STATISTICAL ANALYSIS PLAN
PHASE 3

Version:	3.0
Date:	*10 August, 2023*
Based on:	Protocol Amendment No. 5: 30 November 2021 *Protocol Amendment No. 4: 07 September 2021* *Protocol Amendment No. 3: 27 Apr 2021* *Protocol Amendment No. 2: 01 Feb 2021* *Protocol Amendment No. 1: 08 Apr 2020* *Original Protocol Final Version Date: 20 Dec 2019*
Study Treatment:	Crinecerfont (NBI-74788)
Study Number:	NBI-74788-CAH3003
Study Title:	A Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Efficacy of Crinecerfont (NBI-74788) in Adult 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

NCT04490915

Document History

Version Number	Date	Author(s)	Page Number	Modification(s)
1.0	16 December 2022	[redacted]	N/A	N/A
2.0	04 August 2023	[redacted]	26	The list of analysis sets was updated to include the FAS excluding site 1083, the DB Period SAS, and the OL Period SAS.
2.0	04 August 2023	[redacted]	29	Added On Drug-Missing Endpoint subject classification.
2.0	04 August 2023	[redacted]	29-31	Reference ranges were added for hormones (Section 6.5)
2.0	04 August 2023	[redacted]	33-34	Analysis visit windows were updated to include windows for subjects who discontinue study drug prior to Week 24 or Month 12.
2.0	04 August 2023	[redacted]	34	Safety Follow-Up summary was added to Table 5 and will be implemented for labs, ECGs, vital signs, and the BPRS where summary statistics are summarized by visit.
2.0	04 August 2023	[redacted]	36	The section for start and stop dates for concomitant medications was updated to include all concomitant medications.
2.0	04 August 2023	[redacted]	36	Updated MedDRA version to 26.0.
2.0	04 August 2023	[redacted]	39-40	Demographic and baseline characteristics were updated to include all pre-defined subgroups as well as the stratification factors per IRT.

2.0	04 August 2023	[redacted]	42	A section for Medical History was added to summarize this data.
2.0	04 August 2023	[redacted]	48	A sentence was added to the sensitivity analyses of the primary efficacy endpoint regarding the tipping point analysis along with the planned table and/or figure for these results.
2.0	04 August 2023	[redacted]	49-51	Additional supplementary analyses were added for the primary efficacy endpoint.
2.0	04 August 2023	[redacted]	51-55	Sensitivity analyses were added for each key secondary endpoint (tipping point and complete-case sensitivity analyses).
2.0	04 August 2023	[redacted]	62	Additional exploratory analyses were added for glucocorticoid dose regimen.
2.0	04 August 2023	[redacted]	66	A summary table was added to summarize total lean mass and total body mass in the OL period.
2.0	04 August 2023	[redacted]	67	Added tables to summarize change from the OL baseline for total TART volume as well as for total TART volume expressed as a percentage of total testicular volume.
2.0	04 August 2023	[redacted]	69	Added clarification that all safety outputs for endpoints during the double-blind period will use the DB Period Safety Analysis Set and outputs for safety during the open-label

				period will use the OL Period Safety Analysis Set.
2.0	04 August 2023	[redacted]	69	Added clarification to how duration of exposure will be calculated as it will be an overall measure of exposure over the DB and OL periods.
2.0	04 August 2023	[redacted]	73	The number of subjects meeting the criteria for orthostatic hypotension was added to the vital signs section.
2.0	04 August 2023	[redacted]	75	A section for Prior and Concomitant Medications was added to describe how this data will be summarized.
2.0	04 August 2023	[redacted]	78	For the MOS-12, removed the summarization of the 6 scale scores as they cannot be calculated based on the version of the MOS-12 used on the study.
2.0	04 August 2023	[redacted]	83-87	Appendices 16.5 and 16.6 were added to describe the tipping point analyses for the key secondary efficacy endpoints.
2.0	04 August 2023	[redacted]	Throughout document	Added SI units
3.0	10 August 2023	[redacted]	47-48	Clarified that the CHW method will only be used if there is a data-dependent sample size increase, otherwise the conventional test statistic will be used.

SIGNATURE PAGE

This document has been prepared by:

[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 FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Term
17-OHP	17-hydroxyprogesterone
A4	Androstenedione
ACTH	Adrenocorticotropic hormone
AE	Adverse event
ANCOVA	Analysis of covariance
BID	Twice daily
BLQ	Below the lower limit of quantification
BMD	Bone mineral density
BMI	Body mass index
BPRS	Brief psychiatric rating scale
BSA	Body surface area
CAH	Congenital adrenal hyperplasia
CI	Confidence interval
CMH	Cochran-Mantel-Haenszel
COVID-19	Coronavirus disease of 2019
CP	Conditional power
CSR	Clinical study report
C-SSRS	Columbia-Suicide Severity Rating Scale
CTx	Serum C-terminal telopeptide
DB	Double-blind
DMC	Data Monitoring Committee
DXA	Dual-energy X-ray absorptiometry
ECG	Electrocardiogram
eCRF	Electronic case report form
EMA	European Medicines Agency
EQ-5D-5L	EuroQol 5-Dimensions 5-Levels
EQ-VAS	EuroQol visual analog scale
FAS	Full analysis set
FCS	Fully conditional specifications
FDA	Food and Drug Administration
FSH	Follicle stimulating hormone

Abbreviation	Term
GC	Glucocorticoid
GFI	Global fatigue index
GTT	Glucose tolerance test
HbA1c	glycated hemoglobin
HDL	High-density lipoprotein
HOMA-IR	Homeostatic model assessment of insulin resistance
ICH	International Council for Harmonization
IPD	Important protocol deviation
IRT	Interactive response technology
ISC	Independent statistical center
IUD	Intrauterine device
LDL	Low-density lipoprotein
LH	Luteinizing hormone
LLN	Lower limit of normal
LS	Least squares
MAF	Multidimensional assessment of fatigue
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MOS-12	Medical outcomes study 12-item sleep scale
NBI	Neurocrine Biosciences, Inc.
NTx	urine N-terminal telopeptide
OL	Open-label
OLE	Open-label extension
PCS	Potentially clinically significant
PDHP	Protocol Deviation Handling Plan
PGWBI	Psychological general well-being index
PKAS	PK analysis set
PRA	Plasma renin activity
PT	Preferred Term
qAM	Every morning
QD	Once per day
QID	Four times per day

Abbreviation	Term
qPM	Every evening
SAE	Serious adverse event
SAP	Statistical analysis plan
SAS	Safety analysis set
SD	Standard deviation
SE	Standard error
SEM	Standard error of the mean
SF-36	36-Item Short Form Heath Survey
SHGB	Sex hormone binding globulin
SMQ	Standardized MedDRA Queries
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
δ_{int}	the interim observed treatment effect

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 24; "DB period") and open-label treatment period (Week 24 to Month 12; "OL period") of the Phase 3 Study NBI-74788-CAH3003. The planned analyses and data summaries for the open label or double-blind active-controlled treatment period (Month 12 to Month 18; "OL/DB period") data will be added to the SAP through an amendment. Summaries of data from the open-label extension for continued access to crinecerfont (starting after Month 18) 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 3 milestones throughout the study for the following analyses:

- interim analysis of the Week 24 data,
- final analysis of the Week 24 data and an analysis of all available safety and efficacy data up to Month 12,
- and final analysis of the Month 12 data.

Decisions regarding each of the 3 sets 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:

- To evaluate the efficacy of crinecerfont (100 mg bid), compared with placebo, in reducing daily glucocorticoid dosage while maintaining adrenal androgen control.
- To evaluate the efficacy of crinecerfont, compared with placebo, in reducing adrenal steroid levels following an initial 4-week treatment period.
- To evaluate the effect of crinecerfont, compared with placebo, on clinical endpoints associated with supraphysiologic glucocorticoid dosing.
- To evaluate plasma concentrations of crinecerfont and metabolites.
- To assess the safety and tolerability of crinecerfont.
- To evaluate an alternate dosing regimen of crinecerfont in subjects who have not reduced their glucocorticoid dose by Month 12.

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 the evening meal (doses separated by approximately 12 hours) for 24 weeks in approximately 165 adult subjects with classic congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency. Eligible subjects will be randomly assigned in a 2:1 ratio (active:placebo) to 2 treatment groups, crinecerfont 100 mg bid or placebo. After the 24-week randomized treatment period, there will be a 6-month, open-label treatment period, during which all subjects will receive crinecerfont 100 mg bid. At Month 12, subjects who have not reduced their glucocorticoid dose to ≤11 mg/m^2/day will be re-randomized (2:1) to receive 100 mg every morning (qAM) and 200 mg every evening (qPM) or to continue 100 mg bid, in a blinded fashion. Subjects who have reduced their glucocorticoid dose to ≤11 mg/m^2/day will continue to receive 100 mg bid in an open-label fashion. At Month 18, subjects will review applicable portions of the informed consent form and confirm whether they will participate in the optional open-label extension (OLE) treatment period for continued access to crinecerfont. Starting at Month 18, all subjects who are continuing in the OLE will initially receive crinecerfont 100 mg bid. If the subject has inadequate disease control despite receiving glucocorticoid treatment at their target dose (in the opinion of the investigator), the crinecerfont dose may be increased to 100 mg qAM and 200 mg qPM. After the Month 24 visit, an alternative dosing regimen of once daily 200 mg qPM can be considered per the investigator. Subjects will remain in the OLE until crinecerfont becomes commercially available, the Sponsor elects to discontinue development of crinecerfont for CAH, the Sponsor elects to discontinue the study, or the subject meets one of the study withdrawal criteria. A final study visit will be conducted approximately 4 weeks after the last dose of study drug.

3.1. Randomization

Approximately 165 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 total daily glucocorticoid regimen on Day 1 (<20 mg/m^2/day or ≥20 mg/m^2/day in hydrocortisone dose equivalents adjusted for BSA), glucocorticoid type (hydrocortisone alone; prednisone, prednisolone, methylprednisolone, with or without hydrocortisone; dexamethasone, with or without another glucocorticoid), and sex.

Subjects with glucocorticoid dose >11 mg/m^2/day (hydrocortisone dose equivalents) at Month 12 will be re-randomized 2:1 to a modified regimen of 100 mg qAM and 200 mg qPM or to continue crinecerfont at 100 mg bid. Month 12 randomization will be stratified by the original treatment assignment (crinecerfont versus placebo) and by Month 12 glucocorticoid dose (>14 mg/m^2/day versus ≤14 mg/m^2/day). Subjects who discontinue study drug prior to Month 12 are ineligible for re-randomization.

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 Month 18 data. The

Sponsor will remain blinded until all subjects complete the Week 24 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 Month 18 data.

An independent DMC will periodically review ongoing unblinded clinical and safety data to ensure the safety and well-being of the study subjects. An unblinded interim analysis of Week 24 data will be conducted by an independent statistical center (ISC) and will be evaluated by the DMC (as described in Section 5.1).

3.3. Sample Size Considerations

The protocol-specified sample size of 165 subjects (110 in the crinecerfont treatment group and 55 in the placebo group) is based on a power calculation for the primary endpoint and considerations for the size of the safety database. Based on a 2-sample t-test, an effect size of 0.75 with a sample size of at least 90 subjects (60 in the crinecerfont treatment group and 30 in the placebo treatment group) will have greater than 90% power to detect a treatment difference at a 0.05 level of significance. With the full sample size of 165 subjects, there is greater than 90% power to detect an effect size as small as 0.55. Sample size and power estimates were obtained from nQuery® Advisor Version 8.

A single planned interim analysis detailed in Section 5.1, with a futility analysis and an unblinded sample size re-estimation, will be performed when approximately the first 76 randomized subjects (approximately 46% of the planned number of enrolled subjects) have had the opportunity to be assessed for the primary endpoint. The unblinded sample size re-estimation may result in an increase of the overall sample size to a maximum of 210 subjects.

4. ENDPOINTS

The efficacy, safety, pharmacokinetic, and patient-reported outcome (PRO) endpoints are listed below.

4.1. Primary Efficacy – DB Period

The primary efficacy endpoint in the DB period is the percent change from baseline in glucocorticoid daily dose (in hydrocortisone equivalents adjusted for BSA [mg/m^2/day]) at Week 24, while Week 24 androstenedione is adequately controlled at ≤120% of baseline or ≤upper limit of normal (ULN) for age and sex.

4.2. Key Secondary Efficacy – DB Period

Key secondary efficacy endpoints in the DB period are:

- change from baseline in serum androstenedione at Week 4
- achievement of a reduction in glucocorticoid daily dose to physiologic levels at Week 24
- change from baseline in homeostatic model assessment of insulin resistance (HOMA-IR; in fasting subjects not on insulin) at Week 24
- percent change from baseline in weight at Week 24
- change from baseline in percent total fat mass at Week 24

4.3. Secondary Efficacy – DB Period

Secondary efficacy endpoints in the DB period are:

- change from baseline in 17-hydroxyprogesterone (17-OHP) at Week 4
- change from baseline in blood pressure at Week 24
- change from baseline in glucose tolerance (as measured by the post-GTT [glucose tolerance test] load glucose levels; in subjects without diabetes mellitus) at Week 24
- change from baseline in waist circumference at Week 24
- change from baseline in menstrual regularity (premenopausal female subjects not using hormonal or intrauterine device [IUD] contraception only) at Week 24
- change from baseline in testicular adrenal rest tumor (TART) volume (expressed as a percentage of the total testicular volume) at Week 24 (male subjects only)

4.4. Exploratory Efficacy – DB Period

Exploratory efficacy endpoints in the DB period include:

- change from baseline in all other hormone parameters (adrenocorticotropic hormone [ACTH], total testosterone, follicle stimulating hormone [FSH], luteinizing hormone [LH], progesterone, and sex hormone binding globulin [SHBG]) at Week 24
- change from baseline in urine androgen metabolite levels at Week 24

- change from baseline in metabolic laboratory parameters (total cholesterol, low-density lipoprotein [LDL], high-density lipoprotein [HDL], glycated hemoglobin [HbA1c], triglycerides) at Week 24
- change from baseline in bone mineral density (BMD) at Week 24
- change from baseline in total lean body mass at Week 24
- change from baseline in total body mass (by DXA) at Week 24
- change from baseline in hirsutism (female subjects only) at Week 24
- change from baseline in acne (female subjects only) at Week 24
- change from baseline in bone markers (serum osteocalcin, serum bone-specific alkaline phosphatase, serum C-terminal telopeptide [CTx], urine N-terminal telopeptide [NTx]) at Week 24
- change from baseline in body mass index (BMI) at Week 24
- achievement of a reduction in glucocorticoid dosing regimen instances per day at Week 24 (subjects on a hydrocortisone-alone regimen at baseline)
- achievement of switching to a hydrocortisone-alone glucocorticoid dose regimen at Week 24 (subjects on prednisone [or equivalent] or prednisone [or equivalent] plus hydrocortisone at baseline)
- achievement of switching to a hydrocortisone-alone glucocorticoid dose regimen at Week 24 (subjects on a dexamethasone-containing glucocorticoid regimen at baseline)
- achievement of switching to a dexamethasone-free glucocorticoid dose regimen at Week 24 (subjects on a dexamethasone-containing glucocorticoid regimen at baseline)

4.5. Exploratory Efficacy – OL Period

Exploratory efficacy endpoints in the OL period include:

- percent change from baseline and OL baseline in glucocorticoid daily dose (in hydrocortisone equivalents adjusted for BSA [$mg/m^2/day$]) at Month 12, while Month 12 androstenedione values are adequately controlled at ≤120% of baseline or ≤ ULN for age and sex
- achievement of a reduction in glucocorticoid daily dose to physiologic levels at Month 12
- change from baseline and OL baseline in hormone parameters (androstenedione, 17-OHP, ACTH, total testosterone, FSH, LH, progesterone, SHBG) at Month 12
- change from baseline and OL baseline in metabolic laboratory parameters (glucose tolerance [as measured by the post-GTT load glucose levels; in subjects without diabetes mellitus], HOMA-IR [in fasting subjects not on insulin], total cholesterol, LDL, HDL, HbA1c, triglyceride) at Month 12
- change and percent change from baseline and OL baseline in weight at Month 12
- change from baseline and OL baseline in BMI at Month 12

- change from baseline and OL baseline in waist circumference at Month 12
- change from baseline and OL baseline in percent total fat mass at Month 12
- change from baseline and OL baseline in total lean mass at Month 12
- change from baseline and OL baseline in total body mass (via DXA) at Month 12
- change from baseline and OL baseline in urine androgen metabolite levels at Month 12
- change from baseline and OL baseline in BMD at Month 12
- change from baseline and OL baseline in hirsutism (female subjects only) at Month 12
- change from baseline and OL baseline in acne (female subjects only) at Month 12
- change from baseline and OL baseline in bone markers (serum osteocalcin, serum bone-specific alkaline phosphatase, serum C-terminal telopeptide [CTx], urine N-terminal telopeptide [NTx]) at Month 12
- change from baseline and OL baseline in testicular adrenal rest tumor (TART) volume (expressed as a percentage of the total testicular volume) at Month 12 (male subjects only)
- change from baseline in menstrual regularity (premenopausal female subjects not using hormonal or IUD contraception only) at Month 12
- change in number of menstrual periods (premenopausal female subjects assigned to the placebo group for the DB period, not using hormonal or IUD contraception) at Week 24 to Month 12, compared to Day 1 to Week 24

4.6. Safety

Safety endpoints in the DB and OL periods include:

- Occurrence of treatment-emergent adverse events (TEAEs), treatment-emergent serious adverse events (TESAEs), adrenal insufficiency TEAEs, and TEAEs leading to discontinuation from study
- Observed and changes from baseline in clinical laboratory test values (including hematology, clinical chemistry, plasma renin activity [PRA], and coagulation)
- Observed and changes from baseline in vital sign values
- Observed and changes from baseline in 12-lead electrocardiogram (ECG) intervals
- Observed and changes from baseline in total scores from the Brief Psychiatric Rating Scale (BPRS)
- “Yes” responses to suicidal behavior or ideation at postbaseline assessments from the Columbia-Suicide Severity Rating Scale (C-SSRS)

4.7. Pharmacokinetics

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

4.8. Patient-Reported Outcomes

PRO endpoints in the DB and OL periods include:

- Observed and change from baseline values for the 8 health concepts and the 2 summary measures from the 36-Item Short Form Health Survey (SF-36)
- The 5 levels for each health dimension from the EQ-5D-5L over time
- Observed and changes from baseline in global fatigue index (GFI) values from the Multidimensional Assessment of Fatigue (MAF)
- Observed and changes from baseline in total score values from the Psychological General Well-Being Index (PGWBI)
- Observed and changes from baseline values in the quantity of sleep and the two problem indices from the Medical Outcomes Study 12-Item Sleep Scale (MOS-12)

5. DATA ANALYSES

5.1. Interim Analysis

An unblinded interim analysis on the primary endpoint using the "promising zone" method (Mehta and Pocock, 2011) will be conducted when approximately the first 76 subjects (approximately 46% of the planned number of enrolled subjects) have had the opportunity to complete the Week 24 assessments. The promising zone method as planned in this study was shown to not inflate the type I error, as described in the simulation report attached to the Adaptation Plan. This interim analysis, in conjunction with the final analysis method described in Section 9.3.2, will not require an adjustment to the level of significance for the final analysis of the Week 24 data. The interim analysis will be conducted by an ISC and will be evaluated by the DMC. The DMC will assess futility and consider sample size re-estimation based on the observed treatment effect and the conditional power at the time of the interim analysis. Conditional power (CP) is the probability that, conditional on the current value of the test statistic at the interim analysis, the study will achieve statistical significance at the final analysis.

At the time of the interim analysis, the CP will be calculated and the following actions will be recommended, based on the interim observed treatment effect (δ_{int}) which is the difference in mean percent change in the primary endpoint between the crinecerfont and placebo treatment groups, and the interim CP calculation:

$\delta_{int} < 0.05$	**Futility Zone.** Recommend stopping for futility
$CP < CP_{min}$	**Unfavorable Zone.** Continue to the planned sample size as the results are currently 'unfavorable'
$CP_{min} \leq CP < CP_{max}$	**Promising Zone.** The results are currently in the 'promising zone'. Increase the sample size, up to a maximum of 210 subjects, to achieve a conditional power of CP_{max}
$CP \geq CP_{max}$	**Favorable Zone.** Continue to the planned sample size as the results are currently 'favorable'

The adaptive elements of this interim analysis and the criterion for each DMC recommendation are detailed in an Adaptation Plan. The Adaptation Plan describes the values of CP_{min}, CP_{max}, and the values of CP that result in different sample size increases. In order to maintain study integrity, access to the Adaptation Plan will be limited to the ISC, DMC, and selected Sponsor staff assigned to the crinecerfont program, specifically the biostatisticians, regulatory lead, and program team lead.

Subjects assessed for the primary efficacy endpoint prior to the time of the interim analysis comprise the Stage 1 sample, and those subjects assessed after the interim analysis comprise the Stage 2 sample. Based on the Stage 1 and Stage 2 subject samples, the two independent stage-wise test statistics for superiority, Z_1 and Z_2, will be computed as follows:

Stage 1 test statistic: The Stage 1 test statistic is given by:

$Z_1 = \hat{\delta}_1 / \text{se}(\hat{\delta}_1)$

calculated from an ANCOVA model adjusting for the randomization stratification factors and relevant baseline value as a covariate, using subject data from Stage 1. Here, $\hat{\delta}_1$ denotes the estimate of δ based on Stage 1 data, and $\text{se}(\hat{\delta}_1)$ denotes its standard error.

Stage 2 test statistic: Similarly, the Stage 2 test statistic is defined by

$Z_2 = \hat{\delta}_2 / \text{se}(\hat{\delta}_2)$, calculated from an ANCOVA model adjusting for the randomization stratification factors and relevant baseline value as a covariate using subject data from Stage 2. Here, $\hat{\delta}_2$ denotes the estimate of δ based on Stage 2 data, and $\text{se}(\hat{\delta}_2)$ denotes its standard error.

With the conventions outlined above, CP at the interim analysis will be calculated as follows:

$$CP_{\tilde{\delta}} = P_{\tilde{\delta}}\left(Z_2 \geq \frac{c_2 - \sqrt{w}\,Z_1}{\sqrt{1-w}} \middle| Z_1\right) = 1 - \Phi\left(\frac{c_2}{\sqrt{1-w}} - \frac{\sqrt{1-w}}{\sqrt{w}}\,Z_1 - \frac{\sqrt{w}}{\sqrt{1-w}}Z_1\right)$$

where w is the information fraction at the time of the interim analysis based on the planned number of 165 subjects, and $c_2 = \Phi^{-1}(0.025) = 1.96$ is the efficacy boundary at final analysis with Φ being the cumulative distribution function of the standard normal distribution.

Based on the interim analysis results, one of the following three recommendations will be made by the DMC:

- Stop the study for futility
- Continue the study to the planned sample size
- Increase the sample size, up to a maximum of 210 subjects

The DMC will communicate the recommendation, without disclosing the numerical results, to a designated group of Sponsor personnel (Sponsor Review Committee [SRC]), who are not involved in the crinecerfont program. The SRC members will be firewalled from the crinecerfont study team and will keep the results confidential until the final analysis of the Week 24 data. The membership, responsibilities, and logistical considerations of the SRC are detailed in the SRC Charter.

Details about the final analysis of the primary endpoint, adjusting for the interim analysis, are outlined in Section 9.3.2.

5.2. Week 24 (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 24 visit. Data through Week 24, including the primary, key secondary, secondary, and exploratory endpoints, will be analyzed. At this time, an analysis of all available safety and efficacy data through Month 12 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 24 analysis are included in this SAP.

5.3. Month 12 (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 Month 12 visit. At this time, all data collected up to Month 12 will be analyzed. All planned analyses for the final Month 12 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 24 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 24; "DB period") will be summarized by randomized treatment group at Day 1, defined as placebo or crinecerfont 100 mg bid (DB treatment group). Data from the open-label period (Week 24 to Month 12; "OL period") will be summarized by OL treatment group, defined as placebo/crinecerfont 100 mg bid or crinecerfont 100 mg bid/crinecerfont 100 mg bid to describe their randomized treatment assignment at Day 1 and their OL treatment, as well as all subjects combined.

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.
Full Analysis Set Excluding Site No. 1083	The FAS as defined above, excluding subject data for site 1083.
DB Period Safety Analysis Set	The 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 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 summary 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.

Analyses for the primary efficacy and all key secondary efficacy endpoints will be re-run using the FAS excluding site number 1083 analysis set.

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.

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 24 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. Body composition assessments obtained within 2 weeks after Day 1 may serve as the baseline if needed. Bone mineral density and testicular ultrasound assessments obtained within 1 month after 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. Note that for hormone parameters that are collected both pre- and post-glucocorticoid (GC) dose, baselines will be defined as the last pre- or post-GC dose value available prior to study drug dosing for all pre- and post-GC dose postbaseline values, respectively, where a change from baseline calculation is necessary, unless otherwise specified.

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 24 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 Month 12 of the study. This baseline will be referred to as the "OL baseline" throughout this document. Unless otherwise noted, all efficacy summaries will be repeated using the study baseline (as defined in Section 6.3.1). For both the OL baseline and study baseline, if the respective Week 24 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 24 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 every 6 months, if the Week 24 assessment is not available then the OL baseline and postbaseline values will not be included in the summary.

Body composition assessments obtained within 2 weeks after Week 24 may serve as the OL baseline if needed. Bone mineral density and testicular ultrasound assessments obtained within 1 month after Week 24 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 hormone parameters that are collected both pre- and post-GC dose, OL baselines will be defined as the last pre- or post-GC dose value available prior to study drug dosing at Week 24 for all pre- and post-GC postbaseline values (after Week 24), respectively, where a change from OL baseline calculation is necessary, unless otherwise specified.

6.3.3. Baseline Definition of Androstenedione "Control" for Primary and Second Key Secondary Endpoints

At Day 1 and Week 24, a sample for serum androstenedione is collected before the morning glucocorticoid dose (pre-GC dose) and 2 hours after the morning glucocorticoid dose (post-GC dose). Subjects not on a morning glucocorticoid dose had two samples collected 2 hours apart with the second sample collected before 1100 hours on Day 1 and Week 24. The Week 24 samples were also timed to be collected prior to and 2 hours after the morning study drug dose.

The post-GC dose androstenedione values at Day 1 will be considered the baseline value for the purpose of assessing androstenedione control at Week 24 (also using the post-GC dose value) as part of the primary endpoint of GC dose reduction while maintaining androstenedione control. If the Day 1 value is missing, the last post-GC dose androstenedione value collected prior to Day 1 will serve as the baseline. If a subject is missing their Week 24 post-GC dose androstenedione value, then the pre-GC dose androstenedione values at Day 1 and Week 24 will be used instead. If both pairs of pre- and post-GC androstenedione values at Day 1 and Week 24 (ie, pre-GC dose Day 1/pre-GC dose Week 24 and post-GC dose Day 1/post-GC dose Week 24) are incomplete, then the androstenedione values (for the purpose of the Week 24 endpoint of GC dose reduction with androstenedione control) at Week 24 will be considered missing and will be imputed using the missing data handling approach described in the primary efficacy analysis of this endpoint (Section 9.3.1). If a subject is missing their Week 24 post-GC dose androstenedione value but has a Week 24 pre-GC dose androstenedione value that is ≤ ULN based on age and sex, then the subject will be considered in androstenedione "control." Note that if all Week 24 androstenedione values are missing, the subject will still have an androstenedione value assigned as baseline. This baseline value will be defined as the last post-GC value prior to study drug dosing. If there are no post-GC values available prior to study drug dosing, then the last pre-GC dose value prior to study drug dosing will be used instead. If a subject is missing all androstenedione values prior to study drug dosing, then their baseline androstenedione value will be multiply imputed using the methodology described in Section 16.1.

This same method of determining the baseline and corresponding Week 24 androstenedione values used to assess control will also be applied to the second key secondary endpoint, the achievement of a reduction in glucocorticoid daily dose to physiologic levels (≤11 mg/m^2/day hydrocortisone equivalents adjusted for BSA) while androstenedione is controlled at Week 24. The missing data handling approach is described in Section 9.4.2.

6.3.4. Baseline Definition of Androstenedione for First Key Secondary Endpoint

The first key secondary endpoint is the change from baseline to Week 4 in serum androstenedione based on the pre-GC dose androstenedione values at Day 1 and Week 4. If a subject is missing the Week 4 pre-GC dose androstenedione value, then the post-GC androstenedione values at Day 1 (or last post-GC dose androstenedione value prior to Day 1, if Day 1 is missing) and Week 4 will be used instead. If both pairs of pre- and post-GC androstenedione values at Day 1 and Week 4 (ie, pre-GC dose Day 1/pre-GC dose Week 4 and post-GC dose Day 1/post-GC dose Week 4) are incomplete, then the androstenedione values (for the purpose of the Week 4 androstenedione endpoint) at Week 4 will be considered missing and will be imputed using the missing data handling approach described in primary efficacy analysis method for this endpoint (Section 9.4.1). The last pre-GC dose androstenedione value prior to study drug dosing will be used as the baseline if all Week 4 values are missing. If there are no pre-GC values available prior to study drug dosing, then the last post-GC dose value prior to study drug dosing will be used instead. If a subject is missing all androstenedione values prior to study drug dosing, then their baseline androstenedione value will be multiply imputed using the methodology described in Section 16.2.

6.3.5. Baseline Definition of 17-OHP for Secondary Endpoint

Change from baseline to Week 4 in serum 17-OHP is a secondary endpoint. The same method as described in Section 6.3.4 will be followed for the purpose of determining baseline and Week 4 17-OHP values for this endpoint. Note that missing values will not be multiply-imputed.

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 24). 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 24, 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 but do not have an evaluable measurement for that endpoint.

6.5. Hormone Reference Ranges

The appropriate reference ranges for androstenedione, 17-hydroxyprogesterone, and progesterone in adult female subjects are dependent on the female subject's pre- or post-menopausal status at Day 1 and are shown in Table 3. For males, the reference ranges are dependent on age and the reference ranges are also given in Table 3. For the purposes of analysis, female subjects who are postmenopausal will have their postmenopausal status marked in their medical history. Female subjects who are postmenopausal will use the "postmenopausal phase" reference ranges provided by the central laboratory. All other female subjects will use the "pre-menopausal mid-follicular" reference ranges provided by the central laboratory.

Reference ranges are also given by sex for ACTH, follicle stimulating hormone (FSH), luteinizing hormone (LH), sex hormone binding globulin (SHBG), and total testosterone.

The assignment of these hormone reference ranges will apply to all summaries and analyses that use reference ranges.

Table 3: Hormone Reference Ranges

Hormone Reference Ranges	
Lab Analyte	Reference Ranges: conventional units; SI units
Androstenedione (conversion factor from ng/dL to nmol/L: x 0.0349)	Adult Male: 18-30 years:50-220 ng/dL; 1.745 -7.678 nmol/L 31-50 years: 40-190 ng/dL; 1.396 - 6.631 nmol/L 51-60 years: 50-220 ng/dL; 1.745 - 7.678 nmol/L Adult Female: Mid Follicular: 51-213 ng/dL; 1.7799 - 7.4337 nmol/L Surge: 73-230 ng/dL; 2.5477 - 8.027 nmol/L Mid Luteal: 73-184 ng/dL; 2.5477 - 6.4216 nmol/L Postmenopausal: 20-75 ng/dL; 0.698 - 2.6175 nmol/L
17-OH Progesterone (17-OHP) (conversion factor from ng/dL to nmol/L: x 0.0303)	Adult Male: 18-30 years: 32-307 ng/dL; 0.9696 - 9.3021 nmol/L 31-40 years: 42-196 ng/dL; 1.2726- 5.9388 nmol/L 41-50 years: 33-195 ng/dL; 0.9999 - 5.9085 nmol/L 51-60 years: 37-129 ng/dL; 1.1211 - 3.9087 nmol/L >= 61 years: Not established Adult Females: Pre-menopausal mid follicular: 23-102 ng/dL; 0.6969 - 3.0906 nmol/L Pre-menopausal surge: 67-349 ng/dL; 2.0301-10.5747 nmol/L Pre-menopausal mid luteal: 139-431 ng/dL; 4.2117 - 13.0593 nmol/L Post-menopausal phase: <=45 ng/dL; 1.3635 nmol/L

Progesterone (conversion factor from ng/mL to nmol/L: x 3.18)	Male: <0.15ng/mL; 0.477 nmol/L, <0.48 nmol/L; 1.5264 nmol/L Female: Premenopausal, not pregnant: <23.90 ng/mL, <76 nmol/L Female, pregnant: 11.00 – 214 ng/mL, 34.98 - 680.52 nmol/L Female, postmenopausal: <0.13 ng/mL, <0.41 nmol/L
ACTH (conversion factor from pg/mL to pmol/L: x 0.22)	Males and Females: 7.2 – 63.3 pg/mL; 1.584 -13.926 pmol/L
Follicle stimulating hormone (FSH) (conversion factor from mIU/mL to IU/L: x 1.0)	Male: 1.6 - 11.0 mIU/mL (IU/L) Female: 1.8 - 20.4 mIU/mL (IU/L)
Luteinizing Hormone (LH) (conversion factor from mIU/mL to IU/L: x 1.0)	Male: 1.7 - 8.6 mIU/mL (IU/L) Female: 1.0 - 95.6 mIU/mL (IU/L)
Sex Hormone Binding Globulin (SHBG)	Male: 10.0 - 80.0 nmol/L (same in SI) Female: 20.0 - 130.0 nmol/L (same in SI)
Total Testosterone (Medpace/MRL) (conversion factor from ng/mL to nmol/L: x 3.47)	Male: 2.70- 10.70 ng/mL; 9.37-37.13 nmol/L Female: 0.06 - 0.86 ng/mL; 0.21-2.98 nmol/L
Total Testosterone (Quest) (conversion factor from ng/dL to nmol/L: x 0.0347)	Male: 250 - 1100 ng/dL; 8.675 - 38.17 nmol/L Female: 2 - 45 ng/dL; .0694 - 1.5615 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 0.

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. For values that are above the upper limit of quantification (ULQ), the value will be set equal to the ULQ.

6.7. Study Day

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. 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. For subjects who are randomized but never dosed, Day 1 will be defined as the subject's randomization date.

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, Table 5, Table 6, Table 7, Table 8. 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. For the purpose of this study, months are defined as 4-week intervals.

The primary endpoint of percent change from baseline in glucocorticoid daily dose at Week 24 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 24 visit are a critical component to the assessment of this endpoint as subjects who do not maintain androstenedione control 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 24, the Week 24 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 primary endpoint. If a subject has a Week 24 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 24 glucocorticoid total daily dose. If both hormone and safety labs were missed at the visit, then the vital signs assessment and its respective Week 24 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 (eg, every 6 months). These more-infrequent assessments include measurements obtained via DXA (bone mineral density and body composition), testicular ultrasound measurements, HbA1c, glucose tolerance test, insulin, bone markers, and lipid panel. These assessments will use the analysis visit windows defined in Table 6, Table 7, and Table 8, as applicable. Table 5 includes the analysis visit windows for subjects who discontinue from study drug prior to OL period or OL/DB period and therefore do not have a first dose date in the OL period or OL/DB period.

Table 4: Primary Analysis Visit Windows

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
Week 2	14	2 to 21
Week 4	28	22 to 36
Week 6	42	37 to 53
Week 9	63	54 to 74
Week 12	84	75 to 98
Week 16	112	99 to 126
Week 20	140	127 to 154
Week 24	168	155 to (first dose date in OL period)
Month 7	196	(First dose date in OL +1) to 210
Month 8	224	211 to 238
Month 9	252	239 to 266
Month 10	280	267 to 308
Month 12	336	309 to (first dose date in OL/DB period)

Table 5: Primary Analysis Visit Windows for Subjects Who Discontinue Study Drug Prior to Week 24 or prior to Month 12

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
Week 2	14	2 to 21
Week 4	28	22 to 36
Week 6	42	37 to 53
Week 9	63	54 to 74
Week 12	84	75 to 98
Week 16	112	99 to 126
Week 20	140	127 to 154
Week 24	168	155 to 182

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
Month 7	196	183 to 210
Month 8	224	211 to 238
Month 9	252	239 to 266
Month 10	280	267 to 308
Month 12	336	309 to 350
DB Safety Follow-up[a*]	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 day of last dose of study drug in DB period + 42 days
OL Safety Follow-up[b*]	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 day of last dose of study drug in OL period + 42 days

[*]Note that only safety labs, vital signs, ECG, and BPRS 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 precedent 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 mineral density and testicular ultrasound 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 Mineral Density and Testicular Ultrasound

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
Week 24	168	84 to (first dose date in OL period + 28)
Month 12	336	252 to (first dose date in OL/DB period + 28)

Analysis visit windows for body composition assessments are presented in Table 7.

Table 7: Analysis Visit Windows for Body Composition

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
Week 24	168	84 to (first dose date in OL period + 14)
Month 12	336	252 to (first dose date in OL/DB period + 14)

Analysis visit windows for the glucose tolerance test, HbA1c, bone markers, and lipid panel assessments are presented in Table 8.

Table 8: Analysis Visit Windows for Glucose Tolerance Test, HbA1c, Bone Markers, and Lipid Panel

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
Week 24	168	84 to (first dose date in OL period)
Month 12	336	252 to (first dose date in OL/DB period)

6.8. Handling of Missing Data

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 mediations, 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 2021 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 and were screened (provided in table footnote)
- Number randomized (within column header)
- Number received study drug
- Number completed planned study drug through Week 24
- Number completed study through Week 24
- Number who did not complete study drug through Week 24, including reasons for early discontinuation of study drug
- Number who did not complete the study through Week 24, 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 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 Month 12
- Number completed study through Month 12
- Number who did not complete study drug through Month 12, including reasons for early discontinuation of study drug
- Number who did not complete the study through Month 12, including reasons for early termination from the study

A summary of the number and percentage of subjects 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 at Day 1 will also be provided and will include subject ID, informed consent date, randomization date, stratification factors, and randomized treatment group.

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.

A listing of lot numbers used in the study will be provided.

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 related to the COVID-19 pandemic or major PDs classified as IPDs 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 (18-25 years; 26-35 years; 36-45 years; 46-55 years; 56-65 years; 66-75 years; ≥76 years)
- sex (based on the eCRF)
- ethnicity
- race
- region (OUS vs US)
-

Baseline characteristics (based on eCRF unless otherwise specified):

- height (cm)
- weight (kg)
- body mass index (BMI, kg/m^2)
- body mass index (<30 kg/m^2 vs ≥30 kg/m^2)

- BSA (m^2; DuBois formula where BSA [m^2] = Weight [kg]$^{0.425}$ × height (cm)$^{0.725}$ × 0.007184])
- serum androstenedione
- HOMA-IR
- total fat mass (kg)
- percent total fat mass (%)
- total daily glucocorticoid dose adjusted for BSA (mg/m^2/day; in hydrocortisone equivalents [see Appendix 16.3])
- total daily glucocorticoid dose (mg/day; in hydrocortisone equivalents [see Appendix 16.3])
- total daily glucocorticoid dose categories adjusted for BSA (<20 mg/m^2/day vs. ≥20 mg/m^2/day; in hydrocortisone equivalents [see Appendix 16.3])
- glucocorticoid type (hydrocortisone alone; prednisone, prednisolone, methylprednisolone, with or without hydrocortisone; dexamethasone, with or without another glucocorticoid)
- glucocorticoid type – dexamethasone use (subjects on dexamethasone at baseline versus subjects not on dexamethasone at baseline)
- glucocorticoid type – hydrocortisone vs. synthetic glucocorticoids (subjects on hydrocortisone alone versus subjects on prednisone [or equivalent] or dexamethasone)
- glucocorticoid type (hydrocortisone; prednisone, prednisolone, methylprednisolone; dexamethasone; 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)

Stratification factors as entered in the IRT:

- total daily glucocorticoid dose adjusted for BSA category (<20 mg/m^2/day vs ≥20 mg/m^2/day; in hydrocortisone equivalents)
- glucocorticoid type (hydrocortisone alone; prednisone, prednisolone, methylprednisolone, with or without hydrocortisone; dexamethasone, with or without another glucocorticoid)
- 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"/ "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:

- Impaired fasting glucose
- Impaired glucose tolerance

- Diabetes
- Hypertension
- Hypertriglyceridemia
- Hyperlipidemia
- Osteopenia
- Osteoporosis
- Fragility fracture (not due to trauma)
- Depression
- Anxiety
- Hirsutism (female subjects only)
- Acne (female subjects only)
- Testicular adrenal rest tumors (male 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
- 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
- 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 (female subjects only)
- Menstrual cycles occurring every 21-35 days (Y/N; female subjects of childbearing potential only not on hormonal or IUD contraception)

7.5. Medical History

Medical history data will be summarized descriptively for the FAS, 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.

7.6. Summary of Analysis Sets

A summary of the number and percentage of subjects included in the full analysis set, the full analysis set excluding site number 1083, 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.

8. PHARMACOKINETICS

The plasma concentrations of crinecerfont and its metabolites will be summarized with descriptive statistics by 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 below the lower limit of quantification (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.

The details of the population PK analysis will be described separately and the results provided in a separate report.

9. EFFICACY – DB PERIOD

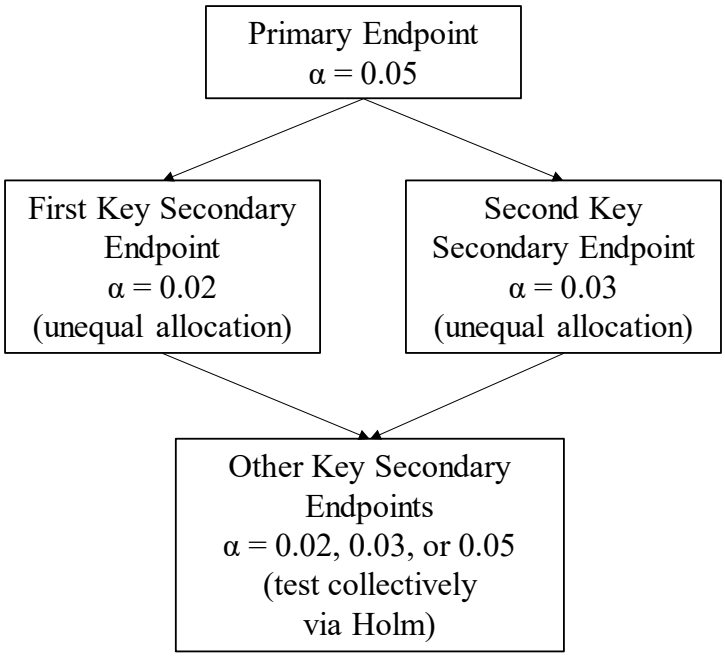
The efficacy endpoints and planned analysis methods for the DB period are described below. Unless otherwise specified, the FAS will be used for all efficacy analyses and descriptive statistics.

9.1. Multiple Comparisons and Multiplicity

The testing procedure described below will be used to control the familywise Type I error rate for the primary and key secondary endpoints (Figure 1). The promising zone method as planned in this study was shown to not inflate the type I error, as described in the simulation report attached to the Adaptation Plan. All significance levels noted below are 2-sided. All other p-values will not be adjusted for multiplicity and should be considered as nominal p-values.

- The primary endpoint will be tested using a significance level of 0.05.
- If the primary endpoint is statistically significant, then the first 2 key secondary endpoints will be tested using an unequal allocation of significance levels. The change from baseline at Week 4 in androstenedione will be tested using a significance level of 0.02. The achievement of a reduction in glucocorticoid daily dose to physiologic levels ($\leq$11 mg/m^2/day hydrocortisone equivalent adjusted for BSA) will be tested using a significance level of 0.03.
- The family of remaining key secondary endpoints will be tested using the Holm procedure (Holm, 1979). The significance level for this family of endpoints will be dependent on whether the first 2 key secondary endpoints are statistically significant (Wiens, 2003). If both are statistically significant, then the significance level for this family of endpoints will be 0.05. If only one is statistically significant, then the significance level for that endpoint will be used for this family of endpoints. If neither is significant, then this family of endpoints will not be tested for statistical significance. However, nominal p-values will be produced.

Figure 1: Procedure to Control for Multiple Comparisons



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 (or percent 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 100 mg bid; 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 GC_TYPE GC_DOSE TRTP;

MODEL CHG = TRTP BASE SEX GC_TYPE GC_DOSE;

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 100 mg bid 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*TRTP*GC_TYPE*GC_DOSE*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: Subjects in the FAS with classic CAH due to 21-hydroxylase deficiency on a stable, supraphysiologic glucocorticoid dose regimen (> 13 mg/m^2/day) at study entry who do not have evidence of glucocorticoid overtreatment. Subjects must have met the other inclusion and exclusion criteria defined in the protocol.

Variable (or endpoint) to be obtained for each subject: percent change from baseline in glucocorticoid daily dose (in hydrocortisone equivalents adjusted for body surface area [BSA; mg/m^2/day]) at Week 24.

The following intercurrent events will be handled using a composite strategy:

1) If a subject experiences a *decrease* in glucocorticoid daily dose but does not maintain androstenedione control (defined as ≤120% of baseline or ≤ULN for age and sex) at Week 24, the subject will be considered a treatment failure and their glucocorticoid daily dose endpoint will be set to zero percent change from baseline.
2) If a subject discontinues study drug but stays on study through Week 24, the subject's observed glucocorticoid daily dose at Week 24 will be used in the analysis.
3) If a subject is on study drug through Week 24 but is missing Week 24 glucocorticoid dose, androstenedione values, or both, the subject's endpoint will be multiply imputed from subjects within the same treatment group with non-missing data at Week 24 (Section 16.1).
4) If a subject in the crinecerfont treatment group discontinues from study drug and is missing glucocorticoid dose or androstenedione values at Week 24, the subject's endpoint will be imputed using data from retrieved dropouts (Section 16.1). 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 16.1). For subjects in the placebo treatment group, missing data will be imputed using non-missing data from subjects in the placebo treatment group (Section 16.1) 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.

Population-level summary: The least-squares mean treatment difference for this endpoint will be estimated using an ANCOVA model. Missing data will be imputed using the multiple imputation procedures defined in Section 16.1.

9.3.2. Primary Efficacy Analysis

The primary efficacy endpoint is the percent change from baseline in glucocorticoid daily dose (in hydrocortisone equivalents adjusted for BSA [mg/m^2/day]) at Week 24, while Week 24 androstenedione is adequately controlled at ≤120% of baseline or ≤ULN for age and sex. Glucocorticoid daily dose will be calculated at baseline (as defined in Section 6.3.1) and at Week 24 using the doses reported on the glucocorticoid medications eCRF (see details in Section 6.7 regarding mapping of the glucocorticoid medication to Week 24). Glucocorticoids other than hydrocortisone will be converted to hydrocortisone equivalents using the equivalency factors defined in Appendix A of the study protocol and Section 16.3. BSA will be calculated using the height from the screening visit and weight from the Day 1 visit, computed using the DuBois formula (Section 7.3). 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.

Baseline androstenedione will be defined as the Day 1 post-morning glucocorticoid dose androstenedione value. Further details of the derivation of this baseline value are described in Section 6.3.3.

Subjects who have a decrease in glucocorticoid daily dose but do not meet the definition of androstenedione control at Week 24 will be considered to have a zero percent change from baseline in the glucocorticoid daily dose at Week 24. Any increases from baseline in glucocorticoid daily dose will be reported as such, regardless of androstenedione control. Subjects who are missing Week 24 glucocorticoid dose data or all Week 24 androstenedione levels will have their data imputed through the multiple imputation procedures described in Section 16.1.

The primary analysis of the primary endpoint will be performed using the CHW test statistic (Cui, Hung, Wang, 1999) comprising test statistics from stage 1 of the trial (the interim dataset), and stage 2 of the trial (the remaining part of the trial). Each test statistic will be generated using an analysis of covariance (ANCOVA) model, as described in Section 9.2.1. If the total sample size was increased in a data-dependent manner as a result of the interim analysis, the primary analysis of the primary endpoint will be performed using the CHW test statistic, as proposed by Cui, Hung, and Wang (1999) and independently by Lehmacher and Wassmer (1999), to determine statistical significance. The CHW test statistic is a weighted test statistic comprised of test statistics from stage 1 of the trial (the interim dataset) and stage 2 of the trial (the remaining part of the trial). Each test statistic will be generated using an ANCOVA model, as described in Section 9.2.1. The independent increments of the Z statistics (i.e., the Wald statistics constructed using model-based treatment effect estimates) of the 2 stages are combined by pre-specified weights that are based on the planned proportion of total number of subjects at which the interim analysis would be taken if there were no change in the design.

In the present case, the study is designed for 165 subjects with an interim analysis after approximately 76 subjects are evaluated for the primary efficacy endpoint. Therefore, the planned proportions are 0.46 and 0.54 for the two stages, respectively. The independent incremental Z statistics for the 2 stages are combined with weights that equal the square root of 0.46 for Stage 1 and square root of 0.54 for Stage 2. Hypothesis test of the null hypothesis H_O: $\delta \leq 0$, tested at one-sided 2.5% level at final analysis, is based on the CHW test statistic, Z_{CHW}, which combines the independent stage-wise test statistics defined in Section 5.1, as follows:

$$Z_{\mathrm{CHW}} = \sqrt{w}\, Z_1 + \sqrt{1-w}\, Z_2.$$

The p-value is defined by $P_{CHW}=\Phi(Z_{CHW})$, where Φ is the cumulative probability based on standard normal distribution. Gao et al (2008) have shown that the type-1 error is preserved at the specified α level despite the possibility of a data-dependent increase in the sample size.

If the total sample size is not increased in a data-dependent manner as a result of the interim analysis, the primary analysis of the primary endpoint will be performed using the conventional test statistic generated from an ANCOVA model (as described in Section 9.2.1) to determine if statistical significance is reached. As a sensitivity analysis, the CHW test statistic will be computed as described above.

All sensitivity and supplementary analyses outlined in Sections 9.3.3 and 9.3.4 will use the same method (ie, CHW test statistic or conventional test statistic) as the primary analysis of the primary endpoint.

Statistical significance for the treatment comparison will be determined using the multiple comparison procedure described in Section 9.1. Descriptive statistics and LS means for the percent change from baseline in total glucocorticoid daily dose (with androstenedione control) at Week 24 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 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).

The analyses described in this section for the primary efficacy endpoint will be repeated for the full analysis set excluding site number 1083.

9.3.3. Sensitivity Analyses of the Primary Efficacy Results

The assumptions relative to the ANCOVA model (e.g. 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 24 glucocorticoid daily dose values (from the MI procedure described in Section 16.1) for subjects in either the crinecerfont treatment group or the placebo treatment group will be delta-adjusted (with a range of plausible penalties for each treatment group) until the treatment difference at Week 24 is no longer statistically significant. The implementation procedure for the tipping point analysis is described in Appendix 16.4. 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 primary endpoint will be repeated using only subjects with observed data at Week 24. Subjects who are missing the Week 24 primary endpoint will be excluded from the analysis.

Descriptive statistics and LS means for the 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:

- As an alternative method to handling missing data at Week 24, the ANCOVA analysis for the primary endpoint will be repeated using the Week 20 daily glucocorticoid dose (based on the Week 20 analysis visit date for androstenedione) for subjects who are missing Week 24 glucocorticoid dose data. This analysis will follow the same conventions as are used for the primary analysis of the primary endpoint except in the case for subjects who have missing glucocorticoid dose at Week 24, their Week 20 glucocorticoid dose will be used instead. The same definition of androstenedione control will be implemented using androstenedione values at baseline and Week 20 for this subset of subjects. For subjects missing both Week 20 and Week 24 glucocorticoid doses, the missing data imputation strategy from the primary analysis of the primary endpoint will be used.
- The ANCOVA analysis for the primary endpoint will be repeated using the average of the Week 20 and Week 24 daily glucocorticoid doses. Post-glucocorticoid dose androstenedione values at Week 20 and Week 24 will be averaged and compared against the baseline post-glucocorticoid dose androstenedione value to assess whether the subject maintained androstenedione "control" (where "control" is defined as the average of the Week 20 and 24 androstenedione values ≤120% of baseline or ≤ULN for age and sex). If a subject is missing Week 20 or Week 24, then the visit with the observed values will be used instead of the average. Subjects who are missing Week 20 and Week 24 glucocorticoid dose data or all Week 20 and Week 24 androstenedione values will have their data imputed through the multiple imputation procedure following the same methods as described in the primary analysis of the primary endpoint.

- The ANCOVA analysis as described for the primary 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 24. All conventions used in the primary analysis of the primary endpoint will be used in this supplementary analysis.
- The ANCOVA analysis (as described above) for the primary endpoint will be repeated using the absolute change from baseline in glucocorticoid daily doses in mg/day (ie, not adjusted for BSA). All conventions used in the primary analysis of the primary endpoint will be used in this supplementary analysis.
- The ANCOVA analysis for the primary endpoint will be repeated using the Week 24 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 24 will be used regardless of the subject's androstenedione values. Missing data at Week 24 will be imputed using the same methods as in the primary analysis of the primary efficacy endpoint.
- The ANCOVA analysis for the primary analysis of the primary 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. Missing data at Week 24 will be imputed using the same methods as in the primary analysis of this endpoint.
- The ANCOVA analysis for the primary analysis of the primary endpoint will be repeated with androstenedione control being defined as ≤180% of the baseline value or ≤ ULN. Missing data at Week 24 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 glucocorticoid total daily dose at Week 12, Week 16 or Week 20 where androstenedione is controlled if control is not met at Week 24. Note that missing data at Weeks 12, 16 and 20 will not be imputed. For subjects who do not meet the androstenedione control definition at any visit between Week 12 and Week 24 or are missing all data between Week 12 and Week 20, their percent change from baseline in glucocorticoid total daily dose will be set to 0. Missing data at Week 24 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 highest glucocorticoid total daily dose at Week 12, Week 16 or Week 20 where androstenedione is controlled if control is not met at Week 24. Note that missing data at Weeks 12, 16 and 20 will not be imputed. For subjects who do not meet the androstenedione control definition at any visit between Week 12 and Week 24 or are missing all data between Week 12 and Week 20, their percent change from baseline in glucocorticoid total daily dose will be set to 0. Missing data at Week 24 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 average of the two screening post-GC dose androstenedione visit values and the Day 1 post-GC dose androstenedione value at baseline as compared to the average of the

Week 20 and Week 24 post-GC dose androstenedione values. If the Week 20 and Week 24 total daily glucocorticoid doses are not equal, then the Week 24 post-GC dose androstenedione value will be used instead. If there is missing data (GC or androstenedione) at Week 20 or Week 24 then the observed data from the visit with non-missing data will be used instead. If both Week 20 and Week 24 have missing GC or androstenedione data then the Week 24 imputed value will be used. Missing data at Week 24 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 weight collected at Week 24 in the calculation of BSA (instead of using Day 1). That is, the baseline glucocorticoid total daily dose in mg/m^2/day will be calculated using height at the screening visit and baseline weight and the Week 24 glucocorticoid total daily dose in mg/m^2/day will be calculated using height at the screening visit and weight at the Week 24 visit in the calculation of Week 24 BSA. If Week 24 weight is missing, baseline BSA will be used instead. Missing data at Week 24 will be imputed using the same methods as 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 Androstenedione

The first key secondary endpoint is the change from baseline in serum androstenedione at Week 4. The pre-morning glucocorticoid dose androstenedione values at baseline and Week 4 will be used in the calculation of this endpoint (as defined in Section 6.3.4). If a subject is missing their baseline or Week 4 pre-glucocorticoid dose androstenedione value, then the post-glucocorticoid androstenedione values at baseline and Week 4 will be used instead. The primary analysis of this endpoint will be performed using an ANCOVA model as described in Section 9.2.1. Subjects who are missing all androstenedione levels at Week 4 will have their Week 4 data imputed through the multiple imputation procedure described in Section 16.2.

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 by treatment group and visit for the serum androstenedione observed and change from baseline values through Week 24. For visits where androstenedione is collected more than once (e.g., Day 1), each visit will be presented with the pre- and post- morning GC dose androstenedione values. The androstenedione observed values from the two screening visits and the average of these values will also be summarized.

Mean (±SEM) values of serum androstenedione at each visit will be summarized in line graphs by treatment group. For visits where androstenedione is collected more than once, each visit will be presented with the pre- and post- morning GC dose androstenedione values on separate plots. Similar graphs will be presented for the changes from baseline.

A tipping point sensitivity analysis of change from baseline in serum androstenedione at Week 4 will be performed to assess the robustness of the missingness assumptions as was described for the primary endpoint in Section 9.3.3. The implementation procedure for the tipping point analysis is described in Appendix 16.5 for each key secondary efficacy endpoint.

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

As a supplementary analysis, the percent change from baseline in serum androstenedione 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 to Week 4 in serum androstenedione will be analyzed using the average of the two screening post-GC dose visit values and the Day 1 post-GC dose value as compared to the Week 4 post-GC dose androstenedione value. If one or more of the screening or Day 1 values is missing or if collected pre-GC dose, then the average of the remaining values will be used instead. This analysis will be performed using an ANCOVA model as described above. 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 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 methods as described in the primary analysis of this endpoint.

9.4.2. Physiologic Glucocorticoid Daily Dose Responder Analysis

The second key secondary endpoint is the achievement of a reduction in glucocorticoid daily dose to physiologic levels (≤11 mg/m^2/day hydrocortisone equivalent adjusted for BSA) at Week 24, while Week 24 androstenedione is adequately controlled at ≤120% of baseline or ≤ULN for age and sex. The selection of the baseline and Week 24 androstenedione values used to assess control for this endpoint is the same method as is used for the primary endpoint (and described in Section 6.3.3). 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. If the model fails to converge due to a small number of observations within a cell, stratification levels may be collapsed to allow for model convergence. Subjects whose androstenedione levels at Week 24 are not controlled will be considered non-responders. For subjects who are missing glucocorticoid dose, androstenedione, or both at Week 24, the multiply-imputed datasets from the primary 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. For the Week 24 visit, the treatment difference and p-value from the CMH analysis will also be displayed.

A tipping point sensitivity analysis of the achievement of reduction in physiologic glucocorticoid daily dose at Week 24 will be performed to assess the robustness of the missingness assumptions as was described for the primary endpoint in Section 9.3.3. The implementation procedure for the tipping point analysis is described in Appendix 16.6 for this key secondary efficacy endpoint.

A complete-case sensitivity analysis will also be performed to test the robustness of the multiple imputation procedure. The CMH analysis for the physiologic glucocorticoid daily dose endpoint will be repeated using only subjects with observed data at Week 24. Subjects who are missing the Week 24 secondary endpoint will be excluded from the analysis.

9.4.3. HOMA-IR

An additional key secondary endpoint is the change from baseline in HOMA-IR at Week 24 (in subjects without diabetes mellitus). HOMA-IR is calculated as: (fasting glucose in mmol/L x fasting insulin in mU/L) / 22.5. Subjects who are concomitantly taking insulin at Day 1 or subjects who are not fasting at the time of the assessment will be excluded from this analysis. The analysis of this endpoint will be performed using an ANCOVA model, as described in Section 9.2.1. Subjects who are missing HOMA-IR at Week 24 will have their data imputed through the multiple imputation procedure described in Section 16.2.

Statistical significance for the treatment comparison will be determined using the multiple comparison procedure described in Section 9.1. Descriptive statistics and LS means for change from baseline to Week 24 in HOMA-IR will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

A tipping point sensitivity analysis of change from baseline in HOMA-IR at Week 24 (in subjects without diabetes mellitus) will be performed to assess the robustness of the missingness assumptions as was described for the primary endpoint in Section 9.3.3. The implementation procedure for the tipping point analysis is described in Appendix 16.5 for each key secondary efficacy endpoint.

A complete-case sensitivity analysis will also be performed to test the robustness of the multiple imputation procedure. The ANCOVA analysis for the change from baseline in HOMA-IR at week 24 will be repeated using only subjects with observed data at Week 24. Subjects who are missing the Week 24 secondary endpoint will be excluded from the analysis.

9.4.4. Weight

An additional key secondary endpoint is the percent change from baseline in weight at Week 24 and will be analyzed using an ANCOVA model, as described in Section 9.2.1. Subjects who are missing weight at Week 24 will have their data imputed through the multiple imputation procedure described in Section 16.2.

Statistical significance for the treatment comparison will be determined using the multiple comparison procedure described in Section 9.1. Descriptive statistics and LS means for percent change from baseline to Week 24 in weight will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

A tipping point sensitivity analysis of percent change from baseline in weight at Week 24 will be performed to assess the robustness of the missingness assumptions as was described for the

primary endpoint in Section 9.3.3. The implementation procedure for the tipping point analysis is described in Appendix 16.5 for each key secondary efficacy endpoint.

A complete-case sensitivity analysis will also be performed to test the robustness of the multiple imputation procedure. The ANCOVA analysis for percent change from baseline in weight will be repeated using only subjects with observed data at Week 24. Subjects who are missing the Week 24 secondary endpoint will be excluded from the analysis.

As a supplementary analysis, the absolute change from baseline in weight at Week 24 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.

Descriptive statistics will be presented by treatment group and visit for weight observed, change from baseline, and percent 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 and percent change from baseline values.

In the subset of subjects classified as overweight or obese (BMI $\geq$ 25 kg/m^2 at baseline), the number and percentage of subjects achieving a greater than 5 percent reduction in total body weight from baseline to Week 24 will also be presented by treatment group.

In addition to weight, BMI observed and change from baseline values will be summarized with descriptive statistics by visit and treatment group. The change from baseline to Week 24 in BMI will be analyzed using an ANCOVA model, as described in Section 9.2.1. Missing data will be imputed using the same methods as described in the primary analysis of the change from baseline in weight (as described above). Descriptive statistics and LS means for the change from baseline to Week 24 in BMI will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

BMI will also be presented in the following categories: underweight (BMI < 18.5 kg/m^2), normal ($\geq$ 18.5 to < 25 kg/m^2), overweight ($\geq$ 25 to < 30 kg/m^2), obese I ($\geq$ 30 to < 35 kg/m^2), obese II ($\geq$ 35 to < 40 kg/m^2), obese III ($\geq$ 40 kg/m^2). Descriptive statistics will be presented by treatment group for the number and percentage of subjects in each BMI category by visit.

Shifts from baseline to Week 24 in collapsed BMI categories will be presented in tables. The categories are defined as follows: underweight or normal (BMI < 25 kg/m^2), overweight ($\geq$ 25 to < 30 kg/m^2), and obese (BMI $\geq$ 30 kg/m^2). Each shift table will have four rows and four columns, with rows reflecting the categories at baseline as well as missing values and columns reflecting the categories at Week 24 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.

9.4.5. Total Fat Mass

An additional key secondary endpoint is the change from baseline in percent total fat mass (calculated as total fat mass [g]/total body mass [g] * 100) at Week 24 and will be analyzed using an ANCOVA model, as described in Section 9.2.1. Subjects who are missing percent total fat mass at Week 24 will have their data imputed through the multiple imputation procedure described in Section 16.2.

Statistical significance for the treatment comparison will be determined using the multiple comparison procedure described in Section 9.1. Descriptive statistics and LS means for change from baseline to Week 24 in percent total fat mass 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 percent total fat mass observed and change from baseline values.

A tipping point sensitivity analysis of change from baseline in percent total fat mass at Week 24 will be performed to assess the robustness of the missingness assumptions as was described for the primary endpoint in Section 9.3.3. The implementation procedure for the tipping point analysis is described in Appendix 16.5 for each key secondary efficacy endpoint.

A complete-case sensitivity analysis will also be performed to test the robustness of the multiple imputation procedure. The ANCOVA analysis for change from baseline in percent total fat mass will be repeated using only subjects with observed data at Week 24. Subjects who are missing Week 24 percent total fat mass data will be excluded from the analysis.

As a supplementary analysis, the change from baseline in total fat mass (in kg) at Week 24 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.

9.5. Analysis of Secondary Efficacy Endpoints

The following sections describe the secondary efficacy endpoints and methods of summarization and analysis. For all analyses of the secondary 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.

9.5.1. 17-OHP

Change from baseline to Week 4 in 17-OHP is a secondary endpoint. The pre-morning glucocorticoid dose 17-OHP values at baseline and Week 4 will be used in the calculation of this endpoint (as described in Section 6.3.5). If a subject is missing their Week 4 pre-glucocorticoid dose 17-OHP value, then the post-glucocorticoid 17-OHP values at baseline and Week 4 will be used instead.

The analysis of this endpoint will be performed using an analysis of covariance (ANCOVA) model, as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 4 in 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 by treatment group and visit for 17-OHP observed and change from baseline values. Mean (±SEM) values of 17-OHP at each visit will be summarized in line graphs by treatment group. For visits where 17-OHP is collected more than once (e.g., Day 1), each timepoint will be presented with the pre- and post- morning GC dose 17-OHP values on separate plots. Similar graphs will be presented for the changes from baseline.

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

As a supplementary analysis, the change from baseline to Week 4 in 17-OHP, derived as a multiple of the ULN (based on age and sex), will be analyzed using the same method as described in the primary analysis of this endpoint.

9.5.2. Blood Pressure

Change from baseline to Week 24 in blood pressure is a secondary endpoint. For the purpose of analysis and summarization, the average of the triplicate sitting blood pressure values at each visit will be used. Systolic blood pressure (SBP), diastolic blood pressure (DBP), and mean arterial pressure (MAP; calculated as MAP = DBP + 1/3[SBP – DBP]) 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 24 in SBP, DBP, and MAP will be presented by treatment group along with the LS-mean treatment differences, 95% confidence intervals, and p-values.

Descriptive statistics will be presented by treatment group and visit for SBP, DBP, and MAP observed and change from baseline values.

9.5.3. Glucose Tolerance

Change from baseline to Week 24 in glucose tolerance (in subjects without diabetes mellitus) is a secondary endpoint. Subjects with a medical history of diabetes mellitus or who meet criteria for diabetes mellitus based on the Day 1 glucose tolerance test (fasting glucose ≥ 126 mg/dL [6.993 mmol/L] or post-glucose load ≥ 200 mg/dL [11.1 mmol/L]) or HbA1c (HbA1c ≥ 6.5%) will be excluded from this analysis. The analysis of this endpoint will be based on the post-GTT glucose values and performed using an ANCOVA model, as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 24 in glucose tolerance will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

The change in glucose values from pre-glucose load to post-glucose load will also be calculated by subject for the baseline and Week 24 visits. Descriptive statistics for the observed and change from baseline glucose values for the pre-glucose load, post-glucose load, and differences between the pre- and post-glucose loads will be summarized by treatment group.

Shifts from baseline to Week 24 in the post-GTT glucose level categories will be presented in tables. The categories are defined as follows: post-GTT glucose levels < 140 mg/dL (7.77 mmol/L) (normal blood glucose level), post-GTT glucose levels ≥ 140 mg/dL (7.77 mmol/L) and < 200 mg/dL (11.1 mmol/L) (impaired glucose tolerance/prediabetes), or post-GTT glucose levels ≥ 200 mg/dL (11.1 mmol/L) (diabetes mellitus). Each shift table will have four rows and four columns, with rows reflecting the categories at baseline as well as missing values and columns reflecting the categories at Week 24 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.

Shifts from baseline to Week 24 in pre-GTT glucose level categories will also be presented in tables. The categories are defined as follows: pre-GTT fasting glucose levels < 100 mg/dL (5.55 mmol/L) (normal fasting blood glucose level), pre-GTT glucose levels ≥ 100 mg/dL (5.55

mmol/L) and < 126 mg/dL (6.993 mmol/L) (impaired fasting glucose/prediabetes), or pre-GTT glucose levels ≥ 126 mg/dL (6.993 mmol/L) (diabetes mellitus). Each shift table will have four rows and four columns, with rows reflecting the categories at baseline as well as missing values and columns reflecting the categories at Week 24 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.

9.5.4. Waist Circumference

Change from baseline to Week 24 in waist circumference (in centimeters) is a secondary endpoint. The analysis of this endpoint will be performed using an ANCOVA model, as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 24 in waist circumference 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 waist circumference observed and change from baseline values.

9.5.5. Menstrual Regularity

Menstrual regularity at Week 24 (in female subjects of childbearing potential who are not on hormonal or intrauterine device contraceptives) is a secondary endpoint. Subjects meeting these criteria 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 this 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 4 to 8 menstrual cycles from baseline to Week 24. The number and percentage of subjects meeting the definition for regular menstrual cycle at Week 24 will be summarized by treatment group. This summary will be stratified based on menstrual cycle regularity at study entry (as collected on the CAH medical history eCRF, "Do menstrual cycles typically occur every 21-35 days?"). 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.

9.5.6. 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. Both local and central readings were performed for the ultrasound data. Only data from the central reading will be analyzed. 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.

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 24 in total TART volume as a percentage of total testicular volume is a secondary endpoint. The analysis of this endpoint will be performed in all male subjects (regardless of TART status at baseline) 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 24 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 24 for subjects with the presence of TARTs at baseline. Response categories and the criteria for each category are detailed in Table 9. The frequency and percentage of subjects meeting each classification at Week 24 (as compared to baseline) will be summarized by treatment group. Complete responder will be analyzed using a CMH test as described in Section 9.2.2. Complete or partial responders (pooled together) 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.

Table 9: Response Categories

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

ACTH, total testosterone, follicle stimulating hormone (FSH), luteinizing hormone (LH), sex hormone binding globulin (SHBG), and progesterone will be summarized with descriptive statistics by visit and treatment group. Note that for hormone parameters that were collected both pre- and post-GC dose, the following summaries will be produced for both pre-GC dose and post-GC dose values. Both observed and change from baseline values will be displayed. All summaries of testosterone and progesterone will be stratified by sex. Summaries of FSH, LH, and SHBG will also be stratified by sex as well as by combined sexes. The change from baseline to Week 24 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 24 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.

Shifts from baseline to Week 4 in gonadotropin values (LH and FSH) will be presented in tables by sex as well as by combined sexes. Postmenopausal women will be excluded from the shift tables. Each shift table will have four rows and four columns, with rows reflecting the reference range category at baseline (low, normal, high, missing), and columns reflecting the reference range category at Week 4 (low, normal, high, missing). 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.

The achievement of response at Week 24 (as defined for each of the following responder analyses described below) 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 (with the exception of sex) used in the randomization. Subjects who are missing Week 24 data will be omitted from the analysis.

A responder definition for the normalization of LH will be defined as subjects with abnormal LH levels at baseline (defined as < LLN or > ULN) and achieve normalization at Week 24 in male subjects (based on pre-GC dose LH levels). The number and percentage of subjects meeting this responder definition at Week 24 will be summarized by treatment group. This responder definition and summarization will also be repeated for FSH.

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 24 based on the average of the pre-glucocorticoid and post-glucocorticoid testosterone levels at the applicable visit. The number and percentage of female subjects meeting this responder definition at Week 24 will be summarized by treatment group.

A responder definition for the normalization of progesterone in female subjects of child-bearing potential not on hormonal contraceptives will be defined as female subjects with a progesterone level >0.6 ng/mL at baseline and decrease to a progesterone level <0.6 ng/mL at any point during the DB period of the study. The average of the progesterone levels will be used at visits where more than one hormone sample is collected. This analysis will be restricted to female subjects

who are of childbearing potential and are not on hormonal contraceptives. The number and percentage of female subjects in this subset meeting this responder definition at each visit in the DB period will be summarized by treatment group.

A responder definition for the ratio of androstenedione to testosterone (A4/T) in male subjects 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 24. The number and percentage of male subjects meeting this responder definition at Week 24 will be summarized by treatment group.

9.6.2. Urine Androgen Metabolite Levels

Urine androgen metabolite levels, including androsterone and etiocholanolone (corrected for urine creatinine), will be summarized with descriptive statistics by visit and treatment group. Summary tables will include observed and change from baseline values for each urine androgen parameter. The change from baseline to Week 24 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 24 each of these parameters will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

9.6.3. Metabolic Assessments

In addition to HOMA-IR and glucose tolerance, other metabolic assessments include a fasting lipid panel (with total cholesterol, low-density lipoprotein [LDL], and high-density lipoprotein [HDL], triglyceride) as well as glycated hemoglobin (HbA1c) (stratified by subjects with or without diabetes mellitus). These metabolic parameters will be summarized with descriptive statistics by visit and treatment group. Summary tables will include observed and change from baseline values for each lab parameter. The change from baseline to Week 24 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 24 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. Bone Mineral Density

BMD (assessed via DXA) will be summarized with descriptive statistics by visit and treatment group. Summary tables will include observed and change from baseline values for BMD (in g/cm^2) for each region of measurement.

The number and percentage of subjects meeting criteria for normal BMD (T-score $\geq$ -1.0), osteopenia (T-score < -1.0 to > -2.5), or osteoporosis (T-score $\leq$ -2.5) will be summarized by visit and treatment group. Descriptive statistics for observed and change from baseline in BMD T-scores will also be summarized by treatment group by visit and region of measurement.

In addition to this, the change from baseline to Week 24 in bone mineral density (for both g/cm^2 and T-score) 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 24 in BMD (for both g/cm^2 and T-score) will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

9.6.5. Body Composition (excluding fat mass)

Total lean body mass and total body mass (assessed via DXA) will be summarized with descriptive statistics by visit and treatment group. The summary table will include observed, change from baseline, and percent change from baseline values. The change from baseline to Week 24 for total lean mass and total body mass will each be analyzed using an ANCOVA model as described in Section 9.2.1. Descriptive statistics and LS means for change from baseline to Week 24 for total lean body mass and total body mass will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

Descriptive statistics for the total body mass (including the individual components of fat mass and lean mass) may also be displayed graphically by treatment group.

9.6.6. 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. 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 values.

The change from baseline to Week 24 in VAS score for hirsutism will be analyzed using an ANCOVA model as described in Section 9.2.1 (with the exception of sex as a stratification factor). Descriptive statistics and LS means for change from baseline to Week 24 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.

9.6.7. Acne

Visual analog scales (VAS) were used to assess the subject's perception of severity of acne 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 acne will be summarized descriptively by visit and treatment group. 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 values.

The change from baseline to Week 24 in VAS score for acne will be analyzed using an ANCOVA model as described in Section 9.2.1 (with the exception of sex as a stratification factor). Descriptive statistics and LS means for change from baseline to Week 24 in the VAS score for acne will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value.

9.6.8. Bone Markers

Serum osteocalcin and serum bone-specific alkaline phosphatase are bone marker measurements to assess bone formation. Serum C-terminal telopeptide (CTx) and urine N-terminal telopeptide (NTx) are bone marker measurements to assess bone resorption. Observed and change from baseline values in these bone markers will be summarized using descriptive statistics by treatment group and visit. The change from baseline to Week 24 for each of these bone marker

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 24 in each of the bone marker parameters will be presented by treatment group along with the LS-mean treatment difference, 95% confidence intervals, and p-value. If there are a sufficient number of postmenopausal women with bone marker values, all of the above analyses will be repeated in this subset of subjects.

9.6.9. 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 24. 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 24. The number and percentage of subjects meeting this responder definition at Week 24 will be summarized by treatment group. A shift table of hydrocortisone dosing frequency from baseline to Week 24 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 24. 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 plus prednisone (or equivalent) regimen at baseline, a responder is defined as a subject who switches to a hydrocortisone alone regimen at Week 24 (i.e, a subject who stops taking prednisone or equivalent prior to the Week 24 visit and is on a hydrocortisone-only regimen at Week 24). The number and percentage of subjects meeting this responder definition at Week 24 will be summarized by treatment group.

In subjects on a dexamethasone regimen at baseline, a responder is defined as a subject who switches to a hydrocortisone alone regimen at Week 24. The number and percentage of subjects meeting this responder definition at Week 24 will be summarized by treatment group.

In subjects on a dexamethasone regimen at baseline, a responder will also be defined as a subject who switches to a regimen that does not contain dexamethasone at Week 24. The number and percentage of subjects meeting this responder definition at Week 24 will be summarized by treatment group.

The achievement of (1) response for hydrocortisone alone dosing frequency reduction; (2) switching from prednisone (or equivalent) regimen to hydrocortisone alone regimen at Week 24; (3) switching from a dexamethasone regimen to hydrocortisone alone regimen at Week 24; and (4) switching from a dexamethasone regimen to a dexamethasone-free regimen at Week 24 (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 24 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)
- Glucocorticoid type – dexamethasone use (subjects on dexamethasone at baseline versus subjects not on dexamethasone at baseline)
- Glucocorticoid type – hydrocortisone vs. synthetic glucocorticoids (subjects on hydrocortisone alone versus subjects on prednisone [or equivalent] or dexamethasone)
- Sex (male versus female; based on the eCRF)
- BMI (< 30 kg/m^2 vs. ≥ 30 kg/m^2)
- Race
- Ethnicity
- Age (< 65 years vs. ≥ 65 years)
- Baseline total daily glucocorticoid dose (<20 mg/m^2/day versus ≥20 mg/m^2/day; based on the eCRF)

If any of these categories are sparse in number of subjects (e.g., <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, baseline glucocorticoid type, and baseline total daily glucocorticoid dose (<20 mg/m^2/day versus ≥20 mg/m^2/day), 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 will be 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 below summaries 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 below analyses and summary displays will start with Month 7 as the first postbaseline visit, unless otherwise specified.

10.1. 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 Month 12. 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 (Day 1) to Month 12 in glucocorticoid total daily dose. For the summary tables, the change from baseline to each postbaseline visit will start with Month 7 and conclude at Month 12. Figures of the mean (±SEM) observed, change from baseline, and percent change from baseline glucocorticoid total daily dose values (with and without the derivation of zero for loss of androstenedione control) will include all visits from baseline (Day 1) to Month 12.

The descriptive statistics summary table will also be produced for the glucocorticoid total daily dose (mg/day) observed, change from OL baseline, and percent change from OL baseline values (with and without the derivation of 0 for loss of androstenedione control).

Similar to the responder analysis defined in Section 9.4.2, 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 a visit, while androstenedione is adequately controlled at ≤120% of baseline or ≤ULN for age and sex, respectively, at that specific visit. Androstenedione control will be assessed based on the androstenedione values at baseline and each respective OL period visit through Month 12 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.

10.2. Hormone Measurements

Descriptive statistics will be presented by treatment group and visit for the observed and change from OL baseline values for the following hormone measurements in the OL period: ACTH, 17-OHP, androstenedione, total testosterone, LH, FSH, progesterone, and SHBG. For visits where the hormone measurement is collected more than once (eg, Month 12), each visit will be presented with the pre- and post- morning glucocorticoid dose hormone values at that visit. All

summaries of testosterone and progesterone will be stratified by sex. Summaries of FSH, LH, and SHBG will also be stratified by sex as well as by combined sexes.

Mean (±SEM) values of the above hormone measurements at each visit in the OL period will be summarized in line graphs by treatment group. For visits where the hormone is collected more than once, each visit will be presented with the pre- and post-morning glucocorticoid dose hormone values 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 (Day 1) to Month 12 in each of the hormone parameters.

Shifts from OL baseline to Month 12 in gonadotropin values (LH and FSH) will be presented in shift tables by sex as well as by combined sexes. Postmenopausal women will be excluded from the shift tables. Each shift table will have four rows and four columns, with rows reflecting the reference range category at OL baseline (low, normal, high, missing), and columns reflecting the reference range category at Month 12 (low, normal, high, missing). 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.

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 postbaseline visit, based on the average of the testosterone values at the applicable visit. Descriptive statistics will be presented by treatment group for the number and percentage of female subjects classified as responders and non-responders at each visit in the OL period.

A responder definition for the normalization of progesterone in female subjects of child-bearing potential not on hormonal contraceptives will be defined as female subjects with a progesterone level >0.6 ng/mL at OL baseline and decrease to a progesterone level <0.6 ng/mL at any postbaseline visit thereafter. The average of the progesterone levels will be used at visits where more than one hormone sample is collected. This analysis will be restricted to female subjects who are of childbearing potential and are not on hormonal contraceptives. Descriptive statistics will be presented by treatment group for the number and percentage of female subjects classified as responders and non-responders at OL baseline and each visit thereafter in the OL period.

A responder definition for the normalization of LH will be defined as subjects with abnormal LH levels at OL baseline (defined as < LLN or > ULN) and achieve normalization at Month 12 in male subjects based on pre-GC dose levels. The number and percentage of subjects meeting this responder definition at Month 12 will be summarized by treatment group. This responder definition and summarization will also be repeated for FSH.

A responder definition for the ratio of androstenedione to testosterone (A4/T) in male subjects 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 postbaseline visit thereafter. Descriptive statistics will be presented by treatment group for the number and percentage of male subjects classified as responders and non-responders at OL baseline and each visit thereafter in the OL period.

In addition to the serum hormone measurement summaries described above, urine androgen metabolite levels, including androsterone and etiocholanolone (corrected for urine creatinine), will be summarized with descriptive statistics by visit and treatment group. Summary tables will include observed and change from OL baseline values for each urine androgen parameter.

10.3. 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 (in subjects not taking insulin and fasting), glucose tolerance (as measured by the post-GTT load glucose levels), fasting lipid panel (including total cholesterol, LDL, triglycerides, and HDL), and HbA1c (stratified by subjects with and without diabetes mellitus). Percent change from baseline will also be summarized for all parameters except HbA1c.

The above summaries will be repeated for the change from baseline (Day 1) to Month 12 in each of the metabolic assessments.

10.4. Weight, Waist Circumference, BMI, Total Lean Mass, Total Fat Mass, and Total Body Mass

Descriptive statistics will be presented by treatment group and visit for the observed and change from OL baseline values for weight, waist circumference, BMI, percent total fat mass, total lean mass, and total body mass. Both absolute and percent change from OL baseline will be presented for weight.

The above summaries will be repeated for the change from baseline (Day 1) to Month 12 in weight, waist circumference, BMI, and percent total fat mass, total lean mass, and total body mass.

In the subset of subjects classified as overweight or obese (BMI $\geq$ 25 kg/m^2 at baseline), the number and percentage of subjects achieving a greater than 5 percent reduction in total body weight from baseline to Month 12 will also be presented by treatment group.

10.5. Menstrual Regularity

Menstrual regularity (in female subjects of childbearing potential who are not on hormonal or intrauterine device contraceptives) will also be summarized for the OL period. The number and percentage of subjects meeting the definition for having regular menstrual cycles (as defined in Section 9.5.5.) from OL baseline to Month 12 will be summarized by treatment group. This summary will be stratified based on menstrual cycle regularity at study entry.

In addition to the above summary, for premenopausal women assigned to the placebo group for the DB period (not using hormonal or IUD contraception), the number of menstrual periods from Week 24 to Month 12 will be compared to the number of menstrual periods from Day 1 to Week 24. The change in the number of menstrual periods between the two treatment periods will be summarized descriptively.

10.6. TARTs

Change from baseline to Month 12 in TARTs will also be summarized for the OL period. 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. Note that this summary will be repeated for the observed and change from OL baseline values.

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. Note that this summary will be repeated for the observed and change from OL baseline values.

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 the echogenicity will be summarized with descriptive statistics by visit and treatment group.

A response assessment will be conducted at Month 12 for subjects with the presence of TARTs at baseline. Response categories and the criteria for each category are detailed in Table 9 within Section 9.5.6. The frequency and percentage of subjects meeting each classification at Month 12 (as compared to baseline) will be summarized by treatment group.

10.7. Bone Mineral Density

BMD (assessed via DXA) will be summarized with descriptive statistics by visit and treatment group. Summary tables will include observed and change from baseline values for BMD (in g/cm^2) for each region of measurement.

The number and percentage of subjects meeting criteria for normal BMD (T-score $\geq$ -1.0), osteopenia (T-score < -1.0 to > -2.5), or osteoporosis (T-score $\leq$ -2.5) will be summarized by visit and treatment group. Descriptive statistics for observed and change from baseline in BMD T-scores will also be summarized by treatment group by visit and region of measurement.

The above summaries will be repeated for the change from OL baseline to Month 12 in BMD.

10.8. Hirsutism and Acne

Visual analog scales (VAS) were used to assess the subject's perception of severity of hirsutism and acne 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 OL baseline values in VAS scores for hirsutism and acne will be summarized descriptively by visit and treatment

group. 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 OL baseline values.

The above summaries will be repeated for the change from baseline to Month 12 in VAS scores for hirsutism and acne.

10.9. Bone Markers

Serum osteocalcin and serum bone-specific alkaline phosphatase are bone marker measurements to assess bone formation. Serum C-terminal telopeptide (CTx) and urine N-terminal telopeptide (NTx) are bone marker measurements to assess bone resorption. Observed and change from OL baseline values in these bone markers will be summarized using descriptive statistics by treatment group and visit. This summary will be repeated for the change from baseline to Month 12 by treatment group and visit.

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 during the DB period will be based on the DB Period Safety Analysis Set. Safety data will be summarized descriptively by treatment group. All outputs for safety endpoints during the OL period will be based on the OL Period Safety Analysis Set and 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 both the DB period and the OL period safety data 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) regardless of when the exposure to crinecerfont began and ended (i.e., duration of exposure to crinecerfont will be an overall measure of exposure over the DB and OL periods and will not be presented separately for the DB and OL periods).

Duration of exposure will be summarized with descriptive statistics. The number and percentage of subjects with the following exposure categories will also be presented:

- > 0 to < 4 weeks
- ≥ 4 weeks to < 24 weeks
- ≥ 24 weeks to < 9 months
- ≥ 9 months to < 12 months
- ≥ 12 months

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. The number and percentage of subjects in each treatment group who are dosing compliant (defined as >120%, ≥ 80% to ≤ 120%, and <80%) will be presented for each postbaseline visit where 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 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 the last dose of study drug in the OL period + 28 days will be considered a TEAE.

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 24) and for the 6-month OL period (Week 24 to Month 12) 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.

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 the DB period and OL period separately.

A listing of TEAEs resulting in premature study drug discontinuation will be provided, which will include subject ID, treatment period, treatment group, study day of the 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 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 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 will be summarized by descending preferred term in the crinecerfont group.

11.3. Clinical Laboratory Data

The hematology, clinical chemistry, coagulation, thyroid stimulating hormone (TSH), free thyroxine (T4), and PRA 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.

Two shift tables will be presented within the double-blind treatment period: shifts from baseline to Week 4 and shifts from baseline to Week 24. 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 Month 12 and shifts from OL baseline (Week 24) to Month 12.

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, including orthostatic blood pressures and heart rate (calculated as standing value minus sitting value), 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. If a subject has multiple standing vital signs collected during the same visit, the last one collected will be used to calculate whether the subject met criteria for orthostatic hypotension (defined as a drop from sitting to standing values in systolic BP of >20 mmHg or a drop from sitting to standing in diastolic BP of >10 mmHg). Observed and change from baseline values for the vital signs data and the number and percentage of subjects meeting the criteria for orthostatic hypotension will be summarized at each visit for the DB and OL periods separately.

11.5. 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 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 (BPRS)

The Brief Psychiatric Rating Scale (BPRS) is a clinician-rated tool designed to assess the severity of psychopathology in subjects with schizophrenia and other psychotic disorders. The BPRS includes 18 items that address somatic concern, anxiety, emotional withdrawal, conceptual disorganization, guilt feelings, tension, mannerisms and posturing, grandiosity, depressive mood, hostility, suspiciousness, hallucinatory behaviors, motor retardation, uncooperativeness, unusual thought content, blunted affect, excitement, and disorientation. The severity of each of the 18 items of the BPRS is rated on a scale of 1 (not present) to 7 (extremely severe) (total score range: 18 to 126). Higher scores represent greater symptom severity. Items that are not scored are left blank (i.e., not score as "0"). Rather than using the EDC total score value, the total score for BPRS will be re-calculated as the sum of the ratings for each of the 18 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 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 and then by preferred name.

12. PATIENT-REPORTED OUTCOMES

Patient-reported outcomes (PROs) capture the status of a subject's health condition and come directly from the subject, without interpretation of the subject's response by a clinician or anyone else. The following section describes the PROs that were administered to subjects throughout the study and the methods of scoring and summarizing these data. All outputs for the PRO summaries will be based on the Full Analysis Set. The analysis of the PRO data will be based on descriptive statistics and presented by treatment group according to the study visit unless otherwise noted. PRO data will not be imputed. No formal hypothesis-testing analysis of PRO data will be performed. Unless otherwise noted, the below summaries will be produced for both 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. 36-Item Short Form Health Survey (SF-36) V2

The SF-36 is a multipurpose, short form health survey with 36 questions that yields an 8-scale profile of functional health and wellbeing, as well as two psychometrically based physical and mental health summary measures. It consists of 8 health concepts: vitality, physical functioning, pain, general health perception, physical role limitations, emotional role functioning, social functioning, and mental health. The standard 4-week recall is used in this study. Each domain is scored based on a weighted sum of the questions in their section, on a 0-100 scale (each section is weighted equally). The lower the score the worse the disability. Calculation of SF-36 V2 score is via a licensed scoring software. Observed and change from baseline values will be summarized descriptively for the 8 health concepts and the 2 summary measures by treatment group for the DB period and the OL period. The study baseline will be used for the DB period summaries, and the OL baseline will be used for the OL period summaries.

12.2. EuroQoL 5 Dimensions 5 Levels (EQ-5D-5L) and Visual Analog Scale (EQ-VAS)

The EuroQol 5 Dimensions 5 Levels (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.

Subjects also rate 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 at each visit and change from baseline to each postbaseline visit will be summarized by treatment group.

The above summaries will be produced for the DB period and the OL period. The study baseline will be used for the DB period, and the OL baseline will be used for the OL period summaries.

12.3. Multidimensional Assessment of Fatigue (MAF)

The Multidimensional Assessment of Fatigue (MAF) is a 16-item scale that measures fatigue according to 4 dimensions: 1-degree and severity, 2-distress that fatigue causes, 3-timing of fatigue (over the past week, when it occurred and any changes), and 4-impact of fatigue on various activities of daily living (household chores, cooking, bathing, dressing, working, socializing, sexual activity, leisure and recreation, shopping, walking, and exercising). Subjects rate each item 1 through 14 using a 10-point numeric rating scale, and items 15 and 16 using categorical responses for timing. The results from the MAF are reported as a single score in the form of a Global Fatigue Index (GFI). Scores range from 1-no fatigue to 50-severe fatigue.

MAF scoring is a sum of the responses to questions 1-15. Item 16 is not included in the GFI. If respondents select 1 to question 1 (To what degree do you experience fatigue = "1 – not at all"), then they are instructed to stop the questionnaire, assign a score of 0 to items 2-16 and their score would be "1-no fatigue". If respondents select anything besides 1, they continue to questions 2-16. The GFI is calculated by converting item 15 to a 0-10 scale by multiplying each score by 2.5, summing items 1, 2, and 3; and averaging items 4-14 and the newly score item 15. A score will not be assigned to items 4-14 if the subject gave a response of "do not do any activity for reasons other than fatigue."

The GFI observed, change from baseline, and percent change from baseline values will be summarized descriptively by treatment group for the DB period and the OL period. The study baseline will be used for the DB period, and the OL baseline will be used for the OL period summaries.

12.4. Psychological General Well-Being Index (PGWBI)

The Psychological General Well-Being Index (PGWBI) is a validated 22-item scale that assesses psychological and general well-being in 6 domains: anxiety (5 questions), depressed mood (3 questions), positive well-being (4 questions), self-control (3 questions), general health (3 questions) and vitality (4 questions). Each question has a response option on 6-point Likert scale (0-5 in value) with a one month recall period. A high score means higher level of psychological wellbeing. The scores for all domains can be summed to provide a maximum score (up to 110 points).

Scoring for the PGWBI consists of 1 total score, ranging from 0-110 (22 questions X 0-5 per question) as well as 6 sub health scores.

Table 10: PGWBI Sub Health Scores

Health State	Number of Questions	Score Range
Anxiety	5	(5 x 5) = 0-25
Depressed Mood	3	(3 x 5) = 0-15
Positive well-being	4	(4 x 5) = 0-20
Self-Control	3	(3 x 5) = 0-15
General health	3	(3 x 5) = 0-15
Vitality	4	(4 x 5) = 0-20

The observed and change from baseline values for the PGWBI total scores and 6 sub scores will be summarized descriptively by treatment group for the DB period and the OL period. The study baseline will be used for the DB period, and the OL baseline will be used for the OL period summaries.

12.5. Medical Outcomes Study 12-Item Sleep Scale (MOS-12)

The Medical Outcomes Study 12-item Sleep Scale (MOS-12) is a reliable, valid tool for assessing changes in sleep. The MOS-12 is a subject-reported, non-disease-specific instrument for evaluating sleep outcomes and consists of 12 items to measure 6 sleep dimensions. The recall timeframe is the past four weeks.

The MOS-12 yields 6 scale scores:

- Sleep disturbance, which measures the ability to fall asleep and maintain restful sleep (4 items)
- Sleep adequacy, which measures the sufficiency of sleep in terms of sleeping enough to provide restoration of wakefulness (2 items)
- Sleep quantity, which measures (in hours) the amount of sleep an individual has had each night (1 item)
- Somnolence, which measures daytime drowsiness or sleepiness (3 items)
- Snoring (1 item)
- Shortness of breath, or headache (1 item)
-

Each item within a domain (except sleep quantity) is recalibrated on a 0-100 scale (with 0 as the lowest possible score and 100 as the highest possible score). Items within each scale (as bulleted above) are averaged together to create the 6 scale scores. Items that are left blank (missing data) are not used when calculating the scale scores. Scores represent the average for all items in the scale that the respondent answered. Note that the 6 scale scores will not be calculated or summarized as the CRFs collected data following the unrevised, older version of the MOS-12 scale.

The sleep problems index II uses 9 of the 12 items of the scale to compute an overall sleep problem summary. This index contains questions from the sleep disturbance, sleep adequacy, respiratory impairment, and somnolence domains but does not include the question on sleep quantity. The 3 questions in index II not included in index I are time to fall asleep, feeling that sleep was quiet, and feeling drowsy or sleepy during the day. Scoring of the MOS-12 will be based on the reference manual.

The observed, change from baseline, and percent change from baseline values for the quantity of sleep and the 2 sleep problem indices will be summarized with descriptive statistics by treatment group for the DB period and the OL period. The study baseline will be used for the DB period, and the OL baseline will be used for the OL period summaries.

13. 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 or collected on study but will not be summarized:

- Physical examination results (excluding waist circumference, weight, height)
- CYP21A2 genotyping
- Pregnancy test
- Urine drug screen
- Urinalysis
- Testicular ultrasound – locally read data
- Glucocorticoid dose diary data

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.

14. 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.

15. REFERENCES

Auchus RJ, Arlt W. Approach to the patient: the adult with congenital adrenal hyperplasia. J Clin Endocrinol Metab. 2013 Jul;98(7):2645-55.

Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. Biometrics. 1999;55(3):853-7.

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

Holm S. A simple sequentially rejective multiple test procedure. Scandanavian Journal of Statistics 1979; 6: 65-70.

Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. Statist Med. 2011;30(28):3267-84.

Speiser PW, Arlt W, Auchus RJ, Baskin LS, Conway GS, Merke DP, et al. Congenital Adrenal Hyperplasia Due to Steroid 21-Hydroxylase Deficiency: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2018 Nov 1;103(11):4043-88.

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.

Wiens BL. A fixed sequence Bonferroni procedure for testing multiple endpoints. Pharmaceut Statist. 2003;2:211-215.

16. APPENDICES

16.1. Implementation of Multiple Imputation for the Primary Endpoint

The primary efficacy endpoint requires the availability of both the glucocorticoid dose at Week 24 as well as androstenedione values used in the assessment of androstenedione control (as defined in Section 6.3.3). For subjects with incomplete or missing data at Week 24 (ie, missing glucocorticoid dose, androstenedione, or both), the following steps will be taken to multiply impute the missing data for use in the interim and final analysis of the primary endpoint.

Subjects with a missing glucocorticoid dose, androstenedione, or both at Week 24 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 24). First, the missing Week 24 post-dose androstenedione values will be imputed using a regression-based multiple imputation model and the model will include baseline androstenedione (post-GC dose), age, and sex (as specified in the eCRF). The subsequent FCS regression model will impute Week 24 glucocorticoid total daily doses (in $mg/m^2/day$) and the model will include baseline glucocorticoid dose ($mg/m^2/day$), baseline androstenedione (post-GC dose), Week 24 androstenedione (post-GC dose), and sex (as specified in the eCRF). Note that a minimum of 0 is set for imputed Week 24 post-GC dose androstenedione values and the Week 24 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 24, androstenedione "control" will be determined based on the imputed Week 24 androstenedione value and the percent change from baseline to Week 24 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 <u>on study drug</u> at Week 24 and are missing one or both components of the primary endpoint, their missing Week 24 data will be imputed from subjects within the same treatment group with non-missing data at Week 24 (Week 24 Completers).
- For subjects who are <u>off study drug</u> at Week 24 and are missing one or both components of the primary endpoint, their missing Week 24 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.

16.2. Implementation of Multiple Imputation for Key Secondary Endpoints of Serum Androstenedione, HOMA-IR, Weight, and Total Fat Mass

For the key secondary endpoints of serum androstenedione, HOMA-IR, weight, and total fat mass that require multiple imputation, the following steps will be taken:

1) For subjects on study drug without 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 and sex.

2) For subjects in the crinecerfont treatment group off study drug without 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 amount of Retrieved Dropouts, the endpoint 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 model will include the baseline value for the endpoint and sex.

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.

16.3. Hydrocortisone Dose Equivalent Conversion Factors

The following conversion factors will be used for non-hydrocortisone glucocorticoid medications to determine hydrocortisone dose equivalents.

Glucocorticoid	Hydrocortisone dose equivalent conversion
Methylprednisolone	4×
Prednisolone	4×
Prednisone	4×
Dexamethasone	60×

Auchus and Arlt, 2013; Speiser et al., 2018.

16.4. 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 16.1), if a subject was originally missing the Week 24 glucocorticoid daily dose value (mg/m^2/day), add *delta* (see rules below) to the imputed Week 24 glucocorticoid daily dose value. Calculate the percent change from baseline to Week 24 in the glucocorticoid daily dose for all subjects:

- If Week 24 glucocorticoid daily dose is non-missing: percent change from baseline = (Week 24 glucocorticoid daily dose value – baseline glucocorticoid daily dose value) / baseline glucocorticoid daily dose value x 100;
- If Week 24 glucocorticoid daily dose is missing and the subject is in either the crinecerfont treatment group or placebo treatment group: percent change from baseline = (Week 24 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 24 (as defined in the primary 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 24 for the glucocorticoid daily dose in each of the 100 imputed datasets using ANCOVA (as specified in 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 (i.e., >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.

Note that if the increment or decrement of delta is too large or too small, then the delta may be adjusted in order to make the number of comparisons reasonable. If significance is not lost or is lost slowly, some values of delta may be omitted from the final table. The tipping point will be defined as the combination of delta values that result in loss of significance.

16.5. Implementation of Tipping Point Analysis for the Key Secondary Endpoints

This section describes the tipping point analysis for the following key secondary endpoints: serum androstenedione, HOMA-IR, weight and percent total fat mass. The timepoint of interest for each endpoint is Week 24 with the exception of serum androstenedione where change from baseline to Week 4 is defined as the endpoint. Note that the guidance given in the text below for the tipping point analyses refers to Week 24; however, Week 4 will be used for analyses of serum androstenedione as defined in Section 9.4.1.

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

1) For the 100 imputed datasets created (as described in Section 16.2), if a subject was originally missing the Week 24 endpoint value, add *delta* (see rules below) to the imputed Week 24 endpoint value. Calculate the percent change or change from baseline (depending on the original endpoint) to Week 24 for all subjects as applicable for each key secondary endpoint:
    - If the Week 24 endpoint value is non-missing: percent change from baseline = (Week 24 value – baseline value) / baseline value x 100; and change from baseline = (Week 24 value – baseline value).
    - If the Week 24 endpoint value is missing and the subject is in either the crinecerfont treatment group or placebo treatment group: percent change from baseline = (Week 24 delta-adjusted imputed value – baseline value) / baseline value x 100 and change from baseline – (Week 24 delta-adjusted imputed value – baseline value).
2) Analyze the percent change from baseline (or change from baseline) to Week 24 for the values in each of the 100 imputed datasets using ANCOVA (as specified in 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 (i.e., >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 for each key secondary endpoint (ie, an analysis of the original, multiply imputed datasets)
- For subjects in the crinecerfont treatment group, the second delta will be defined based on the information in Table 11 below for each key secondary endpoint. If this is significant, continue to increase delta based on Table 11 until significance is lost.
- For subjects in the placebo treatment group, the second delta will be defined based on the information in Table 11 below for each key secondary endpoint. If this is significant, continue to decrease delta based on Table 11 until significance is lost.

Table 11: Delta-Adjustment Values

Key Secondary Endpoint	Second Delta		If still significant after the second delta, increase or decrease delta by the following:	
	Crinecerfont	Placebo	Crinecerfont	Placebo
Serum Androstenedione (nmol/L)	0.1745	-0.1745	0.1745	-0.1745

HOMA-IR	0.1	-0.1	0.2	-0.2
Weight (kg)	1	-1	1	-1
%Total Fat Mass	1	-1	1	-1

Note that if the increment or decrement of delta is too large or too small, then the delta may be adjusted in order to make the number of comparisons reasonable. If significance is not lost or is lost slowly, some values of delta may be omitted from the final table. The tipping point will be defined as the combination of delta values that result in loss of significance.

16.6. Implementation of Tipping Point Analysis for the Key Secondary Endpoint: Physiologic Glucocorticoid Daily Dose Responder

This section will discuss the tipping point analysis for physiologic glucocorticoid daily dose responders which is also a key secondary endpoint as described in Section 9.4.2. This key secondary endpoint is defined as the achievement of a reduction in glucocorticoid daily dose to physiologic levels (≤11 mg/m^2/day hydrocortisone equivalent adjusted for BSA) at Week 24, while Week 24 androstenedione is adequately controlled at ≤120% of baseline or ≤ULN for age and sex.

A tipping point sensitivity analysis will be performed to assess the robustness of the missing data assumptions for the glucocorticoid daily dose at Week 24. The following steps will be taken to implement the tipping point analysis for this key secondary endpoint:

1) For the 100 imputed datasets created (as described in Section 16.1), if a subject was originally missing the Week 24 value (i.e., missing the glucocorticoid daily dose at Week 24), add *delta* (see rules below) to the imputed Week 24 value. Calculate the change from baseline to Week 24 for all subjects as applicable:
    - If the Week 24 endpoint value is non-missing: change from baseline = (Week 24 value – baseline value).
    - If the Week 24 endpoint value is missing and the subject is in either the crinecerfont treatment group or placebo treatment group: change from baseline = (Week 24 delta-adjusted imputed value – baseline value).
2) If the subject does not have androstenedione control at Week 24 (as defined in the primary endpoint) then the change from baseline will be set to 0 prior to the analysis of the delta-adjusted imputed datasets.
3) Determine if each subject is a responder or a non-responder. Analyze the categorical data at Week 24 for the values in each of the 100 imputed datasets using the Cochran-Mantel-Haenszel (CMH) test. The CMH test will compare treatment groups (crinecerfont 100 mg bid vs. placebo) and will include the stratification factors used in the randomization. Use PROC MIANALYZE to combine the results and obtain the p-value.
4) If the p-value in the previous step was not significant (i.e., >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 for each key secondary endpoint (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. If this is significant, continue to decrease delta by 1 until significance is lost.

Note that if the increment or decrement of delta is too large or too small, then the delta may be adjusted in order to make the number of comparisons reasonable. If significance is not lost or is lost slowly, some values of delta may be omitted from the final table. The tipping point will be defined as the combination of delta values that result in loss of significance.