following clinical laboratory variables:

- aspartate aminotransferase (AST),
- alanine aminotransferase (ALT),
- alkaline phosphatase (ALP),
- total bilirubin,
- creatine kinase,
- creatinine,
- white blood cell count,
- absolute neutrophil count,
- hemoglobin,
- platelet count,
- TSH,

- Free T4, and
- PRA.

11.4. Vital Signs

The vital signs data (blood pressure [BP], pulse rate, respiratory rate, and body temperature) will be summarized with descriptive statistics by treatment group and visit. Sitting blood pressure and heart rate measurements were collected in triplicate at each visit throughout the study and the average of the triplicate values will be used for the purpose of summarization. Observed and change from baseline values for pulse rate, respiratory rate, and body temperature will be summarized for the DB period and OL period separately.

Blood pressure data will be summarized categorically based on the guidance published by Flynn et al (Flynn 2017) for subjects with normal blood pressure, elevated blood pressure, stage 1 HTN (hypertension), and stage 2 HTN (as defined in Table 8 below). Percentiles for systolic blood pressure and diastolic blood pressure will be calculated based on the boys and girls tables (by age and height percentile) as published by Flynn et al (Flynn 2017). The number and percentage of subjects in each of the following blood pressure categories will be presented by treatment group and visit by age category as well as all ages combined:

Table 8: Blood Pressure Categories by Age Group

	Ages 2 - < 13 years old	**Subjects ≥ 13 years old**
Normal BP	< 90th percentile	<120/<80 mm Hg
Elevated BP	≥ 90th percentile to < 95th percentile or 120/80 mm Hg to < 95th percentile (whichever is lower)	120/<80 to 129/<80 mm Hg
Stage 1 HTN	≥ 95th percentile to < 95th + 12 mm Hg or 130/80 mm Hg to 139/89 mm Hg (whichever is lower)	130/80 to 139/89 mm Hg
Stage 2 HTN	≥ 95th percentile + 12 mm Hg or 140/90 mm Hg (whichever is lower)	≥ 140/90 mm Hg

Flynn et al., 2017.

The same categorical summaries will be presented in frequency tables for the maximum value at any postbaseline visit within the treatment period (including unscheduled and early termination visits). Each subject will be counted once by their maximum postbaseline blood pressure category. The above categories will be summarized for both the DB and OL periods.

11.5. 6- or 12-lead Electrocardiogram

Descriptive statistics for the observed values and changes from baseline will be presented for each of the ECG parameters (heart rate, PR interval, QRS duration, QT interval, QTcF) by visit and treatment group.

Frequency tables (number and percentage of subjects) will be presented for the investigator overall categorical assessment (ie, Normal, Abnormal not Clinically Significant, or Abnormal Clinically Significant) by visit and treatment group.

Two additional categorical summaries (frequency tables displaying number and percentage of subjects) will be presented for the QTcF interval by treatment group. For the first summary, the observed QTcF values at each visit will be classified as follows:

- Greater than 450 msec
- Greater than 480 msec
- Greater than 500 msec

For the second categorical summary, the changes from baseline to each subsequent visit will be classified as follows:

- Increase greater than 30 msec
- Increase greater than 60 msec

The same categorical summaries will be presented in frequency tables for values at any postbaseline visit (including unscheduled and early termination visits) meeting the aforementioned criteria.

The above ECG summaries will be produced for the DB period and the OL period separately.

11.6. Columbia-Suicide Severity Rating Scale

The C-SSRS is a validated instrument to prospectively assess suicidal ideation and behavior and will be used only for subjects ≥ 6 years of age. The C-SSRS data will be presented in the following summaries:

- Screening/lifetime assessment
- Screening/past 6 months (suicidal ideation items); past 1 year (suicidal behavior items) assessment
- Baseline (Day 1) assessment
- Postbaseline assessments

Each summary will display the number and percentage of subjects who report "Yes" to specific C-SSRS items or categories of items (a category is assigned a "Yes" value if a "Yes" is reported for any item in the category). These C-SSRS items and categories are as follows:

- Suicidal Ideation Items
  - (1) Wish to be dead
  - (2) Non-specific active suicidal thoughts
  - (3) Active suicidal ideation with any methods (not plan) without intent to act
  - (4) Active suicidal ideation with some intent to act, without specific plan
  - (5) Active suicidal ideation with specific plan and intent
- Suicidal Ideation Category: Any of items (1) through (5)
- Suicidal Behavior Items
  - (6) Preparatory acts or behavior

(7) Aborted attempt
(8) Interrupted attempt
(9) Non-fatal suicide attempt
(10) Completed suicide ("Since Last Visit" assessments only)

- Suicidal Behavior Category: Any of items (6) through (10)
- Suicidal Ideation or Behavior Category: Any of items (1) through (10)

For the "all postbaseline assessments" summary, each subject's C-SSRS responses for all postbaseline assessments will be evaluated, and a "Yes" response for any assessment will be considered as a "Yes" for the subject.

In addition to the summaries described above, shift tables comparing postbaseline suicidal ideation scores to baseline scores will be presented. The shift table scores are defined as the following:

0 = No suicidal ideation
1 = Wish to be dead
2 = Non-specific active suicidal thoughts
3 = Active suicidal ideation with any methods (not plan) without intent to act
4 = Active suicidal ideation with some intent to act, without specific plan
5 = Active suicidal ideation with specific plan and intent

The shift tables will display the number and percentage of subjects within each cell of a 6 x 6 table for each treatment group, with the rows representing the baseline score and the columns representing the maximum score recorded across all postbaseline assessments (including both scheduled and unscheduled visits). Subjects missing either a baseline score or all postbaseline scores will not appear in the table.

The above C-SSRS summaries will be produced for the DB period and the OL period separately.

11.7. Brief Psychiatric Rating Scale for Children (BPRS-c)

The Brief Psychiatric Rating Scale for Children (BPRS-c) is a clinician-rated tool designed to assess the severity of psychopathology in subjects with schizophrenia and other psychotic disorders. The BPRS-c is a 21-item, clinician-based rating scale designed for use in evaluating psychiatric problems of children and adolescents. The severity of each item of the BPRS-c is rated on a scale of 0 (not present) to 6 (extremely severe) (total score range: 0 to 126). Higher scores represent greater symptom severity. The BPRS-c will be administered only to subjects ≥3 years of age. Items that are not scored are left blank (ie, not scored as "0"). Rather than using the EDC total score value, the total score for BPRS-c will be re-calculated as the sum of ratings for each of the 21 items. Missing items will be imputed based on the average score of the other items at that assessment; however, if >50% of the items are missing for that subject/visit, then the total score will be set to missing.

BPRS-c total score observed and change from baseline values will be summarized with descriptive statistics by visit and treatment group. This summary will be produced for the DB period and the OL period separately. The study baseline will be used for the DB period summaries while the OL baseline will be used for the OL period summaries.

11.8. Prior and Concomitant Medications

Prior medications and concomitant medications, including glucocorticoids and mineralocorticoids, will be summarized by World Health Organization (WHO) Drug Anatomical Therapeutic Chemical Classification (ATC) Level 3 category (or Level 2 if there is not an applicable Level 3 category) and preferred name. Concomitant medications will be summarized separately for the DB and OL periods.

Medications will be assigned to study periods (prior, concomitant in the DB period, and/or concomitant in the OL period) based on the medication start and stop dates relative to study drug dosing in the DB and OL periods. Medications that were started and stopped prior to the date of the first dose of study drug in the DB period will be assigned as prior. Medications that had a stop date or were ongoing during the DB period will be assigned as concomitant in the DB period. Medications that had a stop date or were ongoing during the OL period will be assigned as concomitant in the OL period.

Within each summary, the number and percentage of subjects using medications in each WHO Drug ATC category (Level 3/preferred name) will be summarized by treatment group. A subject may take the same medication more than once or multiple medications for a subject may be classified under the same ATC level or preferred name. A subject is counted only once for each level of medication classification within a summary. The tables will be sorted alphabetically by ATC code.

12. PATIENT AND CAREGIVER REPORTED OUTCOMES

Patient and caregiver reported outcomes capture the status of a subject's health condition and come directly from the subject or caregiver, without interpretation of the subject's or caregiver's response by a clinician or anyone else. The following section describes the assessments that were administered to subjects and caregivers throughout the study and the methods of scoring and summarizing these data. All outputs for the summaries will be based on the Full Analysis Set for the DB period and on the OL Period Safety Analysis set for the OL period. The analysis of the data will be based on descriptive statistics and presented by treatment group according to the study visit unless otherwise noted. These data will not be imputed. No formal hypothesis-testing analysis of patient and caregiver reported outcomes data will be performed. Unless otherwise noted, the below summaries will be produced for the DB period and the OL period separately and will follow the treatment-period specific treatment group and baseline definitions (as described in Section 6) where the study baseline will be used for the DB period summaries and the OL baseline will be used for the OL period summaries.

12.1. EuroQoL 5 Dimensions (EQ-5D) and Visual Analog Scale (EQ-VAS)

The EQ-5D-Y was administered to subjects 8 to 15 years of age. This instrument has subjects self-report on 5 dimensions of health: Mobility, Looking After Myself, Usual Activities, Pain/Discomfort, and being Worried, Sad, or Unhappy. Each dimension has 3 levels of severity: no problems, some problems, or a lot of problems. The subject indicates his/her health state by checking the box next to the most appropriate statement. The number and percentage of subjects in each of the 3 levels for each dimension at baseline and each postbaseline visit will be summarized by treatment group.

The EQ-5D-5L was administered to subjects ≥ 16 years of age. The EQ-5D-5L is a general, single index measure for describing and valuing health. It defines health in terms of 5 dimensions: Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The subject indicates his/her health state by checking the box next to the most appropriate statement. Number and percentage of subjects in each of the 5 levels for each dimension at baseline and each postbaseline visit will be summarized by treatment group.

For both the EQ-5D-Y and the EQ-5D-5L, subjects rated their overall health on a 0 to 100 hash-marked, vertical visual analogue scale (EQ-VAS). The endpoints are labeled 'The best health you can imagine' (100) and 'The worst health you can imagine' (0). The EQ-VAS score observed and change from baseline values will be summarized at baseline and each postbaseline visit by treatment group and age subset (ie, 8 to 15 years of age vs. ≥ 16 years of age).

The above summaries will be produced for the DB period and the OL period.

12.2. Pediatric Quality of Life Instrument (PedsQL)

The 23-item Pediatric Quality of Life Instrument (PedsQL™) Generic Core Scales encompass: 1) Physical Functioning (8 items), 2) Emotional Functioning (5 items), 3) Social Functioning (5 items), and 4) School Functioning (5 items) (Varni et al., 2001). The instructions ask how much of a problem each item has been during the past one month. A 5-point Likert response scale is used (0=never a problem; 1=almost never a problem; 2=sometimes a problem; 3=often a

problem; 4=almost always a problem). For ages 5 to 7, the response scale is reworded and simplified to a 3-point scale (0=not at all a problem; 2=sometimes a problem; 4=a lot of a problem), with each response choice anchored to a happy-to-sad faces scale. Items are reverse-scored and linearly transformed to a 0 to 100 scale, so that higher scores indicate better health-related quality of life. The mean score will be calculated as the sum of the items divided by the number of items answered. If more than 50% of the items in the scale are missing, the scale scores will not be computed. If 50% or more items are completed, the mean of the completed items in the scale will be calculated. The psychosocial health summary score will be calculated as the sum of the items divided by the number of items answered in the emotional, social, and school functioning scales. The physical health summary score will be calculated as the physical functioning scale score. The total score will be calculated as the sum of all the items divided by the number of items answered on all the scales.

The psychosocial health summary score, the physical health summary score, and the total score observed and change from baseline values will be summarized with descriptive statistics by visit and treatment group. This summary will be produced for the DB period and the OL period. Subject and parent reports will be summarized separately.

12.3. Pediatric Quality of Life Family Impact Module

The PedsQL Family Impact Module is a 36-item instrument that include 6 scales measuring parent/caregiver self-reported functioning: 1) Physical Functioning (6 items), 2) Emotional Functioning (5 items), 3) Social Functioning (4 items), 4) Cognitive Functioning (5 items), 5) communication (3 items), 6), Worry (5 items), and 2 scales measuring parent-reported family functioning; 7) daily Activities (3 items) and 8) Family Relationships (5 items) (Varni et al., 2004). A 5-point response scale is used (0=never a problem; 4=always a problem). Items are reverse-scored and linearly transformed so that higher scores indicate better functioning. The mean score will be calculated as the sum of the items divided by the number of items answered. If more than 50% of the items in the scale are missing, the scale scores will not be computed. If 50% or more items are completed, the mean of the completed items in the scale will be calculated. The parent HRQL summary score (20 items) will be computed as the sum of the items divided by the number of items answered in the physical, emotional, social, and cognitive function scales. The family functioning summary score (8 items) will be computed as the sum of the items divided by the number of items answered in the daily activities and family relationships scales. The total score will be calculated as the sum of all 36 items divided by the number of items answered.

The parent HRQL summary score, the family functioning summary score, and the total score observed and change from baseline values will be summarized with descriptive statistics by visit and treatment group. This summary will be produced for the DB period and the OL period.

13. OTHER DATA

13.1. Ease of Administration and Palatability of Study Drug – DB Period

The acceptability/ease of administration and palatability of the study drug at Week 4 is a secondary endpoint and was assessed via questionnaire. Palatability of the study drug was assessed by asking subjects ≥ 6 years of age to rate the taste and smell of the study drug, as well as how easy the study drug was to take, on a 5-point hedonic scale, with 5 being a favorable response and 1 being an unfavorable response. Each response was anchored to a face on a happy-to-sad scale. For subjects 2 to 5 years of age, caregivers were asked to rate the ease of administration of the study drug both on that day and overall on a 5-point hedonic scale with 5 indicating relative ease and 1 indicating relative difficultly.

The three questions related to ease of administration, taste, and smell will be summarized on the Full Analysis Set with descriptive statistics by treatment group, formulation type (oral solution vs. capsule), and age group (> 2 to < 6 years; ≥ 6 years) at Week 4.

13.2. Ease of Administration and Palatability of Study Drug – OL Period

The three questions related to ease of administration, taste, and smell (as described in Section 13.1) will be summarized with descriptive statistics on the OL Period Safety Analysis Set by treatment group, visit (Week 28 [OL baseline] and Week 52), formulation type (oral solution vs. capsule), and age group (> 2 to < 6 years; ≥ 6 years).

14. DEVIATIONS FROM PROTOCOL PLANNED ANALYSIS

This section will document any deviations from the protocol planned analysis along with any assessments that were performed on study (and reviewed by the study site) but will not be summarized.

The following assessments were performed on study but will not be summarized:

- Physical examination results (excluding height, weight, and Tanner staging)
- CYP21A2 genotyping
- Pregnancy test
- Urine drug screen
- Urinalysis
- Glucocorticoid daily dose from eDiary

The following list includes all deviations from the protocol planned analysis:

- The Safety Analysis Set definition in the protocol specified that the subject must have postbaseline safety data during the double-blind placebo-controlled period. This part of the definition was removed from the SAP due to the potential issue of having a subject dosed with study drug and then becoming lost to follow-up.
- The Safety Analysis Set was split up into the DB Period Safety Analysis Set and the OL Period Safety Analysis Set.

15. PERFORMANCE QUALIFICATION OF SAS® PROGRAMS

The analysis and summary of data from this study will be performed using SAS® 9.4 (or a later release if available). All SAS® programs used in the production of statistical analyses, tables, listings, and figures described in this SAP will undergo performance qualification (verification that the program produces the intended output) in accordance with department standard operating procedures. The performance qualification may include independent programming and/or peer review of the SAS® log files. In addition, tables, figures, listings, and statistical analysis output will be independently reviewed for completeness and accuracy.

16. REFERENCES

Bayley N, Pinneau SR. Tables for predicting adult height from skeletal age: revised for use with the Greulich-Pyle hand standards. J Pediatr. 1952 Apr;40(4):423-41.

Dubois D, Dubois EF. A formula to estimate the approximate surface area if height and weight be known. Arch Intern Med. 1916; 17:863-871.

European Medicines Agency. Points to consider on implications of Coronavirus disease (COVID-19) on methodological aspects of ongoing clinical trials. March 2020.

Flynn JT, Kaelber DC, Baker-Smith CM, et al. Clinical Practice Guideline for Screening and Management of High Blood Pressure in Children and Adolescents. Pediatrics 2017; 140(3):e20171904.

Kuczmarski RJ, Ogden CL, Guo SS, et al. 2000 CDC growth charts for the United States: Methods and development. National Center for Health Statistics. Vital Health Stat 11(246). 2002.

O'Kelly M, Ratitch B. Clinical Trials with Missing Data: A Guide for Practitioners. West Sussex, United Kingdom: John Wiley & Sons, Ltd. 2014; 292-297, 309-310.

United States Food and Drug Administration. Conduct of Clinical Trials of Medical Products During the COVID-19 Public Health Emergency. Guidance for Industry, Investigators, and Institutional Review Boards. March 2020, updated January 2021.

Varni JW, Seid M, Kurtin PS. PedsQL™ 4.0: Reliability and Validity of the Pediatric Quality of Life Inventory Version 4.0 Generic Core Scales in Healthy and Patient Populations. Medical Care. 2001; 39(8):800-12.

17. APPENDICES

17.1. Implementation of Multiple Imputation for the Primary Endpoint and First Key Secondary Endpoint

For the primary and first key secondary endpoint that require multiple imputation, the following steps will be taken.

1) For subjects on study drug and are missing the endpoint at the timepoint of interest, impute missing data using a regression-based multiple imputation model based on data from subjects with non-missing endpoint data at the timepoint of interest within the same treatment group. The model will include the baseline value for the endpoint, sex, and pubertal stage.

2) For subjects in the crinecerfont treatment group off study drug prior to timepoint of interest and are missing the endpoint at the timepoint of interest, impute missing data using a regression-based imputation model based on data from Retrieved Dropouts. If there are an insufficient number of Retrieved Dropouts, the endpoint will be imputed from subjects in the placebo treatment group. For similar subjects in the placebo treatment group, missing data will be imputed using non-missing data from subjects in the placebo treatment group. The model will include the baseline value for the endpoint, sex, and pubertal stage.

Note that for the regression-based multiple imputation models described above, the values recorded on the eCRF will be used instead of those recorded at the time of randomization in the IRT.

The number of imputed datasets created from the steps above for each endpoint is 100. After the multiply-imputed datasets have been created, each of the datasets will be analyzed using the ANCOVA model described in Section 9.2.1. The results of these analyses will be combined using PROC MIANALYZE.

17.2. Implementation of Multiple Imputation for Second Key Secondary Endpoint

The second key secondary endpoint requires the availability of both the glucocorticoid dose at Week 28 as well as androstenedione values used in the assessment of androstenedione control (as defined in Section 6.3.5). For subjects with incomplete or missing data at Week 28 (ie, missing glucocorticoid dose, androstenedione, or both), the following methods will be used to multiply impute the missing data for use in the analysis of the endpoint. The methods of handling each case of missing data are described in text below.

Subjects with a missing glucocorticoid dose, androstenedione, or both at Week 28 will have their data multiply imputed using the fully conditional specifications (FCS) method. This method will be implemented in SAS® 9.4 PROC MI using a series of conditional models (one for each incomplete variable at Week 28). First, the missing Week 28 pre-GC dose androstenedione values will be imputed using a regression-based multiple imputation model and the model will include baseline androstenedione (pre-GC dose), age, and sex (as specified in the eCRF). The subsequent FCS regression model will impute Week 28 glucocorticoid total daily doses (in $mg/m^2/day$) and the model will include baseline glucocorticoid dose ($mg/m^2/day$), baseline androstenedione (pre-GC dose), Week 28 androstenedione (pre-GC dose), and sex (as specified

in the eCRF). Note that a minimum of 0 is set for imputed Week 28 pre-GC dose androstenedione values and the Week 28 glucocorticoid daily dose (mg/m^2/day) as it is biologically implausible for either of these values to be less than 0. A random-number generator seed value of 1234 will be used. For subjects missing androstenedione at Week 28, androstenedione "control" will be determined based on the imputed Week 28 androstenedione value and the percent change from baseline to Week 28 in glucocorticoid total daily dose will be set to 0 if the imputed androstenedione value does not indicate "control." The number of imputed datasets created from the steps above is 100.

The above multiple imputation procedure will be conducted in more than one PROC MI step so that subjects with missing data are correctly imputed from subjects with complete data based on the rules defined below:

- For subjects who are on study drug at Week 28 and are missing one or both components of the primary endpoint, their missing Week 28 data will be imputed from subjects within the same treatment group with non-missing data at Week 28 (Week 28 Completers).
- For subjects who are off study drug at Week 28 and are missing one or both components of the primary endpoint, their missing Week 28 data will be imputed from subjects who are classified as Retrieved Dropouts (Section 6.4). If there are an insufficient number of Retrieved Dropouts, the missing endpoint values will be imputed from subjects in the placebo treatment group. For subjects in the placebo treatment group, missing data will be imputed using non-missing data from subjects in the placebo treatment group.

The multiply-imputed datasets generated from the steps above will be set back together prior to the final analysis steps.

After the multiply-imputed datasets have been created, each of the 100 datasets will be analyzed using the ANCOVA model described in Section 9.2.1. The results of these analyses will be combined using PROC MIANALYZE.

17.3. Implementation of Tipping Point Analysis for the Primary Endpoint

A tipping point sensitivity analysis will be performed to assess the robustness of the missing data assumptions for the primary endpoint. The following steps will be taken to implement the tipping point analysis:

1) For the 100 imputed datasets created (as described in Section 17.1), if a subject was originally missing a Week 4 androstenedione value, add *delta* (see rules below) to the imputed Week 4 androstenedione value. Calculate the change from baseline to Week 4 in androstenedione for all subjects:
    - If Week 4 androstenedione is non-missing: change from baseline = Week 4 androstenedione value – baseline androstenedione value;
    - If Week 4 androstenedione is missing and the subject is in either the crinecerfont treatment group or placebo treatment group: change from baseline = Week 4 delta-adjusted imputed androstenedione value – baseline androstenedione value.
2) Analyze the change from baseline to Week 4 in androstenedione in each of the 100 imputed datasets using ANCOVA (as specified Section 9.2.1) using PROC MIXED. Use

PROC MIANALYZE to combine the results and obtain the LS mean statistics (within treatment, and treatment differences).

3) If the p-value in the previous step was not significant (ie, >0.05), then no additional analyses are needed. If the p-value was significant, repeat the previous steps using the next values of *delta*. Continue to do this until significance is lost, or all combinations of deltas (defined for each treatment group) have been analyzed.

The following rules will be used for defining *delta*:

- The initial delta will be = 0 (ie, an analysis of the original, multiply imputed datasets)
- For subjects in the crinecerfont treatment group, the second delta will be = 0.1745 nmol/L. If this is significant, continue to increase delta by 0.1745 until significance is lost.
- For subjects in the placebo treatment group, the second delta will be = -0.1745 nmol/L. Delta will be decreased by 0.1745 until significance is lost.

Delta may be adjusted to approach insignificance faster due to the wide range of values.

17.4. Implementation of Tipping Point Analysis for the First Key Secondary Endpoint

A tipping point sensitivity analysis will be performed to assess the robustness of the missing data assumptions for the first key secondary endpoint. The following steps will be taken to implement the tipping point analysis:

1) For the 100 imputed datasets created (as described in Section 17.1), if a subject was originally missing a Week 4 17-OHP value, add *delta* (see rules below) to the imputed Week 4 17-OHP value. Calculate the change from baseline to Week 4 in 17-OHP for all subjects:
    - If Week 4 17-OHP is non-missing: change from baseline = Week 4 17-OHP value – baseline 17-OHP value;
    - If Week 4 17-OHP is missing and the subject is in either the crinecerfont treatment group or placebo treatment group: change from baseline = Week 4 delta-adjusted imputed 17-OHP value – baseline 17-OHP value.
2) Analyze the change from baseline to Week 4 in 17-OHP in each of the 100 imputed datasets using ANCOVA (as specified Section 9.2.1) using PROC MIXED. Use PROC MIANALYZE to combine the results and obtain the LS mean statistics (within treatment, and treatment differences).
3) If the p-value in the previous step was not significant (ie, >0.05), then no additional analyses are needed. If the p-value was significant, repeat the previous steps using the next values of *delta*. Continue to do this until significance is lost, or all combinations of deltas (defined for each treatment group) have been analyzed.

The following rules will be used for defining *delta*:

- The initial delta will be = 0 (ie, an analysis of the original, multiply imputed datasets)

- For subjects in the crinecerfont treatment group, the second delta will be = 15.15 nmol/L. If this is significant, continue to increase delta by 15.15 until significance is lost.
- For subjects in the placebo treatment group, the second delta will be = -15.15 nmol/L. Delta will be decreased by 15.15 until significance is lost.

Delta may be adjusted to approach insignificance faster due to the wide range of values.

17.5. Implementation of Tipping Point Analysis for the Second Key Secondary Endpoint

A tipping point sensitivity analysis will be performed to assess the robustness of the missing data assumptions for the second key secondary endpoint. The following steps will be taken to implement the tipping point analysis:

1) For the 100 imputed datasets created (as described in Section 17.2), if a subject was originally missing the Week 28 glucocorticoid daily dose value (mg/m^2/day), add *delta* (see rules below) to the imputed Week 28 glucocorticoid daily dose value. Calculate the percent change from baseline to Week 28 in the glucocorticoid daily dose for all subjects:
    - If Week 28 glucocorticoid daily dose is non-missing: percent change from baseline = (Week 28 glucocorticoid daily dose value – baseline glucocorticoid daily dose value) / baseline glucocorticoid daily dose value x 100;
    - If Week 28 glucocorticoid daily dose is missing and the subject is in either the crinecerfont treatment group or placebo treatment group: percent change from baseline = (Week 28 delta-adjusted imputed glucocorticoid daily dose value – baseline glucocorticoid daily dose value) / baseline glucocorticoid daily dose value x 100.
2) If the subject does not have androstenedione control at Week 28 (as defined in the second key secondary endpoint) then the percent change from baseline will be set to 0 prior to the analysis of the delta-adjusted imputed datasets.
3) Analyze the percent change from baseline to Week 28 in glucocorticoid daily dose in each of the 100 imputed datasets using ANCOVA (as specified Section 9.2.1) using PROC MIXED. Use PROC MIANALYZE to combine the results and obtain the LS mean statistics (within treatment, and treatment differences).
4) If the p-value in the previous step was not significant (ie, >0.05), then no additional analyses are needed. If the p-value was significant, repeat the previous steps using the next values of *delta*. Continue to do this until significance is lost, or all combinations of deltas (defined for each treatment group) have been analyzed.

The following rules will be used for defining *delta*:

- The initial delta will be = 0 (ie, an analysis of the original, multiply imputed datasets)
- For subjects in the crinecerfont treatment group, the second delta will be = 1. If this is significant, continue to increase delta by 1 until significance is lost.
- For subjects in the placebo treatment group, the second delta will be = -1. Delta will be decreased by 1 until significance is lost.

Delta may be adjusted to approach insignificance faster due to the wide range of values.