STATISTICAL ANALYSIS PLAN

Study Title:	A Randomized, Double-Blind, Placebo-Controlled, Dose-Ranging Study to Evaluate the Efficacy and Safety of SPR001 (Tildacerfont) in Adult Subjects with Classic Congenital Adrenal Hyperplasia
Study Number:	Study SPR001-203
Investigational Drug:	SPR001 (Tildacerfont)
Indication:	Treatment of Congenital Adrenal Hyperplasia
Investigators:	Multicenter
IND Number:	
Sponsor:	Spruce Biosciences, Inc. 611 Gateway Drive, Suite 740 South San Francisco, CA 94080
Plan Version:	2.0, 2 February 2024
Plan Prepared by:	Inka Leprince, MS Statistical Consultant PharmaStat, LLC

STATISTICAL ANALYSIS PLAN

Study SPR001-203

A Randomized, Double-Blind, Placebo-Controlled, Dose-Ranging Study to Evaluate the Efficacy and Safety of SPR001 (Tildacerfont) in Adult Subjects with Classic Congenital Adrenal Hyperplasia

Plan Version: 2.0, 2 February 2024

Plan version reviewed: 6 February 2024

Signature Page

DocuSigned by:
Saba Sile
714FA14B8C454FF..

06-Feb-2024
Date

Saba Sile, MD
VP of Clinical Development
Spruce Biosciences Inc.

DocuSigned by:
Will Charlton
41EB0D56E6FF4EF...

06-Feb-2024
Date

Will Charlton, MD
Chief Medical Officer
Spruce Biosciences, Inc.

DocuSigned by:
Kevin Carroll
45953F27ACF14DA...

06-Feb-2024
Date

Kevin Carroll, PhD.
External Statistical Expert
KJC Statistics Ireland Ltd.

Table of Contents

1 REVISION HISTORY 10
2 RELATED DOCUMENTS: PROTOCOL AND CASE REPORT FORMS .. 10
3 COMMITMENT TO GOOD STATISTICAL PRACTICE 10
3.1 Definition of Good Statistical Practice 10
3.2 Use of Standards 11
4 PURPOSE OF THE ANALYSIS PLAN 11
5 STUDY DESIGN 12
5.1 Randomization and Blinding 13
5.2 Study Treatment 14
5.2.1 Study Drug Administration 14
5.2.2 Glucocorticoid Changes 15
5.2.3 Stress Dosing of Glucocorticoid 16
5.2.4 Assessments 16
5.2.5 Efficacy Assessments 16
5.2.6 Safety Assessments 18
6 STUDY OBJECTIVES AND ENDPOINTS 23
7 SAMPLE SIZE AND POWER 26
8 ANALYSIS SETS 27
8.1 Intent-To-Treat Analysis Set 27
8.2 Modified Intent-To-Treat Analysis Set 27
8.3 Per-Protocol Analysis Set 27
8.4 Safety Analysis Set 27
8.5 Randomized Safety Analysis Set 28
8.6 Pharmacokinetic Analysis Set 28
9 GENERAL CONSIDERATIONS 28
9.1 Presentation of Summary Statistics 28
9.2 Presentation of p-values 29
9.3 Definitions and Derived Variables 29
9.3.1 Screened and Enrolled Subjects 29
9.3.2 Study Day 30
9.3.3 Treatment Study Day 30
9.3.4 End of Study Treatment Definition 30
9.3.5 End of Study Definition 30
9.3.6 Age 30
9.3.7 Age at Diagnosis of CAH 30
9.3.8 Body Mass Index 31

9.3.9 Body Surface area 31
9.3.10 Baseline Values 31
9.3.11 Change from Baseline 32
9.3.12 Percent Change from Baseline 32
9.3.13 Change from Baseline on the Log Scale 32
9.3.14 Geometric Mean Ratio and Percent Change 33
9.3.15 Upper Limit of Normal or Target by Efficacy Biomarker 33
9.3.16 Responder Definitions 33
9.3.17 HADS Total Score 34
9.3.18 SF-36 Total Score 34
9.3.19 Modified Ferriman-Gallwey Total Score 35
9.3.20 Homeostatic Model Assessment of Insulin Resistance 35
9.3.21 Testicular Adrenal Rest Tumor Volume 35
9.3.22 Cardiovascular Risk Factors 35
9.3.23 Study Drug Exposure Variables 36
9.3.24 Prior, Concomitant and Post treatment Medications 39
9.3.25 Adverse Events 40
9.3.26 Glucocorticoid Therapy 40
9.3.27 Variables Related to Menstrual Cycles 42
9.3.28 Geographic Regions 43
9.4 Analysis Windows 43
10 STATISTICAL AND ANALYSIS ISSUES 44
10.1 Adjustments for Covariates 44
10.2 Handling Dropouts or Missing Data 45
10.2.1 Handling of Efficacy Data 45
10.2.2 Handling of Questionnaire Data 45
10.2.3 Handling of Laboratory Data 45
10.2.4 Handling of Safety Data 46
10.3 Primary Analyses and Safety Data Monitoring 47
10.4 Multicenter Considerations 47
10.5 Multiple Comparisons, Multiplicity 48
10.6 Active-Control Studies 48
10.7 Examination of Subgroups 48
11 STUDY SUBJECTS 49
11.1 Subject Enrollment and Disposition 49
11.1.1 Screened and Enrolled Subjects 49
11.1.2 Randomized Subjects 49

11.2 Protocol Deviations 49
11.3 Demographics and Baseline Characteristics 50
11.3.1 Demographics 50
11.3.2 Baseline Characteristics 50
11.4 Medical History 51
11.4.1 General Medical History 51
11.4.2 Congenital Adrenal Hyperplasia History 51
12 STUDY DRUG AND OTHER MEDICATIONS 52
12.1 Exposure to Study Drug and Study Treatment Compliance 52
12.2 Glucocorticoid Therapy 53
12.3 Prior and Concomitant Medications 53
13 EFFICACY ANALYSES 54
13.1 Primary Efficacy Analysis 54
13.2 Final Efficacy Analysis 55
13.3 Primary Efficacy Analysis 55
13.3.1 Definition of A4 Dose Response Analysis 55
13.3.2 Dose Response Statistical Hypothesis 56
13.3.3 Dose Response Modeling and Testing 57
13.3.4 Summary of Descriptive Statistics for A4 58
13.3.5 Sensitivity Analyses 58
13.4 Secondary Efficacy Analyses 62
13.4.1 First Secondary Efficacy Analysis 62
13.4.2 Second Secondary Efficacy Analysis 64
13.4.3 Third Secondary Efficacy Analysis 65
13.4.4 Forth Secondary Endpoint Analysis 65
13.4.5 Fifth Secondary Endpoint Analysis 65
13.5 Exploratory Efficacy Analyses 66
13.5.1 A4 Responder Analysis 66
13.5.2 17-OHP Responder Analysis 69
13.5.3 TART Lesion Volume CFB at Week 18 69
13.5.4 Correlation of Tildacerfont Plasma Concentration vs Dose 70
13.5.5 Correlation of A4 Level versus 17-OHP Level 70
13.5.6 SF-36 Change from Baseline at Week 18 Analysis 71
13.5.7 Change from Baseline Analyses through Week 76 72
13.5.8 Evaluation of ATCH 73
13.5.9 Within Subject Correlation of A4 and 17-OHP with tildacerfont Plasma Concentration 73

13.5.10 Additional Exploratory Analyses 74
14 SAFETY ANALYSES 75
14.1 Adverse Events 76
14.1.1 Overall Summary of TEAEs 77
14.1.2 Summary of TEAEs by System Organ Class and Preferred Term 77
14.2 Clinical Laboratory Evaluation 77
14.3 Vital Signs 79
14.4 12-Lead Electrocardiogram 79
14.5 Psychiatric Evaluations 80
14.5.1 The Hospital Anxiety and Depression Scale 80
14.5.2 Columbia–Suicide Severity Rating Scale 80
14.6 Quality of Life 81
14.6.1 Patient Global Impression of Change 81
14.6.2 Clinical Global Impression – Improvement Scale 81
15 CHANGES RELATIVE TO THE PROTOCOL-SPECIFIED ANALYSIS 81
16 REFERENCE 83
17 APPENDIX A: SCHEDULE OF ASSESSMENTS 84
18 APPENDIX B: CLINICAL LABORATORY PARAMETERS 96
19 APPENDIX C: CALCULATING TOTAL SCORE FOR SHORT-FROM 3696
20 APPENDIX D: A4 DOSE RESPONSE OVER TIME BY RANDOMIZED TREATMENT ARM 96
21 APPENDIX E: CALCULATING ABSOLUTE CFB AND ABSOLUTE TREATMENT EFFECTS FOR A4 AT BASED ON ANALYSES CONDUCTED ON THE LOG SCALE. 102

List of Tables

Table 1 Study 203 Objectives and Endpoints 23
Table 2 Classification of Body Mass Index 31
Table 3 Study SPR001-203 Baseline Specifications for Specific Variables 32
Table 4 Upper Limit of Normal Values by Subgroup and Key Biomarker 33
Table 5 Responder Definition by Endpoint 34
Table 6 Spruce 203 Cardiovascular Risk Factors by Sex 36
Table 7 First Dose of Treatment Period by Dispensation 37
Table 8 Relative Potencies of Glucocorticoids in Hydrocortisone Equivalents 41
Table 9 Analysis Visit Windows (Clinic Visits Only) 43
Table 10 Potentially Clinically Significant Abnormalities by Laboratory Parameter 78
Table 11 Potentially Clinically Significant Abnormalities by Vital Sign Parameter 79
Table 12 Pre-Specified Threshold Levels for ECG Parameters 80
Table 13 Study SPR001-203 Protocol Schedule of Events at Screening (Protocol Version 7.0, March 14, 2022 84
Table 14 Study SPR001-203 Protocol Schedule of Post-Screening Events (Protocol Version 7.0 March 14, 2022) 86
Table 15 Study SPR001-203 Protocol Schedule of Post-Screening Events (Protocol Version 7.0 March 14, 2022) 94
Table 16 Clinical Laboratory Parameters 96

List of Figures

Figure 1 Schema of Study SPR001-203 Study Visits 21
Figure 2 Schema of Study SPR001-203 Optional Open-Label Extension Period Study Visits 22

List of Abbreviations and Definition of Terms

Abbreviation	Description
17-OHP	17-hydroxyprogesterone
A4	androstenedione
ACTH	adrenocorticotropic hormone, corticotropin
ADAM	analysis data model
AE	adverse event
AESI	adverse event of special interest
AIC	Akaike information criterion
BID	Twice daily
BLQ	Below level of quantification
BMI	body mass index
BP	blood pressure
C-SSRS	Columbia–Suicide Severity Rating Scale
CAH	congenital adrenal hyperplasia
CGI-I	Clinical Global Impression – Improvement Scale
CI	confidence interval
CFB	change from baseline
CRF	corticotropin-releasing factor
CRO	contract research organization
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
ECG	electrocardiogram
eCRF	electronic case report form
GC	glucocorticoid
GM	geometric mean
GMR	geometric mean ratio
HADS	Hospital Anxiety and Depression Scale
HbA1c	hemoglobin A1c
HC	hydrocortisone
HCe	hydrocortisone equivalent(s)
HDL	high-density lipoprotein
HOMA-IR	homeostatic model assessment of insulin resistance
HR	heart rate
IGA	Investigator's Global Assessment (of acne severity)
ITT	Intent to Treat (Population)

Abbreviation	Description
LDL	low-density lipoprotein
LFT	liver function test
log	logarithmic function
LS	least squares
MCP-Mod	Multiple Comparison Procedure – Modelling
MedDRA	Medical Dictionary for Regulatory Activities
MDRI	Multi Domain Response Index
mFG	modified Ferriman Gallwey (score for hirsutism)
mITT	modified Intent to Treat (Population)
MMRM	mixed-model for repeated measures
PCSA	potentially clinically significant abnormality(ies)
PGIC	Patient Global Impression of Change
PK	pharmacokinetic
PP	per-protocol
PT	preferred term
QD	once daily
QoL	quality of life
QTcF	Fridericia-corrected QT interval
REML	restricted maximum likelihood
RTSM	randomization and trial supply management
SAE	serious adverse event
SAP	statistical analysis plan
SF-36	Short Form 36
SOC	system organ class
TART	testicular adrenal rest tumor
TEAE	treatment-emergent adverse event
ULN	upper limit of normal
WHO DD	World Health Organization Drug dictionary

1 REVISION HISTORY

Version	Date	Document Owner	Revision Summary
1.0	31 December 2023	Inka Leprince	First version

2 RELATED DOCUMENTS: PROTOCOL AND CASE REPORT FORMS

Version	Date
Protocol Version 1.0	19DEC2019
Protocol Version 2.0	30JAN2020
Protocol Version 3.0	14FEB2020
Protocol Version 4.0	14APR2020
Protocol Version 5.0	31AUG2020
Protocol Version 6.0	31MAR2021
Protocol Version 7.0	14MAR2022
Case Report Forms	02JAN2023

3 COMMITMENT TO GOOD STATISTICAL PRACTICE

3.1 Definition of Good Statistical Practice

The International Council for Harmonisation Guideline on Statistical Principles for Clinical Trials (ICH E9) implicitly defines good statistical practice. Good statistical practice includes both appropriate statistical designs to minimize bias and maximize precision of analysis plus operational excellence to assure credibility of results. The scientific design associated with any clinical trial is found in the protocol. More detailed, pre-specified statistical analysis methods can be found in the statistical analysis plan.

We interpret the operational side of good statistical practice as a transparent, reproducible, and validated approach to acquiring and analyzing clinical trial data. Reproducible research depends upon process transparency and also provides auditability of the statistical analysis. Analysis transparency requires that a navigable electronic process chain exists from defining the objective of the analysis to creating the results.

3.2 Use of Standards

Data standards are foundational for creating an environment where tools can be leveraged at different points in the analysis process. Data standards for clinical development of drugs have been defined and are maturing under various initiatives through the Clinical Data Interchange Standards Consortium (CDISC). Spruce Biosciences uses standard data tabulation model (SDTM) data sets and Analysis Data Model (ADaM) statistical analysis files for producing analysis results. Other applicable standards include regulatory guidance from the Food and Drug Administration (FDA) and ICH:

- ICH Guideline on the Structure and Content of Clinical Study Reports (ICH E3)
- ICH Guideline for Good Clinical Practice (ICH E6)

4 PURPOSE OF THE ANALYSIS PLAN

This statistical analysis plan (SAP) pre-specifies the statistical analysis methods for supporting the completion of the clinical study report (CSR) of Study SPR001-203 for investigational product SPR001 (referred to as tildacerfont hereafter), an investigational drug candidate designed to treat classic congenital adrenal hyperplasia (CAH). This SAP will be used to analyze the primary efficacy, final efficacy, and safety data collected during the study. The analysis of the open label extension period will be covered in a separate SAP. Dose-exposure-response analyses combining data across multiple studies will be conducted using a separate pharmacokinetics SAP. The planned analyses identified in this SAP may be included in regulatory submissions and/or future manuscripts.

The analysis methods described in this plan are considered *a priori*, in that they have been defined prior to clinical database lock and treatment unblinding. Exploratory analyses that are not defined in this SAP, may be performed to support the clinical development program. Any post-hoc or unplanned analyses that are performed for the CSR, but not defined in this SAP, will be documented in the CSR. Changes from the planned analyses stated in the study protocol are described in Section 15. Should the SAP and the protocol be inconsistent with respect to any further planned analyses, the language of the SAP is governing.

The analysis of pharmacokinetic (PK) concentration data and its parameters, as well as the analysis of the optional open-label extension period will be captured in separate SAPs.

5 STUDY DESIGN

This is a multi-center, randomized, double-blind, placebo-controlled, dose-ranging clinical study with a 3-part Treatment Period and an optional Open-Label Extension Period. The 3-part Treatment Period will evaluate the efficacy and safety of up to 70 weeks of treatment with tildacerfont in adult subjects with classic CAH who have elevated biomarkers at baseline. At least 72 adult subjects will be randomized into the Treatment Period of the study.

A 6-week, single-blind, placebo Run-in Period, will assess subjects for adherence to their existing GC therapy and daily ingestion of placebo with food. Upon completion of a 6-week single-blind, placebo Run-in Period, subjects who meet all eligibility criteria will be randomized into the Treatment Period. Subjects will be randomized in a 1:1:1:1 ratio to receive either tildacerfont at a dose of 50 mg (once daily) QD, 100 mg QD, and 200 mg QD or matching placebo along the randomization stratum (Week 4 A4 $\leq 4 \times$ upper limit of normal [ULN] versus Week 4 A4 $> 4 \times$ ULN). The first 12 weeks of the treatment period constitute the randomized, double blind, placebo-controlled, dose finding treatment period (Part A).

Following Part A, all subjects will receive tildacerfont for duration of 12 weeks (Part B) with opportunities to up-titrate their tildacerfont dose up to 200 mg QD based on treatment response. Placebo subjects will receive the 100 mg tildacerfont QD dose at the beginning of Part B. Part B will have possible dose escalations at Week 18 and Week 24 for eligible subjects not meeting androstenedione (A4) response criteria as an assessment of dose titration for non-responders (A4 > upper limit of normal [ULN]).

Part C is a 46-week treatment period in which all subjects are escalated to receive 200 mg tildacerfont QD at Week 30, with opportunities for glucocorticoid (GC) reduction for eligible subjects (subjects meeting A4 $\leq$ ULN).

Subjects continuing into the optional Open-Label Extension Period will receive 200 mg tildacerfont QD for up to 240 weeks of treatment, with opportunities for GC adjustment for eligible subjects at each visit. Upon completion of the Open-Label Extension Period or completion of Treatment Period Part C for those subjects who do not continue into the Open-Label Extension Period, subjects will return to the clinic within 30 days following their last dose for a Follow-up visit.

Clinical visits during Parts A, B, and C in the Treatment Period, the Open-Label Extension Period, and Follow-up Period will include efficacy, biomarkers, safety, and pharmacokinetic (PK) assessments. Figure 1 depicts study visits from Screening Period to Follow-up Period and Figure 2 depicts study visits during the optional Open-Label Extension Period.

The maximum duration per subject of the main Treatment Period of Study SPR001-203 is expected to be up to 87 weeks (i.e., a maximum of 45-day Screening Period, a 6-week Run-in Period, a 70-week Treatment Period, and a 30-day safety Follow-up Period). Including the optional Open-Label Extension Period (240 weeks), the maximum duration per subject is expected to be up to 327 weeks. Subjects randomized to tildacerfont in Part A will receive up to 70 weeks of treatment with tildacerfont in the main study. For subjects randomized to placebo in Part A, this will include up to 58 weeks of treatment with tildacerfont in the main study. Assessments and procedures for evaluation of safety, efficacy, PK, and biomarkers will be conducted per the protocol-specified schedule (see Appendix A: Table 13, Table 14, and Table 15).

5.1 Randomization and Blinding

All subjects will receive single-blinded placebo during the Run-in Period. Eligible subjects will be centrally randomized to study intervention in Part A and assigned to the appropriate dose of tildacerfont in Parts B and C using a randomization and trial supply management (RTSM) system. Subjects will be randomized to study drug treatment in a 1:1:1:1 ratio to receive placebo or tildacerfont at dose levels of 50 mg, 100 mg, or 200 mg, stratified by the Week 4 A4 value (A4 ≤ 4 × upper limit of normal [ULN] versus A4 > 4 × ULN). In Parts B and C, subjects will

receive tildacerfont with possible tildacerfont dose escalation at Weeks 18, 24, and 30 as described in Section 5.2.1.

Investigators, subjects, and the Sponsor's contract research organization (CRO) will be blinded to assigned study drug treatment assignment in Part A and to dose of tildacerfont in Part B. To maintain the blind, all subjects will receive the same number of study drug tablets per dose, with the appropriate number of tildacerfont and/or placebo tablets for the assigned study intervention. The subject level blind will be maintained until the completion of the final analyses. At the time of the primary efficacy analysis—after all subjects have completed the Week 18 visit of the Part A double-blind Treatment Period and the data has been hard-locked for analysis —the Sponsor will be unblinded to group level results, but Investigators, study personnel who conduct the study, and the subject-level blind will be maintained until the completion of the final analyses.

5.2 Study Treatment

5.2.1 Study Drug Administration

During the Run-in Period, subjects will take placebo for 6 weeks. In Part A of the study (the first 12 weeks of the Treatment Period), subjects will receive randomized study drug treatment (i.e., tildacerfont 50, 100, or 200 mg QD or matching placebo) without dose escalation. During Part B (the second 12 weeks of the Treatment Period) and Part C (the last 46 weeks of the Treatment Period), all subjects will receive tildacerfont and the dose level will be determined as follows.

- Subjects who were randomized to placebo in Part A will receive 100 mg tildacerfont QD at the beginning of Part B (Week 18).
- Subjects who were randomized to 200 mg tildacerfont QD in Part A will continue to receive tildacerfont at 200 mg QD throughout Parts B and C.
- Based on A4 results from the Week 18 and Week 24 visits, eligible subjects may undergo a tildacerfont dose escalation after each of these study visits. The RTSM system will assign appropriate dose levels of tildacerfont based on subjects' A4 results (i.e., A4 ≤ ULN or A4 > ULN) at Weeks 18 and 24. Specifically:

  - Subjects taking tildacerfont 50 mg QD will either continue to receive the same dose (if A4 ≤ ULN) or be escalated to 100 mg tildacerfont QD (if A4 > ULN and the subject is tolerating his/her current tildacerfont dose)
  - Subjects taking 100 mg tildacerfont QD will either continue to receive the same dose (if A4 ≤ ULN) or be escalated to 200 mg tildacerfont QD (if A4 > ULN and the subject is tolerating his/her current tildacerfont dose).

- At Week 30, all subjects will be escalated to 200 mg tildacerfont QD.

During the optional Open-Label Extension Period, subjects will continue to receive tildacerfont at 200 mg QD.

Tildacerfont is produced as a 50 mg tablet. The placebo is produced to look identical to the tildacerfont tablet. To maintain the blind, all subjects will receive the same number of study drug tablets, and take 4 tablets a day with the appropriate number of tildacerfont and/or placebo tablets for the assigned. The 200 mg tildacerfont dose will contain 4 tildacerfont tablets in each compartment; the 100 mg tildacerfont dose will contain 2 tildacerfont tablets and 2 placebo tablets; the 50 mg tildacerfont dose will contain 1 tildacerfont tablet and 3 placebo tablets; and the placebo dose will contain 4 placebo tablets.

5.2.2 Glucocorticoid Changes

Subjects will use their own physician-prescribed supply of glucocorticoids (GCs) as prescribed during the study, including during the Run-in Period. During Part C of the Treatment Period, subjects with A4 ≤ ULN at Week 38 or Week 46 will begin to reduce their daily GC dose level by increments of no more than 5 mg/day in hydrocortisone equivalents (HCe), according to the protocol-defined conversion ratios of GC regimens summarized in Table 8, down to a minimum of 15 mg/day. Subjects may reduce to a dose below 15 mg HCe with medical monitor approval. GC reductions may stop at a higher GC level (>15 mg) based on Primary Investigator (PI) judgment. More details on GC dose adjustments are provided in Protocol Section 6.5.1.4.

5.2.3 Stress Dosing of Glucocorticoid

The Sponsor will provide a supply of oral hydrocortisone (HC) tablets for periods of stress dosing. During times of clinically significant physical stress, subjects are advised to take stress doses of HC, under advisement of their physician. Per the protocol, subjects who are currently stress dosing should reschedule efficacy and safety assessments until the event that initiated the stress dosing has resolved (see Protocol Section 6.5.1.5).

5.2.4 Assessments

Appendix A: Table 13, Table 14, and Table 15 show the schedule of events for the study.

5.2.5 Efficacy Assessments

Serum A4 collected at clinic visits for this study.

Additional efficacy assessments collected at clinic visits in Part A of the Treatment Period and/or the optional Open-Label Extension Period include:

- Serum 17-hydroxyprogesterone (17-OHP)
- Lesion volume of testicular adrenal rest tumors (TARTs) assessed by complete scrotal ultrasounds in in male subjects
- Quality of life assessments (Short Form-36 [SF-36] total score)
- Acne assessments (Investigator's Global Assessment [IGA] of acne severity score)
- Hirsutism assessments (modified Ferriman Gallwey [mFG] score for hirsutism)
- Total daily dose of GC regimen(s)
- Cardiovascular risk factors, as collected by metabolic assessments

5.2.5.1 Serum Biomarkers

All blood samples for hormones and androgens will be collected at 8 AM (7 – 9 AM window) prior to administering the morning GC dose after an overnight fast. A4, 17 OHP, adrenocorticotropic hormone (ACTH), and testosterone will be measured. Hormone assessments will be collected at all study visits.

5.2.5.2 Testicular Adrenal Rest Tumors

Scrotal ultrasounds will be obtained for adult male subjects to detect and to evaluate the size (volume) and number of TARTs. The initial scrotal ultrasound may be scheduled for any time up to the Week 4 visit and will be considered the baseline measurement for the TART endpoint. Only subjects with detectable TART(s) at baseline will be followed for TART changes at Week 18 and Week 76 of the Treatment Period, as shown in Table 14.

5.2.5.3 Quality of Life Assessments: Short Form 36

Quality of life will be measured using the SF-36, which will be administered on Day 1 of the Run-in Period; Week 6 (baseline), Week 18, and Week 76 of the Treatment Period, as shown in Table 14.

5.2.5.4 Hyperandrogenic Symptoms

5.2.5.4.1 Acne

Acne severity will be evaluated using an IGA score by the Investigator, which will be assessed at Week 6 (baseline), Week 18, and Week 76 of the Treatment Period, as shown in Table 14.

5.2.5.4.2 Hirsutism

Hirsutism will be evaluated in female subjects using an mFG score by the Investigator, which will be assessed at Week 6 (baseline), Week 18, and Week 76 of the Treatment Period, as shown in Table 14.

5.2.5.5 Glucocorticoid Dose Process and Documentation

At each visit, the prior day's type of GC, total GC dose, and distribution of daily doses, including timing of dosing will be recorded on a CRF called Key Glucocorticoid Dosing. If an adjustment in GC dose or change in distribution of the daily GC regimen is warranted, a new record will be entered in the Prior & Concomitant Daily Glucocorticoids/Mineralocorticoids eCRF. Changes to GC (for example: switch from short to intermediate) type during Part A of the study will be considered a protocol deviation.

5.2.5.6 Metabolic Assessments

Metabolic assessments include assessments of hypertension (systolic blood pressure [SBP] and diastolic blood pressure [DBP]), waist circumference, fasting assessments of insulin resistance (hemoglobin A1c [HbA1c], fasting glucose, fasting insulin, and homeostatic model assessment of insulin resistance [HOMA-IR]), assessments of dyslipidemia (total cholesterol, low-density lipoprotein [LDL] cholesterol, high-density lipoprotein [HDL] cholesterol, and triglycerides).

HbA1c, fasting glucose, fasting insulin, HOMA-IR, total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides will be assessed as part of clinical laboratory, described in Section 5.2.6.5. SBP, DBP, and waist circumference will be assessed as part of vital signs, described in Section 5.2.6.2.

5.2.6 Safety Assessments

Safety will be assessed by repeated clinical evaluations including adverse events (AEs), treatment-emergent adverse events (TEAEs), serious AEs (SAEs), AEs leading to discontinuation/withdrawal, AEs of special interest (AESIs), vital signs, physical examination, electrocardiograms (ECGs), clinical laboratory tests (serum chemistry, hematology, and urinalysis), and psychiatric evaluations for suicide risk by Columbia–Suicide Severity Rating Scale (C-SSRS).

5.2.6.1 Adverse Events

Investigators will collect information related to AEs will be collected throughout this clinical trial. All AEs occurring in all subjects will be collected following signing of the first informed consent until approximately 30 days after the last study drug treatment.

5.2.6.2 Vital Signs

The following vital signs will be assessed after the subject has been sitting for approximately 5 minutes: blood pressure (systolic and diastolic; mmHg); pulse rate (beats per minute); respiration rate (breaths per minute); and body temperature (°C); weight (kg); height (cm); waist circumference (cm). Vital signs will be obtained at Screening, and at every clinic visit during the Treatment Period. Height is measured at Screening. From these height and weight assessments,

body mass index (BMI; kg/m2) and body surface area (BSA; m^2) will be calculated. Waist circumference is measured at Screening, Week 6 (baseline), Week 18, and Week 76 of the Treatment Period, as shown in Table 14.

5.2.6.3 Physical Examinations

A full physical examination should include assessments of the cardiovascular, respiratory, GI, neurological, and musculoskeletal systems; head, eyes, ears, neck, and throat (HEENT); thyroid; skin; and extremities. The full physical examination may exclude rectal, genitourinary, and breast exams.

An abbreviated physical examination includes the following components: cardiovascular, respiratory, abdomen, musculoskeletal, HEENT, and skin.

A full physical examination will be performed during the Screening Visit and at Day 1, Week 18, and Week 76, as shown in Table 14.

5.2.6.4 Electrocardiograms

All 12-lead electrocardiogram (ECG) assessments will include heart rate, QRS, QT, and QTc intervals using Fridericia's formula. ECGs will be performed at the following visits: Screening, Week 6, Week 18, Week 30, and Week 76, as shown in Table 14.

5.2.6.5 Clinical Laboratory and Urinalysis

Clinical laboratory assessments include hematology, clinical chemistry, coagulation, lipid panel, thyroid panel, luteinizing hormone (LH), follicle-stimulating hormone (FSH), sex hormone-binding globulin (SHBG), renin, aldosterone, inhibin B for males only, and estradiol, prolactin, and progesterone for females only. Estimated glomerular filtration rate (eGFR) for screening will be calculated from blood creatinine measured as part of screening clinical chemistry. Clinical lab assessments will be performed at the following visits: Week 6, Week 10, Week 14, Week 18, Week 24, Week 30, Week 38, Week 46, Week 56, Week 58, Week 70, Week 76 and Week 80, as shown in Table 14.

5.2.6.6 Psychiatric Evaluations

5.2.6.6.1 Columbia–Suicide Severity Rating Scale

The C-SSRS will be used during the study to monitor suicidal ideation and behavior. The Baseline/Screening Version of the C-SSRS, which assesses both lifetime history and history from the last 12 months, will be used at screening to determine subject eligibility. The Since Last Visit Version of the C-SSRS will be used at all subsequent visits: Week 6, Week 18, Week 30, and Week 76, as shown in Table 14.

5.2.6.6.2 Hospital Anxiety and Depression Scale

Subject anxiety and depression will be monitored during the study using the HADS, assessed at Week 6, Week 18, Week 30, Week 58, Week 70, and Week 76, as shown in Table 14.

Figure 1 Schema of Study SPR001-203 Study Visits

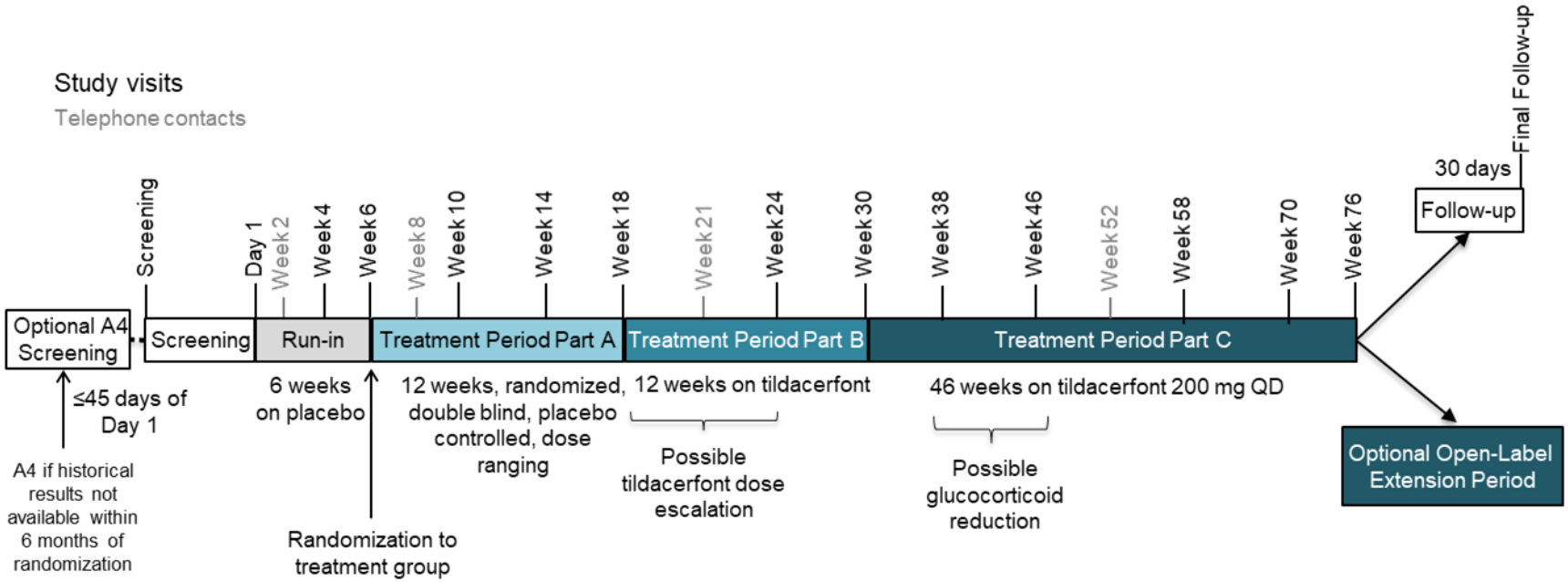
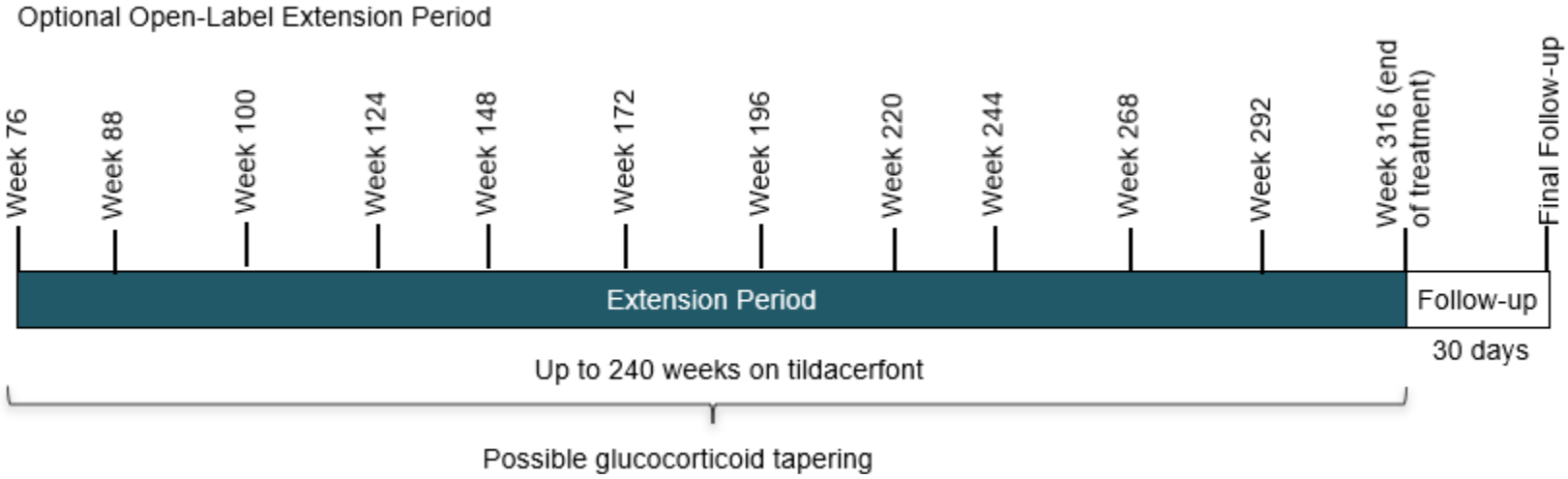


Figure 2 Schema of Study SPR001-203 Optional Open-Label Extension Period Study Visits



6 STUDY OBJECTIVES AND ENDPOINTS

This study will evaluate the potential of tildacerfont to reduce and control biomarkers (percent and absolute change in A4 and 17-OHP) in adult subjects with classic CAH who have elevated biomarkers at baseline on their current, supraphysiologic dose of GC replacement therapy. Part B and Part C will provide further monitoring for clinical outcomes with dose escalation to determine whether a higher dose may be necessary for hormone control in subjects with higher baseline hormone levels. Part C of this study and the optional Open-Label Extension Period will explore the potential GC-sparing effects of tildacerfont in subjects with elevated baseline A4 who are able to achieve A4 control on tildacerfont.

Study SPR001-203 will characterize clinical outcomes after up to 12 weeks of double-blind treatment and after up to 76 weeks of treatment with tildacerfont. The optional Open-Label Extension Period will provide additional characterization of up to 240 weeks of treatment with tildacerfont (see Table 1 for a summary of study objectives).

Table 1 Study 203 Objectives and Endpoints

Objective	Efficacy Analysis	
Primary Efficacy		
To evaluate the effect of tildacerfont in reducing A4 in subjects with classic CAH over 12 weeks	1.1.	Assessment of dose response for change from baseline in log A4 after 12 weeks on double-blind, placebo-controlled treatment (Week 18)
Secondary Efficacy		
To evaluate the effect of tildacerfont in reducing A4 in subjects with classic CAH over 12 weeks	2.1.	Change from baseline in A4 as assessed on the log scale after 12 weeks on double-blind, placebo-controlled treatment (Week 18)
To evaluate the effect of tildacerfont in reducing A4 in subjects with classic CAH over 12 weeks	2.2.	Absolute change from baseline in A4 as determined via application of the delta theorem to the log scale analysis results after 12 weeks on double-blind, placebo-controlled treatment (Week 18)

Objective	Efficacy Analysis	
To evaluate the effect of tildacerfont in reducing 17-OHP in subjects with classic CAH over 12 weeks	2.3.	Assessment of dose response for change from baseline in log 17-OHP after 12 weeks on double-blind, placebo-controlled treatment (Week 18)
To evaluate the effect of tildacerfont in reducing 17-OHP in subjects with classic CAH over 12 weeks	2.4.	Change from baseline in 17-OHP as assessed on the log scale after 12 weeks on double-blind, placebo-controlled treatment (Week 18)
To evaluate the effect of tildacerfont in reducing 17-OHP in subjects with classic CAH over 12 weeks	2.5.	Absolute change from baseline in 17-OHP as determined via application of the delta theorem to the log scale analysis results after 12 weeks on double-blind, placebo-controlled treatment (Week 18)
Exploratory Efficacy		
To evaluate the effect of tildacerfont in reducing A4 in subjects with classic CAH over 12 weeks	3.1.	Proportion of subjects who achieve adequate control of A4 (A4 $\leqslant$ 1x, 1.25x, 1.5x, 1.75x and 2x ULN) after 12 weeks on double-blind, placebo-controlled treatment (Week 18) for each treatment group
To evaluate the effect of tildacerfont in reducing 17-OHP in subjects with classic CAH over 12 weeks	3.2.	Proportion of subjects who achieve adequate control of 17-OHP (17-OHP $\leqslant$ 1200 ng/dL) after 12 weeks on double-blind, placebo-controlled treatment (Week 18) for each treatment group
To evaluate the effect of tildacerfont in reducing testicular adrenal rest tumor(s) (TART[s]) in male classic CAH subjects with TART(s) at baseline after 12 weeks on treatment	3.3.	Change in the lesion volume of TART(s) from baseline after 12 weeks on double-blind, placebo-controlled treatment (Week 18) for each treatment group
To characterize dose-exposure of Tildacerfont in CAH	3.4.	Correlation of Tildacerfont plasma concentration vs dose
To evaluate the effect of tildacerfont on biomarkers relative to each other in subjects with classic CAH over 12 weeks	3.5.	Estimate the within subject correlation of A4 and 17-OHP log scale values over time
To evaluate the effect of tildacerfont in improving quality of life (QoL) in subjects with classic CAH after 12 weeks on treatment	3.6.	Change from baseline in SF-36 total score after 12 weeks on double-blind, placebo-controlled treatment (Week 18) for each treatment group

Objective	Efficacy Analysis	
To evaluate the effect of tildacerfont in improving QoL in subjects with classic CAH	3.7.	Change from baseline at Week 76 in SF-36 total score across all ITT subjects
To evaluate the effect of tildacerfont in improving acne in classic CAH subjects with acne at baseline	3.8.	Change from baseline at Week 76 in IGA of acne severity score across all ITT subjects with acne at baseline
To evaluate the effect of tildacerfont in improving hirsutism in female classic CAH subjects with hirsutism at baseline	3.9.	Change from baseline at Week 76 in modified Ferriman Gallwey (mFG) score for hirsutism across all female ITT subjects with hirsutism at baseline
To evaluate the proportion of subjects who can reduce GC use in subjects with classic CAH	3.10.	Proportion of subjects with at least a 5 mg/day hydrocortisone equivalent (HCe) reduction from baseline in GC dose and A4 ≤ ULN at Week 76
To evaluate the effect of tildacerfont in reducing TART(s) in male classic CAH subjects with TART(s) at baseline	3.11.	Change in the total lesion volume of TART(s) from baseline at Week 76 across all male ITT subjects with TART(s) at baseline
To evaluate the effect of tildacerfont in reducing A4 in subjects with classic CAH over 12 weeks	3.12.	Assessment of dose response for change from baseline in log ATCH after 12 weeks on double-blind, placebo-controlled treatment (Week 18)
To evaluate the effect of tildacerfont in reducing ATCH in subjects with classic CAH over 12 weeks	3.13.	Absolute change from baseline in ATCH as determined via application of the delta theorem to the log scale analysis results after 12 weeks on double-blind, placebo-controlled treatment (Week 18)
To evaluate the effect of tildacerfont in reducing ATCH in subjects with classic CAH over 12 weeks	3.14.	Change from baseline in ATCH as assessed on the log scale after 12 weeks on double-blind, placebo-controlled treatment (Week 18)

Objective	Efficacy Analysis	
To characterize dose-exposure of Tildacerfont in CAH	3.15.	Estimate the within subject correlation of A4 and 17-OHP log scale values with tildacerfont plasma concentrations over time
Exploratory Efficacy (Optional Open-Label Extension Period) - Outside the Scope of this SAP		
To evaluate the effect of tildacerfont in reducing GC use to near-physiologic levels in subjects with classic CAH	4.1.	Proportion of subjects with GC dose ≤ 25 mg/day in HCe and A4 ≤ ULN at end of treatment (EOT)
To evaluate the percentage change in GC use in subjects with classic CAH	4.2.	Percent change from baseline in GC dose at EOT
To evaluate the effect of tildacerfont in reducing cardiovascular risk in subjects with classic CAH and at least one cardiovascular risk factor at baseline	4.3.	Proportion of subjects with improvement in at least one cardiovascular risk factor at EOT
To evaluate the effect of tildacerfont in eliminating TART(s) in male classic CAH subjects with TART(s) at baseline	4.4.	Proportion of male subjects without TART(s) at EOT who had TART(s) at baseline
Safety		
To evaluate the safety of tildacerfont in subjects with classic CAH	5.1.	AEs, SAEs

7 SAMPLE SIZE AND POWER

According to the protocol, a sample size of n = 15 subjects per group will provide at least 80% power to detect a difference of at least 50% in A4 mean percent change from baseline between a given dose level of tildacerfont and placebo, assuming a common SD of 30%.

An assumed dropout rate of 15% during the Treatment Period results in a target of 18 subjects per group and a total sample size of 72 subjects who enter the Treatment Period.

8 ANALYSIS SETS

8.1 Intent-To-Treat Analysis Set

The Intent-To-Treat (ITT) Analysis Set will include all randomized subjects regardless of Treatment Period eligibility or completion. The ITT Analysis Set will be the basis for demographics, baseline characteristics, and efficacy analyses.

8.2 Modified Intent-To-Treat Analysis Set

The following modified ITT (mITT) Analysis Sets will include all randomized subjects who receive at least 1 dose of study drug (tildacerfont or placebo), have at least 1 post-baseline A4 assessment, and did not change background GC through Week 18.

8.3 Per-Protocol Analysis Set

The Per-Protocol (PP) Analysis Set for Part A of the Treatment Period will include all eligible subjects who have no major protocol deviations that would affect the analysis of efficacy data. The PP Analysis Set may be used for supportive secondary analyses of the primary and secondary efficacy endpoints.

Major protocol deviations that would affect the analysis of efficacy data are listed as the following, but not limited to:

- Randomization error
- Subject did not meet efficacy-based inclusion criteria (elevated biomarkers)
- Subject met efficacy-based exclusion criteria (use of dexamethasone, night-shift worker), without waiver
- Subject's aggregate drug compliance was < 80% over the Treatment Period

Additional deviations will be identified prior to database lock and updated in this section.

8.4 Safety Analysis Set

The Safety (SAF) Analysis Set is defined as all subjects who received at least one dose of study drug (tildacerfont or placebo) during the Treatment Period, summarized by actual drug received.

8.5 Randomized Safety Analysis Set

The Randomized Safety (RSAF) Analysis Set is defined as all subjects who received at least one dose of randomized study drug (tildacerfont or placebo) during the Treatment Period. The randomized safety analysis set, which will be based on the actual treatment received, will be used for evaluation of general medical history, study drug exposure, concomitant medication, and safety.

8.6 Pharmacokinetic Analysis Set

The Pharmacokinetic (PK) analysis set will include all randomized subjects who receive at least one dose of tildacerfont and have at least one evaluable PK sample.

9 GENERAL CONSIDERATIONS

Data summarization and presentation conventions are documented in a separate document for mock-shells.

For all summaries of Part A by treatment, the columns will minimally include the randomized treatment groups of placebo, 50 mg tildacerfont, 100 mg tildacerfont, and 200 mg tildacerfont. All other summaries will use actual dose level unless otherwise specified.

9.1 Presentation of Summary Statistics

For most summary statistics, data will be analyzed and displayed in tabular format.

Unless otherwise specified, continuous serum hormones (A4, 17-OHP, ACTH) will be summarized using a 11-point descriptive statistics (i.e., n, mean, standard deviation [SD], median, 25% quartile [Q1], 75% quartile [Q3], minimum, maximum, geometric mean, geometric coefficient of variance [CV%], 95% confidence interval [CI] for geometric mean [including geometric mean ratio and its 95% CI]).

All other continuous variables will be summarized using an 8-point descriptive summary (n, mean, SD, median, Q1, Q3, minimum, and maximum) unless otherwise specified.

The same number of decimal places as in the observed value will be presented when reporting minimum and maximum; 1 additional decimal place than in the observed value will be presented when reporting mean, median, Q1, Q3, geometric mean, 95% CI; 2 additional decimal places than in the observed value will be presented when reporting SD. Geometric CV% will be reported to one decimal place. All percentages will be presented to 1 decimal place, unless otherwise specified.

All categorical/qualitative data will be presented using the frequency of events and percentages. Percentages equal to 100 will be presented as 100% and percentages will not be presented for zero frequencies. For summaries of AEs and concomitant medications (CMs), the percentages will be based on the number of subjects who received the study drug dose within the period that is being summarized.

All analyses and summaries will be produced using SAS® version 9.4 or higher.

9.2 Presentation of p-values

Results of statistical analyses will be reported using summary tables, listings, and figures (TLFs). The ICH of Technical Requirements for Pharmaceuticals for Human Use numbering convention will be used for all CSR TLFs.

Unless otherwise noted, all statistical testing will be two-sided and will be performed at the 0.05 significance level. Tests will be declared statistically significant if the calculated p-value is < 0.05, unless otherwise specified. P-values will be reported with 4 significant digits except when reporting p-values less than 0.001, reported as < 0.001.

9.3 Definitions and Derived Variables

9.3.1 Screened and Enrolled Subjects

9.3.1.1 Screened Subjects

Subjects who signed an informed consent form are considered Screened Subjects

9.3.1.2 Enrolled Subjects

Subjects who received at least one dose of placebo drug during the Run-in Period are considered enrolled.

9.3.2 Study Day

In this study, subjects will receive placebo in a 6-week single-blind placebo Run-in Period. Day 1 will be the first dose date of placebo in the Run-in Period. Study Day, which follows the Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) standard, is defined as (Assessment date - date of first study drug dosing + 1 day), where the assessment date is on or after the date of first study drug dosing; (Assessment date - date of first study drug dosing), where the assessment date is before the date of first study drug dosing.

9.3.3 Treatment Study Day

Randomized treatment will not begin until Week 6, Treatment Study Day 1 is therefore defined as the first dose date in Period A following the same format as Study Day. Visits will be summarized using Study Day, and treatment day will be included in select listings.

9.3.4 End of Study Treatment Definition

A subject is considered to have completed the main study treatment period if they completed the Week 76 visit. The end of study treatment is defined as the date of the last dose of study drug.

9.3.5 End of Study Definition

A subject is considered to have completed the main study if the subject has completed the three parts of the Treatment Period, including the final follow-up visit

9.3.6 Age

Age (years) is captured by the Demographics eCRF form at the Screening Visit.

9.3.7 Age at Diagnosis of CAH

Age (in years) at diagnosis of CAH is captured in the Congenital Adrenal Hyperplasia History eCRF form at the Screening Visit.

9.3.8 Body Mass Index

Body mass index (BMI, kg/m^2) is derived as weight (kg) / [height (m) × height (m)]. Height measured at Screening will be used to calculate BSA. Body mass index may be summarized using the classifications defined in Table 2.

Table 2 Classification of Body Mass Index

Parameter	BMI (kg/m^2)
Underweight	< 18.8
Normal Range	18.5 -< 25
Overweight	25 -< 30
Obese Class I	30 -< 35
Obese Class II	35 - < 40
Obese Class III	≥ 40

9.3.9 Body Surface area

Body surface area (BSA) is calculated per the Mosteller formula as sqrt ((height (cm) × weight (kg))/3600). Height measured at Screening will be used in all BSA calculations.

9.3.10 Baseline Values

Baseline values of serum biomarkers (i.e., A4, 17-OHP, and ACTH) are defined as the average of the Day 1, Week 4 and Week 6 assessments that are on or prior to the date of the first dose of double-blind study drug during Part A, allowing for unscheduled assessments to be windowed and replace missed scheduled visits. Visit windowing conventions will be used to identify the assessments that are used to construct baselines (see Section 9.4). Day 1 values are defined as the last non-missing assessment up to Day 14 for hormone assessments. For all other parameters, baseline values are defined as the last non-missing assessment prior to the first dose of randomized, double-blind study drug in Part A of the study. The following Table 3 provides a guide for the baselines to use in analysis.

Table 3 Study SPR001-203 Baseline Specifications for Specific Variables

Parameter	Screening	Study Week 6 (Baseline)
Patient reported outcomes (HADS, SF-36, C-SSRS)		X
Laboratory safety test (Hematology, Chemistry [includes insulin profile and lipid profile], Urinalysis, Thyroid Panel, Non-Efficacy Hormones)		X
Vital signs (includes weight profile and blood pressure profile)		X
Hirsutism (mFG score)		X
Acne (IGA score)		X
Scrotal ultrasound	X	
ECG		X
GC dose level		X

The initial scrotal ultrasound may be scheduled for any time up to the Week 4 visit and will be considered the baseline measurement for the TART endpoint.

9.3.11 Change from Baseline

Change from baseline is calculated as the post-baseline assessment subtracted by the baseline assessment.

9.3.12 Percent Change from Baseline

Percent change from baseline is calculated for the individual subject as the change from baseline divided by the baseline assessment and multiplied by 100%.

9.3.13 Change from Baseline on the Log Scale

The change from baseline on the log scale is calculated as the difference of the log (post-baseline assessment) subtracted by the log (baseline assessment). It is also mathematically equivalent to the log of the ratio of the post-baseline assessment divided by the baseline assessment.

$$change\ from\ baseline\ on\ the \log scale = \log(post\ baseline\ assessment) - \log(baseline)$$
$$= \log\left(\frac{post\ baseline\ assessment}{baseline\ assessment}\right)$$

9.3.14 Geometric Mean Ratio and Percent Change

For efficacy variables of A4, 17-OHP, and ACTH, the geometric mean ratio (GMR) and percent change from baseline will be calculated by treatment group and dose level. The GMR is derived from the natural logarithm of the post-baseline value divided by the baseline value [log(post-baseline/baseline)]. The GMR is the exponential function of the mean of log(post baseline/baseline). The 95% CI of GMR is the exponential function of the 95% CI of the mean [log(post baseline/baseline)]. The percent change is derived as 100 × (GMR - 1). The 95% CI of percent change is 100 × (95% CI of GMR - 1).

9.3.15 Upper Limit of Normal or Target by Efficacy Biomarker

The ULN and Target values that are displayed in Table 4 will be used in data analysis. The age at screening will be used to determine those age specific ULN values.

The following table lists the pre-specified, Sponsor-provided upper limit of normal values.

Table 4 Upper Limit of Normal Values by Subgroup and Key Biomarker

Baseline Characteristic	ACTH ULN	17-OHP ULN	17-OHP Target	A4 ULN
Males: 18-30	63.3 ng/dL	200 ng/dL	1200 ng/dL	220 ng/dL
Males: 31-50	63.3 ng/dL	200 ng/dL	1200 ng/dL	190 ng/dL
Males: 51+	63.3 ng/dL	200 ng/dL	1200 ng/dL	220 ng/dL
Females: pre-menopausal	63.3 ng/dL	200 ng/dL	1200 ng/dL	285 ng/dL
Females: post-menopausal	63.3 ng/dL	200 ng/dL	1200 ng/dL	75 ng/dL

9.3.16 Responder Definitions

The exploratory variables are proportions of responders based on each of the serum hormone assessments and GC. Indicator variables will be created for each of the three responder criteria as depicted in Table 5. Missing assessments are counted as non-responders.

Table 5 Responder Definition by Endpoint

Endpoint	Definition
Proportion of subjects who achieve A4 $\leq \boldsymbol{x}^{1}$ × ULN after 12 weeks on treatment (Week 18)	$\begin{cases} 1 & A4 \leq \boldsymbol{x} \times ULN \\ 0 & A4 > \boldsymbol{x} \times ULN \; OR \; missing \end{cases}$
Proportion of subjects who achieve 17-OHP ≤ 1200 ng/dL after 12 weeks on treatment (Week 18)	$\begin{cases} 1 & 17-OHP \leq 1200\, ng/dL \\ 0 & 17-OHP > 1200\, ng/dL \; OR \; missing \end{cases}$
Proportion of subjects with at least a 5 mg/day HCe reduction from baseline in GC dose and A4 ≤ ULN	$\begin{cases} 1 & \Delta GC\ dose \leq -5 mg/day\ in\ HCe \text{ AND } A4 \leq ULN \\ 0 & \Delta GC\ dose > -5\ mg/day \text{ OR } A4 > ULN\ OR\ missing \end{cases}$

1 = $\boldsymbol{x} = 1, 1.25, 1.5, 1.75$ and 2.

9.3.17 HADS Total Score

The scale consists of 14 items, 7 items each for anxiety and depression. The anxiety score is the sum of the odd-numbered items (questions 1, 3, 5, 7, 9, 11, 13), and the depression score is the sum of the even-numbered items (questions 2, 4, 6, 8, 10, 12, 14).

Each item is rated on a 4-point scale based on the frequency of symptoms over the preceding week and ranging from 0 (not at all) to 3 (very often). The sub-scores are the sum of individual item scores within each subcategory (anxiety or depression) and ranges from 0 to 21, with higher scores corresponding to higher levels of anxiety or depression. The total score of the individual item scores and ranges from 0 to 42, with higher scores corresponding to higher levels of anxiety or depression.

9.3.18 SF-36 Total Score

The SF-36 consists of 36 items in the following 8 domains: physical functioning, physical role functioning (limitations in usual role activities because of physical health problems), bodily pain, general health perceptions, vitality, social functioning, emotional role functioning (limitations in usual role activities because of emotional problems), and mental health. Domain scores range from 0 to 100, with higher scores corresponding to better subjective health status.

The SF-36 total score and 8 domain scores will be computed by the CRO providing the clinical trial database and provided, analysis ready, in relevant output datasets.

9.3.19 Modified Ferriman-Gallwey Total Score

The mFG scores each of 9 body areas on a scale of 0 (no hair) to 4 (hairiness typical of a man), with the sum of the separate scores providing a hormonal hirsutism score.

The total score is the sum of the individual item scores and ranges from 0 to 36, with higher scores corresponding to body hair patterns more typical of a man.

A subject with hirsutism at baseline is defined as an mFG score of 8 or more.

9.3.20 Homeostatic Model Assessment of Insulin Resistance

HOMA-IR is calculated using the following equations:

$$HOMA\text{-}IR = \frac{glucose\ (mg/L) * insulin(mIU/L)}{405}\ or\ \frac{glucose\ (mmol/L) * insulin(mIU/L)}{22.5}$$

9.3.21 Testicular Adrenal Rest Tumor Volume

For each TART assessment, the total number of TARTs will be captured as well as the size (lesion volume) of each TART individually. The total lesion volume of all detectible TARTs per subject will be calculated and will be compared against baseline total lesion volume.

9.3.22 Cardiovascular Risk Factors

Cardiovascular risk factors are derived from the metabolic risk factors. There are seven assessments spanning eight criteria that fit one of five cardiovascular risk factors.

Table 6 Spruce 203 Cardiovascular Risk Factors by Sex

Risk Factor	Assessment	Female Criterion	Male Criterion
Blood Pressure	Systolic Blood Pressure Diastolic Blood Pressure	≥ 120 mmHg; OR ≥ 70 mmHg	≥ 120 mmHg; OR ≥ 70 mmHg
Cholesterol	LDL Cholesterol Total Cholesterol	≥ 110 mg/dL; OR ≥ 170 mg/dL	≥ 130 mg/dL; OR ≥ 200 mg/dL
HOMA-IR	Fasting HOMA-IR	≥ 2.5	≥ 2.5
Waist Circumference	Waist Circumference	≥ 88 cm	≥ 102 cm
Body Mass Index	Body Mass Index	≥ 30 kg/m^2; OR at least one additional risk factor and BMI ≥ 27 kg/m^2	≥ 30 kg/m^2; OR at least one additional risk factor and BMI ≥ 27 kg/m^2

9.3.23 Study Drug Exposure Variables

9.3.23.1 Study Drug

Study drug is defined as either tildacerfont or placebo.

9.3.23.2 Study Treatment

Study treatment is defined more broadly as study drug or HC for stress dosing provided by the Sponsor.

9.3.23.3 Study Drug Exposure Duration (days)

The duration (days) of study drug exposure is derived from the first dose date and the last dose date of a given period of time. Specifically, the duration (days) is calculated as (LASTDAY - FIRSTDAY+ 1 day), where LASTDAY is the date of last dose, and FIRSTDAY is the date of first dose. If it becomes known that a patient has multiple consecutive missed doses, an adjustment to the calculation of exposure duration may be made.

The date of the first dose is captured by the CRF question, “Date first dose of study drug from this supply was taken by the subject?” For each treatment period, the first dose date is specifically collected from the following protocol-specified dispensations:

Table 7 First Dose of Treatment Period by Dispensation

Treatment Period	Dispensation Visit	Subsequent Visit
Run-in	Week 0	Week 4
Part A	Week 6	Week 10
Part B	Week 18	Week 24
Part C	Week 30	Week 38
Optional Open-Label Extension Period	Week 76	Week 88

In the absence of a scheduled dispensation, the first dose of study drug taken from an unscheduled dispensation that is windowed to the corresponding visit may be used**.** Otherwise, the date of the first dose of study drug from subsequent dispensation visit may be used.

The last dose date for a treatment period can be the last dose date of study drug treatment if the subject did not continue to the following treatment period, or one day prior to the date of the first dose date of the following treatment period.

9.3.23.4 Total Tildacerfont Dose (mg) Taken

Total tildacerfont dosage (mg) is defined as the dosage of tildacerfont taken over a specific study period *i* (where subject is receiving a constant dose), calculated as follows:

$$Total\ Tildacerfont\ Dose_i = \left((\#Dispensed - \#Returned - \#Lost\ or\ Destroyed) \times Tablet\ Strength\right)_i$$

and total tildacerfont dose over the entire study period is calculated as:

$$Total\ Tildacerfont\ Dose = \sum_{i=1}^{k} (Total\ Tildacerfont\ Dose)_i$$

where

Σ represents the summation operator and the value in parentheses is summed over the sequence $i = 1$ to k, where k = number of time intervals subject is intended to receive study drug, and tablet strength

Tablet strength is 0 mg per tablet if placebo or 50 mg tildacerfont per tablet if active drug.

9.3.23.5 Study Drug Compliance per Treatment Period

Study drug compliance is defined as the percentage of study drug actually taken compared to what was expected based on the randomized or titrated dose for time period *i* (where subject is receiving a constant dose). Using the drug accountability data, study drug compliance is calculated as follows:

$$\%Compliance_i = 100 \times \left(\frac{\#\ Dispensed - \#\ Returned - \#Lost\ or\ Destroyed}{\#\ Expected\ tablets\ to\ be\ taken}\right)_i$$

and total % compliance over a treatment period is calculated as:

$$\%Compliance = 100 \times \left(\frac{\sum_{i=1}^{k}(\#Dispensed - \#Returned - \#Lost\ or\ Destroyed)_i}{\sum_{i=1}^{k}(\#Expected\ tablets\ to\ be\ taken)_i}\right)$$

where

Σ represents the summation operator and the value in parentheses is summed over the sequence $i = 1$ to k, where

k = number of time intervals subject is receiving a constant dose. Possible constant dose levels are placebo, 50 mg QD tildacerfont, 100 mg QD tildacerfont, or 200 mg QD tildacerfont. At 50 mg QD tildacerfont dose level, 1 active tablet and 3 placebo tablets are taken per administration (i.e., 4 tablets per day); at 100 mg QD tildacerfont dose level, 2 active tablets and 2 placebo tablets are to be taken per administration (i.e., 4 tablets per day); at 200 mg QD tildacerfont, 4 active tablets are to be taken per administration (i.e., 4 tablets per day).

The expected number of tablets to be taken is the number of expected tablets per day (4 tablets) multiplied by the number of days in time period *i*.

9.3.23.6 Mean Daily Tildacerfont Dose (mg/day)

Mean Daily tildacerfont Dose (mg/day) is calculated as Total Tildacerfont Dose (mg) (Section 9.3.23.4) divided by Treatment Duration (days) for each treatment period (Part A, Part B, and Part C).

9.3.23.7 Diary Compliance

Subjects will fill out the daily GC and study drug diary through the Run-in Period and Part A. For all diary compliance, missing entries are counted as non-compliance. Daily study drug diary

compliance is calculated as the total count of number of 'Yes' responses to taking study drug as directed divided by (LASTDAY – FIRSTDAY+ 1 day), where LASTDAY is the date of last dose, and FIRSTDAY is the date of first dose.

Daily food diary compliance is calculated as the total count of number of 'Yes', with evening meal' responses to taking study drug with an evening meal, divided by (LASTDAY – FIRSTDAY+ 1 day).

Daily GC diary compliance is calculated as the total count of number of 'Yes' or 'Took Stress Dosing' responses to taking GC as directed and number of stress dosing days, divided by (LASTDAY – FIRSTDAY+ 1 day).

Weekly diary compliance is calculated as the total count of number of 'Yes', 'Yes, with evening meal' and/or 'Took Stress Dosing' responses to taking study drug as directed within the duration {Study Day (7n+1) to Study Day (7[n+1])}, divided by 7. Where n has values {0, 1, 2, ..., 18}. Missing entries are counted as non-compliance.

Aggregate diary compliance is calculated as (study drug compliance + food compliance + GC diary compliance) divided by 3.

9.3.24 Prior, Concomitant and Post treatment Medications

A prior medication is considered to be any medication that is taken within a year prior to the first randomized, double-blind study drug dosing.

A concomitant medication (CM) is considered to be any medication with start dates or stop dates within the Treatment Period. Specifically, CMs are medications: that are continued from Screening and continued after the first double-blinded study drug dosing; with start dates within the first dose date through last dose date + 24 hours; with end dates within the first dose date through last dose date + 24 hours; having a missing CM end date, or ongoing).

Post medications are considered to be any medication: taken after the last study drug dose date + 24 hours, having a missing CM end date, or ongoing.

9.3.24.1 Prohibited Concomitant Medications

Prohibited and cautionary concomitant medications are those medications with their potential for metabolic interactions with tildacerfont. Specific definitions can be found in Protocol Sections 6.5.2 and 13.1. Appendix 13.1 in the protocol provides a non-exhaustive list of medications.

9.3.25 Adverse Events

AEs are those AEs with onset date from the signing of the ICF until the end of the follow up period.

Treatment-emergent adverse events (TEAEs) are defined as any AEs, regardless of relationship to study drug, that have an onset or worsening in severity on or after the first dose of double-blinded study drug (within Treatment Period) until 30 days after the final dose of study drug (safety follow-up visit).

If a subject discontinues study drug but remains in the study, AEs with onset date after 30 days of last dose of study drug until the end of the follow up period will be considered post-treatment adverse events.

Related TEAEs are those reported by investigators as possibly related, probably related, or definitely related to the study drug.

AESIs are events that must be monitored on an ongoing basis as defined in Protocol Section 8.3.8 and identified in the AE eCRF page. The AESIs that are serious AEs (SAEs) will be handled according to guidelines from the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE 2017) version 5.0, as described in Protocol Section 8.3.3.1.

9.3.26 Glucocorticoid Therapy

GC therapies are concomitant medications as collected in the Prior & Concomitant Daily Glucocorticoids/Mineralocorticoids eCRF form.

9.3.26.1 Glucocorticoid Dose in Hydrocortisone Equivalents

Table 8 defines the potencies of various GCs relative to HC in treating CAH, as pre-specified and provided by the Sponsor. These conversion factors will be used to standardize and determine whether a subject's daily GC dose satisfies study eligibility criteria and provide general guidelines for GC tapering in Part C.

Table 8 Relative Potencies of Glucocorticoids in Hydrocortisone Equivalents

Glucocorticoid	Potency in CAH (HCe)
Hydrocortisone	1
Prednisone	4
Prednisolone	5
Methylprednisolone	5
Fludrocortisone	0
Dexamethasone	70
Cortisone Acetate	0.8

9.3.26.2 Glucocorticoid Total Daily Dose in Hydrocortisone Equivalents

Since a subject may be taking more than one dose (Morning, Midday, Midday 2, and/or Evening) of more than one GC medication, the total daily GC dose of a given subject at a given day is calculated as the sum of doses in HCe per day of concomitant GC therapies, excluding stress dosing.

9.3.26.3 Stress Event

A stress event is defined as an entry in the diary indicating a stress dose was taken. Consecutive days of stress dosing are counted as a single stress event.

9.3.26.4 Baseline Glucocorticoid Dose

Based on the GC therapies collected in the Prior & Concomitant Daily Glucocorticoids/Mineralocorticoids, the baseline GC dose is the GC Total Daily Dose in HCe at Treatment Study Day 1.

9.3.26.5 Glucocorticoid BSA Based Dose

The baseline GC BSA-based dose is calculated as the GC dose (Section 9.3.26.5) divided by the body surface area (Section 9.3.9) at the corresponding time point, rounded to two decimal places.

9.3.26.6 Glucocorticoid Dose at Visit

Based on the GC therapies collected in the Prior & Concomitant Daily Glucocorticoids/Mineralocorticoids, GC dose at visit is calculated as the sum of the total daily dose in HCe of all concomitant GC therapies at the date of visit as collected in the Date of Visit eCRF.

9.3.26.7 Change in Glucocorticoid Dose and A4 ≤ ULN

A change in GC dose meeting A4 ≤ ULN criterion is defined as follows:

GC dose change is calculated as the sum of the total daily dose in HCe of all concomitant GC therapies at Visit subtracted by the Baseline GC Dose.

If at Visit X A4 ≤ ULN, the GC dose change is recorded as observed. If at Visit X A4 > ULN, the GC dose change at the visit is recorded as missing.

9.3.27 Variables Related to Menstrual Cycles

9.3.27.1 Length (Days) of a Menstrual Cycle

Menstrual cycle length is calculated as (LASTDAY – FIRSTDAY+ 1 day), where FIRSTDAY is the date of the first day in a menstrual cycle where the subject records 'Yes' they menstruated today, and LASTDAY is the date of the last consecutive day a patient answered 'Yes' they menstruated today within the same menstrual cycle.

9.3.27.2 Time (Days) Between Menstrual Cycles

Time (days) between menstrual cycles is calculated as ($FIRSTDATE_i$ – $LASTDATE_{i-1}$ + 1 day), where $FIRSTDATE_i$ is the first date of current menstrual cycle and $LASTDATE_{i-1}$ is the last date of the previous menstrual cycle.

9.3.28 Geographic Regions

About 75 study centers will participate in this study. Three geographic regions will be classified by the countries where these study centers are located: North America, European Union and United Kingdon, and the rest of the world.

9.4 Analysis Windows

Analysis visits will match the protocol-specified visits (Screening, Day 1, Week 4, etc.). Unscheduled clinical visits may be used to capture missed or partially completed clinical visits; therefore, analysis visits and their windows are defined using Study Day (See Section 9.3.2). For the purposes of data analysis and summary, assessments and/or measurements, the protocol-specified visit will be preferentially flagged for analysis. In the absence of a protocol-specified visit, the analysis flag will be based on the collection date/time that is closest to the protocol-scheduled time point (or target Study Day). Analysis visit windows are presented in Table 9.

Table 9 Analysis Visit Windows (Clinic Visits Only)

Protocol Specified Visit Number	eCRF Visit Label	Analysis Visit	Target Study Day	Start (days)	Stop (days)
Visit 1a	SCR	Screening	-15	-45	-7
Visit 1b	SCR	Screening	-15	-45	-7
Visit 2	WK0	Day 1	1	-6	14
Visit 3	WK4	Week 4	29	15	36
Visit 4	WK6	Week 6	43	37	57
Visit 5	WK10	Week 10	71	58	85
Visit 6	WK14	Week 14	99	86	113
Visit 7	WK18	Week 18	127	114	148
Visit 8	WK24	Week 24	169	149	190
Visit 9	WK30	Week 30	211	191	239
Visit 10	WK38	Week 38	267	240	295
Visit 11	WK46	Week 46	323	296	365
Visit 12	WK58	Week 58	407	366	449
Visit 13	WK70	Week 70	491	450	512
Visit 14	WK76	Week 76	533	513	575
Visit 15	WK80	Follow-up	Last dose	Last dose +1	Last dose +30
Visit 1 OLE	WK76	Week 76	533	513	575

Protocol Specified Visit Number	eCRF Visit Label	Analysis Visit	Target Study Day	Start (days)	Stop (days)
Visit 2 OLE	WK88	Week 88	617	576	659
Visit 3 OLE	WK100	Week 100	701	660	785
Visit 4 OLE	WK124	Week 124	869	786	953
Visit 5 OLE	WK148	Week 148	1037	954	1121
Visit 6 OLE	WK172	Week 172	1205	1122	1289
Visit 7 OLE	WK196	Week 196	1373	1290	1457
Visit 8 OLE	WK220	Week 220	1541	1458	1625
Visit 9 OLE	WK244	Week 244	1709	1626	1793
Visit 10 OLE	WK268	Week 268	1877	1794	1962
Visit 11 OLE	WK292	Week 292	2046	1963	2130
Visit 12 OLE	WK316	Week 316	2213	2131	
Visit 12 OLE		Follow-up	Last dose	Last dose +1	Last dose +30

10 STATISTICAL AND ANALYSIS ISSUES

10.1 Adjustments for Covariates

The categorical, randomization stratification variable (i.e., Week 4 A4 value [0 = A4 ≤ 4 × ULN; 1 = A4 > 4 × ULN]) will be used in all efficacy analyses. The categorical variables treatment group (0 = placebo, 50 = 50 mg QD tildacerfont, 100 = 100 mg QD tildacerfont, and 200 = 200 mg QD tildacerfont) and sex (0=male, 1=female) will also be included efficacy analyses. The log baseline serum A4, as a continuous covariate, will be included in a model for the analysis of change from baseline in A4 without the randomization stratification variable. The baseline GC dose, as a continuous covariate, will be included in models for the analysis of serum A4 and the change from baseline in GCs. The baseline value of continuous efficacy endpoints will be included in statistical models, such analysis of covariance (ANCOVA) or mixed-model for repeated measures (MMRM) as a continuous covariate for exploratory endpoints.

When fitting a logistic regression model, the treatment will be coded as 0 = placebo, 50 = 50 mg QD tildacerfont, 100 = 100 mg QD tildacerfont, and 200 = 200 mg QD tildacerfont. The treatment will be an explanatory variable in the model.

10.2 Handling Dropouts or Missing Data

Every effort will be made by the Sponsor to ensure completeness of data collection. Missing data will be imputed as described in the following sections.

In the subject listing, all collected and all imputed values, if any, will be presented.

10.2.1 Handling of Efficacy Data

For missing A4 data, a retrieved dropout approach will be used which missing endpoint data for subjects who discontinued early will be imputed using data collected from subjects after discontinuation of study drug. If insufficient data exists ($n \leq 5$ per treatment group) to perform a retrieved drop out approach assuming missing at random (MAR), a multiple imputation approach will be used assuming missing not at random (MNAR) (Section 13.4).

10.2.2 Handling of Questionnaire Data

A total score for questionnaires will be calculated so long as > 50% of the items are non-missing. Otherwise, the total score will be set to missing.

10.2.3 Handling of Laboratory Data

A retest value will be used if the first test is invalidated, e.g., specimen hemolyzed. For non-PK laboratory values that are continuous in nature but are presented as either above or below the respective quantitation limits, the following imputations will be made for the purposes of summarization:

- If value is listed as < X, then the imputed value will be X/2
- If a value is listed as ≤ X, then the imputed value will be X
- If value is listed as > X, then the imputed value will be X+1
- If a value is listed as ≥ X, then the imputed value will be X

10.2.5 Handling of Safety Data

10.2.5.1 Adverse Events

If the time of onset (before or after intake of study drug) cannot be determined whether an AE is treatment-emergent because of a partial onset date, the event will be counted as a TEAE.

Adverse events with incomplete onset and end dates will have their dates imputed and then will be evaluated for their treatment-emergent status.

Incomplete adverse event onset date/time:

- If day is missing:
  - If month and year are the same as the month and year of first study drug exposure, then impute to the date of first study drug exposure.
  - Else if day is missing and year or month is not the same as the year and month of first study drug exposure, then impute as Day 1 of the month.
- If month is missing:
  - If year is the same as the year of first study drug exposure, then impute to the date of first study drug exposure.
  - Else if year is not the same as the year of first study drug exposure, then impute as January 1.
- If year is missing then impute to the date of first study drug exposure.
- If imputed start date is after AE end date, then impute the start date to the AE end date.

Incomplete adverse event end date/time:

If only have a YEAR, impute as December 31.
Else if only have YEAR and MONTH, then impute to last day of the month.
Otherwise missing, no imputation.If severity or relationship of an AE to study drug is not recorded, the severity or relationship will be imputed as “severe” or relationship as “possibly related,” for analysis purposes. All efforts will be made to ensue no missing severity or relationship of an AE to study drug prior to database lock finalization.

10.2.5.2 Concomitant Medications

If the start date of a medication is missing, the medication will be considered to have started prior to the study. Such a medication may also be considered concomitant, depending on the stop date or lack thereof. If the stop date of a concomitant medication is missing, then the medication

will be treated as ongoing. If the start date of a medication is missing, the stop date will be used to determine whether or not it is concomitant. Medications with other incomplete start dates or end dates will be imputed as follows:

Incomplete medication start date/time:

- If only have a YEAR, impute as January 1.
- Else if only have YEAR and MONTH, impute as Day 1 of month.
- Otherwise missing, no imputation.

Incomplete medication end date/time:

- If only have a YEAR, impute as December 31.
- Else if only have YEAR and MONTH, then impute to last day of the month.
- Otherwise missing, no imputation.

10.3 Primary Analyses and Safety Data Monitoring

The primary efficacy analysis will be conducted as when all subjects complete the Week 18 visit of the Treatment Period in Part A and the data has been hard-locked for analysis. The final analysis will occur after the last subject has completed the Week 76 visit or if the Sponsor discontinues the study.

An external, independent data safety monitoring board (DSMB) will review the available safety data (both blinded and unblinded) from this study periodically. The DSMB will advise the Sponsor of any trends or safety issues that may impact the study or study subjects and provide recommendations to the Sponsor if needed per the DSMB Charter.

10.4 Multicenter Considerations

This trial will be conducted at approximately 110 study centers in the United States, Europe, and other regions. Data from all study centers will be pooled for efficacy and safety analyses. Because the number of subjects at each center is likely to be small, no analyses will be performed by center.

10.5 Multiple Comparisons, Multiplicity

To control family-wise error rate, statistical hypothesis testing for the primary endpoint will be hierarchically dose ordered, testing first the 200mg dose versus placebo at 0.05 2-sided; then followed by the 100mg dose versus placebo at 0.05 2-sided if the 200mg dose reaches $p \leq 0.05$; and finally, the 50mg dose versus placebo at 0.05 2-sided if the 100mg dose reaches $p \leq 0.05$.

Secondary and exploratory endpoints will not be subject to alpha control, rather these endpoints will be interpreted nominally at the 2-sided 5% level.

10.6 Active-Control Studies

The placebo group will serve as a comparator in the primary analysis of this study.

10.7 Examination of Subgroups

Subjects will be categorized into the following common subgroups for the purposes of evaluating the primary and secondary efficacy endpoints.

- Baseline A4 (≤ 4 × ULN, > 4 × ULN)
- Baseline GC dose in HCe (≤ 30 mg/day, > 30 mg/day)
- Sex (Male, Female)
- Age
    - ≤ 30 years, > 30 years
    - ≤ median age of the ITT Analysis Set, > median age of the ITT Analysis Set
- Geographic Region ((i) North America, Not North America and (ii) NA, Latam, Europe, APAC)

Forest plots will be generated to display the primary efficacy, secondary efficacy, and select safety evaluations from the above subgroups, along with the overall results.

Additional subgroups specific to only select endpoints are specified in the appropriate section in the SAP.

11 STUDY SUBJECTS

11.1 Subject Enrollment and Disposition

11.1.1 Screened and Enrolled Subjects

Screening, enrollment and disposition will be summarized by country and by investigator for each treatment group and overall.

The percentages will be based on total subjects who enrolled in Run-in Period of the study. A subject listing will be provided with the above information, as well as geographic region, country, and site ID for subjects who were screened and enrolled in the Run-in Period, but not randomized.

11.1.2 Randomized Subjects

Dispositions of randomized subjects will be summarized by treatment group (including dose level) and overall. An additional summary will be included for subjects who were randomized and dose escalated. The subject disposition summary will include Parts A, B, and C of the Treatment Period.

A listing of disposition will be provided for all randomized subjects. A separated listing will present subjects who are excluded from SAF, RSAF, ITT, mITT or PP analysis sets, along with reasons for the exclusions.

11.2 Protocol Deviations

Protocol deviations will be listed by deviation type (e.g., major) and category (e.g, noncompliance with study procedures or restrictions). All deviations will be identified prior to database lock and will be summarized and presented in listing(s); listings will include flags for deviation type (major or minor). Major protocol deviations are defined per *Guideline for Industry: Structure and Content of Clinical Study* from International Council for Harmonisation

(ICH E3) guidance. All protocol deviations and major protocol deviations will be summarized by deviation category and type. In addition, a listing of all deviations and major protocol deviations will be provided.

11.3 Demographics and Baseline Characteristics

Demographics and baseline characteristics will be summarized and listed for all randomized subjects (ITT).

11.3.1 Demographics

Demographic characteristics will include age, age category (≤ 30 years, –> 30 years; ≤ median age of ITT Analysis Set, > median age of ITT Analysis Set), sex, race, ethnicity, and geographic region. A subject listing of demographics will be provided.

11.3.2 Baseline Characteristics

The following baseline characteristics will be summarized. They will be listed in subject listings in the following categories:

- Serum hormones (baseline value [Section 9.3.10])
- Metabolic baseline characteristics
    - Weight profile
        - Weight
        - Body mass index
        - Body mass index classification
        - Body surface area
    - Insulin Profile
        - HbA1c
        - Fasting glucose
        - Fasting insulin
        - HOMA-IR
    - Lipid Profile
        - Total cholesterol
        - LDL cholesterol
    - Blood Pressure Profile
        - Systolic blood pressure
        - diastolic blood pressure
- Quality of life assessments (SF-36 total score)

- Acne assessments
- Hirsutism assessments
- Lesion volume of TART(s)

11.4 Medical History

Medical history and CAH history will be summarized by randomized treatment group and dose level for ITT subjects. Individual subject listings will include subject identification number and randomized treatment/dose level (if a subject is randomized), along with general history and CAH history data collected from eCRF, for all enrolled subjects (i.e., who participated in the Run-in Period).

11.4.1 General Medical History

General medical history, besides CAH, will be mapped to preferred terms and system organ classes using the Medical Dictionary for Regulatory Activities (MedDRA®) Dictionary (version 23.0). General medical history will be summarized by system organ class (SOC) and preferred term (PT). A subject listing of general medical history will include start date/end date, verbatim medical history term, SOC, PT, and ongoing status.

11.4.2 Congenital Adrenal Hyperplasia History

11.4.2.1 Signs and Symptoms

Age at CAH diagnosis, type of CAH, and the signs and/or symptoms of CAH will be summarized. The most bothersome signs and/or symptoms of CAH will also be summarized. Signs and/or symptoms of CAH will be included in a subject listing.

11.4.2.2 Glucocorticoid History

The Baseline (Week 6) GC dose level in HCe, GC therapy medication(s), reasons for change to the current background GC therapy (if changed therapy) will be summarized. A subject listing will include all summarized variables, as well as any higher/lower dose of GC during adulthood, reason to reduce/increase to the current GC dose.

11.4.2.3 Hospital and Stress Dosing History

Hospitalization and stress dosing history will be summarized separately and will include the frequency and percentage of subjects with at least: a) one hospitalization for adrenal insufficiency/crisis in the past 5 years, b) at least one hospitalization for adrenal insufficiency/crisis in the past 12 months, c) at least one stress dosing event in the past 12 months. A subject listing will be included.

12 STUDY DRUG AND OTHER MEDICATIONS

12.1 Exposure to Study Drug and Study Treatment Compliance

Summaries of study drug exposure and compliance will be summarized overall and separately for those taken during Part A, Part B, and Part C of the study. The source for study drug dosing and compliance is drug accountability.

The summaries will include the total number of doses taken, mean daily dose (mg/day), duration (days) of exposure, and study drug compliance (see definitions in Section 9.3.23.4) by treatment group and dose level. A subject who received multiple dose levels of tildacerfont will be included in each of the dose levels for the purpose of summarization in separate rows for subjects with a dose escalation.

Furthermore, number and percentage of subjects with and without dose escalation of study drug will be summarized by randomized treatment group and dose level received for the overall summary. The denominators for percentages will be the number of subjects who received the randomized study drug treatment.

The drug accountability information, including number of tablets dispensed, number of tablets returned, number of tablets lost or destroyed, along with all summarized variables, will be listed in subject listings. Study drug diary compliance will be listed from the Run-in Period through the end of Part A, and all diary entry records will be provided in a separate listing.

12.2 Glucocorticoid Therapy

Concomitant GC therapy will be summarized overall and separately for those taken during Part A, Part B and Part C of the study. The source of GC dosing is the Prior & Concomitant Daily Glucocorticoids/Mineralocorticoids eCRF.

The following items will be summarized, overall GC diary compliance (Section 9.3.23.7), GC compliance by GC therapy type (i.e., hydrocortisone, prednisolone, prednisone, methylprednisolone, or other), proportion of subjects with a stress dosing event, number of unique stress dosing events (Section 9.3.26.4), duration of stress dosing events. All the above information will be included in a subject listing.

12.3 Prior and Concomitant Medications

All medication verbatim terms reported on the electronic case report form (eCRFs) will be mapped according to the World Health Organization (WHO) Drug Dictionary (WHO DD March 2020 B3). The medications will be mapped to Anatomical/Therapeutic/Chemical (ATC) class and preferred names.

Prior and concomitant medications (see definition in Section 9.3.24), other than glucocorticoids, will be summarized using WHO DD ATC class and preferred name. The summary results will be presented by treatment/dose level and Treatment Period (Part A, Part B, and Part C). Prior medications will be summarized by randomized treatment group. Concomitant medications will be summarized by actual dose taken concomitantly during each treatment period. For Part A, medications will be summarized by the actual randomized dose. For Part B, medications will be summarized under the treatment at the start of the medication. A subject who does-escalates in Part B will be counted within the denominator of each treatment group that s/he received. For example, a subject who started at 50 mg of tildacerfont, escalated to 100 mg of tildacerfont, and escalated again to 200 mg tildacerfont will be counted in all three dose levels; the medication that the subject starts during the treatment period will be counted once under the dose level that

the subject was taking at the start of the medication. For Part C, all subjects will receive 200 mg tildacerfont.

These summaries will present the number and percentage of subjects using each medication. Subjects may have more than one medication within an ATC class and preferred name. At each summary level subjects are counted once if they reported one or more medications at that level. Each summary will be ordered by descending frequency of incidence of ATC class and preferred name within each ATC class.

All medications will be provided in a subject listing with flags indicating study drug dose level. The variables include, but not limited to: Start date/end date/ongoing, medication name, ATC class and preferred name, indication, dose, unit, form, frequency, and route.

13 EFFICACY ANALYSES

13.1 Primary Efficacy Analysis

The Primary Efficacy Analysis is planned after all subjects have completed the study Week 18 visit in Part A of the study. During this first 12-week double-blind, placebo-controlled treatment period, there will be no dose escalation. The scheduled visits included in this treatment period include study Week 10, Week 14, and Week 18 corresponding respectively to 4, 8 and 12 weeks of double-blind randomized treatment. Efficacy Analyses will focus on inferences between tildacerfont treated subjects and placebo treated subjects. As shown in Table 1, the Primary Efficacy Analysis (1.1), the five Secondary Efficacy Analyses (2.1, 2.2, 2.3, 2.4 and 2.5), and selected Exploratory Efficacy Analyses (3.1 to 3.6 and 3.15) will be analyzed at the primary analysis. These Efficacy Analyses will be performed using the ITT Analysis Set (Section 8.2) as the main analyses.

13.2 Final Efficacy Analysis

The Final Efficacy Analysis will occur after the last subject has completed the study Week 76 visit or if the study is discontinued by the Sponsor. The data from study Week 18 through Week 76 will be frozen prior to the final analysis.

The scheduled visits included in this treatment period include study Week 24, Week 30, Week 38, Week 46, Week 58, Week 70, Week 76 visits. Exploratory Efficacy Analyses 3.7 to 3.11 and Safety Endpoints 5.1, as shown in Table 1, will be analyzed at the final analysis.

13.3 Primary Efficacy Analysis

The primary objective of the study is to evaluate the effect of tildacerfont in controlling A4 levels in adult subjects with classic CAH who have elevated biomarkers at baseline on their current, dose of GC replacement therapy over the 12-week, double-blind, placebo-controlled Part A of the Treatment Period. This objective will be achieved via A4 dose response analysis.

13.3.1 Definition of A4 Dose Response Analysis

The A4 dose response analysis is defined as follows:

Population: The analysis will be executed on a “treatment policy” basis, i.e. in the ITT analysis set, including all classic CAH subjects who were randomized into the study.

Variable: The analysis variable is the log-transformed change from baseline in serum A4 over the 12 week double blind period. The log-transformed change is calculated as log (A4 post baseline value) - log (A4 baseline value)

Intercurrent events: For the Primary Efficacy Analysis (and all Secondary Efficacy Analyses), intercurrent events such as changes in GC type or lack of treatment compliance or GC compliance, treatment interruption or discontinuation, use of prohibited medications are irrelevant as primary endpoint data are collected regardless and are included in the primary analysis. Stress dosing is also an intercurrent event that can substantially suppress A4 values and confound treatment effect. To adjust for this intercurrent event, the protocol design (Protocol

Section 6.5.1.5) has measures to minimize the impact of stress dosing on the primary endpoint, such as the delay of scheduled lab draws until 5 days after the subject has returned to their prescribed GC regimen.

13.3.2 Dose Response Statistical Hypothesis

For the Primary Efficacy Analysis, the null hypothesis is that there is no difference in the profile of log A4 change from baseline through 12 weeks of blinded randomized treatment between tildacerfont and placebo. The alternative hypothesis is that there is difference in the profile of log A4 change from baseline through 12 weeks of blinded randomized treatment between tildacerfont and placebo.

To assess this dose response hypothesis, repeated measures random coefficients mixed-effects model will be used:

$$y_{ijt} = (\beta_0 + \xi_i) + (\beta_1 + \eta_i)t + (\beta_2 + \theta_i)t^2 + \beta_{4j}I_j + \beta_{5j}I_j t + \beta_{6j}I_j t^2 + \beta_7 \kappa + \beta_8 G + \beta_9 D + e_{ijt}$$

where:

y_{ijt} is the log A4 change from baseline value for subject $i, i = 1, \ldots N$, at time t in treatment group j, $j = 1, 2, 3, 4$ ($1 = 200$mg tildacerfont, $2 = 100$mg tildacerfont, $3 = 50$mg tildacerfont and $4 =$ placebo), I_t is the indicator for randomised treatment (I_t= 1 for tilacerfont 200mg, =2 for 100mg, =3 for 50mg and =4 for placebo), κ is log baseline A4 value, G is the indicator for sex ($G = 1$ male, $G = 0$ female) and D baseline GC dose (in HCe),

β_0 is the fixed intercept effect.
β_1 is the fixed effect of the linear component of time
β_2 is the fixed effect of the quadratic component of time
β_{4j} is the fixed effect of treatment j parametrized such with $\beta_{44} = 0$
β_{5j} is the effect of treatment j on the linear component of time parametrized such that $\beta_{54} = 0$
β_{6j} is the effect of treatment j on the quadratic component of time parametrized such that $\beta_{64} = 0$
β_7 is the fixed effect of baseline log A4
β_8 is the fixed effect of gender

β_9 is the fixed effect of baseline GC dose

ξ_i is the random effect of subject i on the intercept

η_i is the random effect of subject i on the linear component of time

θ_i is the random effect of subject i on the quadratic component of time

e_{ijt} is the random error for subject i on treatment j and time t.

The random effects, ξ_i, η_i, and θ_i are assumed independent and identically normally distributed with variance components $\sigma_\xi^2, \sigma_\eta^2,$ and σ_θ^2, and, independently, the random error e_{ij} is normally distributed with variance σ^2.

The dose response Primary Efficacy Analysis therefore corresponds to the formal hypothesis statement,

$H_{0L}: \beta_{51} = \beta_{52} = \beta_{53} = 0 \quad vs \quad H_{1L}: \beta_{5j} \neq 0$ for at least one j

and

$H_{0Q}: \beta_{61} = \beta_{62} = \beta_{63} = 0 \quad vs \quad H_{1Q}: \beta_{6j} \neq 0$ for at least one j

The union intersection hypothesis $H_{0L} \cap H_{QL}$ represents the null that there is no dose response for tildacerfont versus placebo.

13.3.3 Dose Response Modeling and Testing

Statistical testing for dose response will be performed using the repeated measures random coefficients mixed-effects model described in Section 13.3.2. A4 values will be log transformed prior to analysis. The model will comprise of the log A4 change from baseline through to study Week 18, i.e. though 12 weeks of double blind randomized treatment; note the actual time, in days, from baseline to each post baseline A4 value through to study week 18 will be used in this analysis. Randomized treatment group and sex will be included in the model as class factors and baseline GC dose (in HCe) as continuous covariate. Subject specific intercept, time (in days)

and time x time random effects will be included. Further, randomized treatment by time and randomized treatment by time x time interactions will be included. The model will be fitted using maximum likelihood estimation. If the variance component for time or time x time are estimated to be zero then the corresponding term (or terms) will be retained in the model as fixed effects and the analysis re-run.

To test the union intersection hypothesis $H_{0L} \cap H_{QL}$, the model will be fitted with and without the randomized treatment by time and randomized treatment by time squared interactions terms. The difference in -2 log likelihood between models will be compared to the χ^2_{df} cumulative distribution function on the appropriate degrees of freedom (i.e. $df = 6$), this will provide the p-value to determine if the union intersection hypothesis can be rejected at the 10% alpha level; if so then there is evidence for a treatment effect of tildercerfont over placebo.

The fitted A4 change from baseline versus time profile will be extracted from the model for each randomized treatment and displayed graphically, superimposed upon the A4 time profiles for individual subjects.

Note, the programming of the repeated measures random coefficients mixed-effects model is particularly complex, hence example SAS codes perform the analysis are presented in Appendix D.

13.3.4 Summary of Descriptive Statistics for A4

Descriptive statistics (11-point) for the serum A4 assessments, change from baseline in serum A4, and % change from baseline in serum A4 will be presented by scheduled time point (i.e., Baseline [serum A4 only], study Week 10, Week 14, and Week 18). Geometric means and geometric mean change from baseline, along with the associated log scale SDs will also be presented.

13.3.5 Sensitivity Analyses

The Primary Efficacy Analysis will be supported by a number of sensitivity analyses.

13.3.5.1 Imputation of A4 values for Missing Data

Missing data can have an adverse effect on the estimation of endpoints. As such, multiple imputation of missing data via retrieved drop-out method will be used for the primary analysis. Retrieved drop-outs are defined as subjects who prematurely discontinue study drug treatment but remain in the study and still have assessments collected at the remaining scheduled visits in the study. These subjects can provide data that is reflective of subjects who discontinue treatment early without collection of assessments. Prior to performing the MI with retrieved drop-out, all intermittent (non-monotone) missing data will be imputed assuming missing at random (MAR).

A total of 20 datasets will be stochastically imputed for each intermittent missing A4 value at Week 18 using a mixed model repeated measures (MMRM) model. In this model A4 values will be log transformed prior to analysis. The model will comprise of the log A4 values at visits 5, 6 and 7 (= study weeks 10, 14 and 18) as the dependent variable with subject as random effect; randomized treatment group (0 = placebo, 50 = 50 mg QD tildacerfont, 100 = 100 mg QD tildacerfont, and 200 = 200 mg QD tildacerfont), time (=visit), sex (0=male, 1=female), and the interaction of treatment and time as fixed effects; and the log-transformed baseline serum A4 and the baseline GC dose (in HCe) as continuous covariates in the model. Within-subject correlations will be modeled using an unstructured covariance structure. Time ordering is a repeated measure within subjects. Errors for different subjects are assumed to be independent with an unstructured covariance structure.

In the event the MMRM model with an unstructured covariance structure does not converge, the following covariance structures will be used as substitution in the order below. Each subsequent covariance structure will be used only if each previous covariance structure was used and the model did not converge.

a) Toeplitz covariance structure (assuming measurements from samples taken closer together in time are more highly correlated than those from samples taken farther apart)

b) First order of auto-regressive [AR(1)] covariance structure (assuming measurements from samples taken closer together in time are more highly correlated than those from samples taken farther apart)
c) Compound symmetry covariance structure (assuming equal correlation for measurements from a subject, regardless of how far apart in time when they were taken)

Then a multiple imputation (MI) of the monotone missing data will be performed using information from retrieved dropouts, if data are available. If $n \geq 5$ subjects in a treatment group have discontinued study drug early and have available data from which to draw values, then a retrieved dropout approach will be utilized for that treatment group using MI by fully conditional specification and predictive mean matching assuming missing not at random (MNAR).

A total of *m* datasets will be stochastically imputed using multiple imputation with chained equations and predictive mean matching. Multiple imputation inference involves the following three steps:

1. The non-monotone missing observations are filled in m times to generate m complete datasets using SAS® PROC MI based on the multiple imputation method.
    a. m will be set to 20. The random seed will be set to 40
    b. Using the Markov chain Monte Carlo method with a non-informative prior (Jeffreys) to perform a monotone-data imputation (IMPUTE=MONOTONE) by the randomized treatment group
2. Each of the resulting *m* imputed datasets will then be imputed again. The monotone missing observations are filled in *m* times to generate *m* complete datasets using SAS® PROC MI based on the multiple imputation method.
    a. The random seed will be set to 50
    b. Using the retrieved drop out subjects with complete data, predict Week 18 A4 change using a set of predictors resulting in a set of linear coefficients *b*

$$log(A4\ postbaseline) = Treatment + Sex + Visit + Treatment * Visit + \log(Baseline\ A4) + Baseline\ GC\ Dose + Last\ available(weeks\ 10\ or\ 14)\ log(A4\ postbaseline)$$

c. Do a random draw from the *posterior predictive distribution* of b producing a new set of coefficients $b*$
d. Use $b*$ to generate predicted values for the Week 18 for all subjects
e. For each missing Week 18 A4 value, identify a set of candidates with observed Week 18 change whose predicted values are close to the predicted value for the case with missing data
f. From the candidates, randomly choose one Week 18 value and assign its observed value to the missing value
g. Repeat process for every subject with missing data

3. The analysis methods described in Section 13.3.3 will be applied to the *m* completed datasets.

   The results from the m completed datasets are combined using Rubin's rules via SAS® PROC MIANALYZE The imputed treatment effect, 95% CI and 2-sided p-value will be presented.

13.3.5.2 Per-Protocol Analysis and mITT Sets

The primary analysis of A4 as described in Section 13.3.3 will be repeated in the PP and mITT Analysis Sets (Section 8.3) as supportive analyses.

13.3.5.3 Subgroup Analyses

The primary analysis of A4 as described in Section 13.3.3 repeated for each subgroup specified in Section 10.7 for the ITT Analysis Set. A forest plot of the estimated treatment effects at Week 18 by dose level and subgroups will be provided.

13.3.5.4 Handling A4 Assessments Collected During GC Stress Dosing

13.4 Secondary Efficacy Analyses

The analysis population for Secondary Efficacy Analyses is the ITT analysis set, including all classic CAH subjects who were randomized into the study.

13.4.1 First Secondary Efficacy Analysis

The First Secondary Efficacy Analysis (see Table 1, Analysis 2.1) relates to the change from baseline in A4 as assessed on the log scale after 12 weeks on double-blind, placebo-controlled treatment. This analysis will be achieved as an integral part of the dose response modelling as described in Section 13.3.3.

The estimated mean change from baseline in log A4 values at 4, 8 and 12 weeks of double blind randomized treatment will be extracted from the model by treatment arm as will the difference and associated 95% CI and p-value between each tildacerfont dose and placebo at each of 4, 8 and 12 weeks. To facilitate ease of interpretation, the estimated mean change from baseline in log A4 for each tildacerfont dose and placebo will be back transformed resulting in the estimated ratio of geometric means to baseline, i.e. the back transformed log scale estimated mean change becomes {estimated A4 geometric mean at time x}/{estimated A4 geometric mean at baseline} = GMR. Hence, GMR − 1 becomes the percentage CFB in geometric means. Further, the difference between each tildacerfont dose and placebo in estimated mean change from baseline in log A4 (along with the associated 95% CI) will also back transformed resulting in the ratio of {A4 GMR tilda at time x}/{A4 GMR placebo at time x} as the measure of treatment effect. And, again, {A4 GMR tilda at time x}/{A4 GMR placebo at time x} −1 becomes the percentage reduction (or increase) in geometric mean A4 values at at time x for tildacerfont compared to placebo.

In addition orthogonal linear and quadratic contrasts between treatments at 12 weeks of double blind randomized treatment will also be estimated as a supportive measure of dose response at this time point.

Example SAS codes in relation to the First Secondary Efficacy Analysis are provided in Appendix D.

A serial plot will be presented of the estimated mean change from baseline in log A4 over time. The y-axis will be the percent CFB (i.e 100{exp(estimated change from baseline in log A4) -1}) and the x-axis will represent the Part A visits, baseline, Week 4, Week 8, and Week 12 of double blind randomized treatment. All treatment groups will be presented on the graph.

13.4.1.1 Supportive MMRM Analysis

A supportive analysis of the log change from baseline in A4 will be achieved via conventional MMRM modelling as described in Section 13.3.5.1. The log scale change from baseline (CFB) least squares (LS) means will be extracted from the model for each tildacerfont dose and placebo at study Weeks 10, 14 and 18, along with the associated SEs and 95% CIs. Further, the treatment effect for each tildacerfont dose versus placebo, being the corresponding difference in log scale LS means, will also be extracted along with the associated SE, 95% CI and 2-sided p-values.

To facilitate ease of interpretation, LS means, differences in LS means and associated 95% Cis s will be back transformed in the same fashion described in Section 13.4.1.

13.4.1.2 Supportive MMRM Tipping Point Analysis

A supportive MMRM sensitivity analysis using the tipping-point approach will be conducted. Tipping point MI analysis is a method to explore the influence of missingness on statistical inference by imputing the missing data in the drug arm with an increasing degree of penalization (shift parameter). The analysis finds a (tipping) point in this wider range of imputed values, at which conclusions change from being favorable to the experimental treatment to being unfavorable. After such a tipping point is determined, clinical judgment can be applied as to the plausibility of the assumptions underlying this tipping point. Tipping point MI is performed as follows:

1. Repeat MI Step 1 as described in Section 13.3.5.1 to generate multiple imputed data sets, with a specified shift parameter that adjusts the imputed values for observations in the treatment group, not the placebo group.
2. Repeat MI Step 2 for the imputed data sets with shift parameter applied.
3. Repeat MI Step 3 to obtain the p-value to see if the p-value is still ≤ 0.05.
4. Repeat steps 1-3 with a more stringent shift parameter applied until the p-value > 0.05.

13.4.2 Second Secondary Efficacy Analysis

The preceding estimated mean change from baseline in A4 is conducted on the log scale due to the inherent skewness and non-Normality of untransformed A4 values. While back transformation of log scale analysis results leads to interpretable quantities on the ratio scale, interest still resides in appreciation of the results on the absolute scale in ng/dL units.

Given the non-Normality of A4 values, a simplistic analysis of CFB = (Week 18 A4 value – baseline A4 value) is inappropriate as this simple CFB will also be non-Normal, resulting in upwardly biased estimates of mean A4 level and inflated SDs and SEs; consequently power will also be eroded.

Thus, to address the Second Secondary Efficacy Analysis (see Table 1, Analysis 2.2) and facilitate the presentation of results on both the ratio (i.e. % change) and absolute scales, the Primary Analysis repeated measures random coefficients mixed-effects model will be re-run exactly as described in Section 13.3.3 but with log A4 as the dependent variable as opposed to the change (i.e. as opposed to log A4 value minus log baseline value). Note, this re-run of the model with log A4 as the dependent variable will give precisely the same results, parameter estimates and variance-covariance matrix estimate as the analysis with log A4 minus log baseline as the dependent variable.

As described with example SAS codes in Appendix E, the delta theorem will be applied to the log scale results to provide consistent estimates of the CFB in A4 values on the absolute scale

together with as estimate for the difference in LSmeans and the associated 95% CI. Results will be presented for the absolute change in A4 at 4, 8 and 12 weeks of double blind randomized treatment. A serial plot will be presented of the estimated absolute mean change from baseline in log A4 over time. The y-axis will be the CFB in ng/dL and the x-axis will display baseline, week 4, week 8 and week 12 of double blind randomized treatment. All treatment groups will be presented on the graph.

13.4.3 Third Secondary Efficacy Analysis

The Third Secondary Efficacy analysis (see Table 1, Analysis 2.3) is, in principal, identical to the Primary Efficacy Analysis except that now the dependent variable in 17-OHP.

To evaluate the dose response for change from baseline in log 17-OHP through 12 weeks of double blind randomized treatment, the analysis and presentation of results will be conducted in exact concordance with the Primary Efficacy Analysis of A4 levels as described in Section 13.3 with the only difference being the removal of log A4 as a covariate and the inclusion of the randomization stratification variable (study Week 4 A4 value [0 = A4 ≤ 4 × ULN; 1 = A4 > 4 × ULN]) as a fixed factor effect.

13.4.4 Forth Secondary Endpoint Analysis

To address the Forth Secondary Efficacy Analysis regarding the change from baseline in 17-OHP on the log scale (see Table 1, Analysis 2.4), the analysis and presentation of results will be conducted in exact concordance with the Second Efficacy Analysis of A4 as described in Section 13.4.1 with the only difference being the removal of log A4 as a covariate and the inclusion of the randomization stratification variable (study Week 4 A4 value [0 = A4 ≤ 4 × ULN; 1 = A4 > 4 × ULN]) as a fixed factor effect.

13.4.5 Fifth Secondary Endpoint Analysis

To address the Fifth Secondary Efficacy Analysis regarding the change from baseline in 17-OHP on the absolute scale (see Table 1, Analysis 2.5), the analysis and presentation of results will be conducted in exact concordance with the Third Efficacy Analysis of A4 as described in

Section 13.4.2 with the only difference being the removal of log A4 as a covariate and the inclusion of the randomization stratification variable (study Week 4 A4 value [0 = A4 ≤ 4 × ULN; 1 = A4 > 4 × ULN]) as a fixed factor effect.

13.5 Exploratory Efficacy Analyses

13.5.1 A4 Responder Analysis

The Exploratory Efficacy Analysis 3.1 in Table 1 is the proportion of classic CAH subjects who were randomized into the study who achieve adequate control of serum A4 levels (defined as A4 ⩽ 1x, 1.25x, 1.5x, 1.75x and 2x ULN) 12 weeks from randomization. The efficacy of tildacerfont is evaluated by estimating the odds of meeting a given criterion in the tildacerfont treatment groups to the odds of meeting the criterion in the placebo treatment group. Subjects with missing A4 response data will be considered to non responders.

The proportion of subjects with A4 response will be summarized for each randomized treatment group as the number of subjects at study Weeks 10, 14 and 18 who have achieved a response divided by the total number of subjects randomized to that specific treatment.

13.5.1.1 Statistical Hypotheses

The null hypothesis for a given treatment comparison is that at Week 18 (i.e after the 12 week double-blind, placebo-controlled treatment period), the odds of response to tildacerfont to that of placebo is unity versus the alternative that the relative odds is not unity,

$$H_{0j}: \left\{\frac{\pi_j}{1-\pi_j}\right\}\left\{\frac{1-\pi_{plaecbo}}{\pi_{plaecbo}}\right\} = 1 \quad vs \quad H_{1j}: \left\{\frac{\pi_j}{1-\pi_j}\right\}\left\{\frac{1-\pi_{plaecbo}}{\pi_{plaecbo}}\right\} \neq 1$$

where j represents tildacerfont dose, $j = 50, 100$ and 200, π_j representes the true probability of response for tildacerfont dose j and similarly for $\pi_{plaecbo}$.

13.5.1.2 Statistical Modeling and Testing

A generalized estimating equations (GEE) model via SAS PROC GENMOD will be used to evaluate A4 response (defined as A4 ⩽ 1x, 1.25x, 1.5x, 1.75x and 2x ULN, with a separate

analysis for each ULN cutoff). The model will include response (= 0 or 1) as the dependent variable; treatment group (0 = placebo, 50 = 50 mg QD tildacerfont, 100 = 100 mg QD tildacerfont, and 200 = 200 mg QD tildacerfont), time (=study week 10, 14 and 18), sex (0=male, 1=female), and the interaction of treatment and time as categorical covariates; log baseline A4 and baseline GC dose (in HCe) as continuous covariates. The link function will be logit and distribution will be binomial. Subject IDs are used to identify the repeated measures of visits within subjects with an exchangeable working correlation structure. If the model fails to converge, the following working correlation structures will be tried in order until convergence is reached: unstructured, independent and AR(1). The model will use a binomial distribution with the logit link function and a significance level of 0.05 2-sided.

The odds ratio for response at study Week 18 for each tildacerfont dose level versus placebo group, together with the associated 95% CI and 2-sided p-value will be extracted from the model. The same will also be extracted for the odds ratio of response at study Weeks 10 and 14. SAS code will be of the form,

```
PROC GENMOD DATA=x;
CLASS SUBJECT TREATMENT VISIT;
MODEL RESPONSE = TREATMENT VISIT TREATMENT*VISIT / DIST=MULTINOMIAL
                    LINK=LOGIT TYPE3 AGGREGATE=TREATMENT;
REPEATED SUBJECT=PATIENT/TYPE=IND;
LSMESTIMATE TREATMENT 'OR 50MG V PCBO' -1 1 0 0 / EXP CL ALPHA=0.05;
LSMESTIMATE TREATMENT 'OR 100MG V PCBO' -1 0 1 0 / EXP CL ALPHA=0.05;
LSMESTIMATE TREATMENT 'OR 200MG V PCBO' -1 0 0 1 / EXP CL ALPHA=0.05;
RUN;
```

13.5.1.3 Graph Presentation

The proportion of responders from the model-based analyses will be displayed graphically over time with the y-axis representing the proportion meeting the responder definition and the x-axis the study Week 10, Week 14 and Week 18.

13.5.1.4 Supportive Analyses: Proportional Odds Analysis

In addition to serial evaluation of A4 response defined as A4 $\leq$ 1x, 1.25, 1.5x, 1.75x and 2x ULN, a combined A4 ordinal response will be analyzed, again via a GEE approach. Study Week

18 A4 response will be categorized for each subject as {A=[⩽ 1x]; B=[>1x to ⩽1.25x]; C=[>1.25x to ⩽1.5x]; D = [>1.5 to ⩽1.75x]; E = [>1.75 to ⩽2.0x]; and F=[>2x]} ULN. If there are too few subjects (n<5) in any of the categorizations A to F, the categorized will be collapsed to {$\tilde{A}$=[⩽ 1x]; $\tilde{B}$=[>1x to ⩽1.5x]; $\tilde{C}$=[>1.5x to ⩽2.0x]; and $\tilde{D}$=[>2x]}.

The model will include ordinal A4 response (= A, B, C, D, E, F; or $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, $\tilde{D}$ if categories are collapsed) as the dependent variable; treatment group (0 = placebo, 50 = 50 mg QD tildacerfont, 100 = 100 mg QD tildacerfont, and 200 = 200 mg QD tildacerfont), time (=study week 10, 14 and 18), sex (0=male, 1=female), and the interaction of treatment and time as categorical covariates; log baseline A4 and baseline GC dose (in HCe) as continuous covariates. The link function will be cumulative logit and distribution will be multinomial. Subject IDs are used to identify the repeated measures of visits within subjects with an exchangeable working correlation structure. If the model fails to converge, the following working correlation structures will be tried in order until convergence is reached: unstructured, independent and AR(1). The significance level will be 0.05 2-sided.

The cumulative odds ratio for response at study Week 18 to each tildacerfont dose level versus placebo group, together with the associated 95% CI and 2-sided p-value will be extracted from the model. The same will also be extracted for the cumulative odds ratio of response at 10 and 14 weeks. SAS code will be of the form,

```
PROC GENMOD DATA=x;
CLASS SUBJECT TREATMENT VISIT;
MODEL CATRESPONSE = TREATMENT VISIT TREATMENT*VISIT / DIST=MULTINOMIAL
                    LINK=CLOGIT TYPE3 AGGREGATE=TREATMENT;
REPEATED SUBJECT=PATIENT/TYPE=IND;
LSMESTIMATE TREATMENT 'CUM OR 50MG V PCBO' -1 1 0 0 / EXP CL ALPHA=0.05;
LSMESTIMATE TREATMENT 'CUM OR 100MG V PCBO' -1 0 1 0 / EXP CL ALPHA=0.05;
LSMESTIMATE TREATMENT 'CUM OR 200MG V PCBO' -1 0 0 1 / EXP CL ALPHA=0.05;
RUN;
```

13.5.2 17-OHP Responder Analysis

The Exploratory Efficacy Analysis 3.2 in Table 1 is the proportion of classic CAH subjects who were randomized into the study who achieve adequate control of serum 17-OHP levels (defined as 17-OHP ≤ 1200 ng/dL) after 12 weeks of randomized, double-blinded treatment.

These data will be analyzed in the exact same manner as Exploratory Efficacy Analysis 3.1 with the only difference being the exclusion of log A4 as a covariate and the inclusion of the randomization stratification variable (study Week 4 A4 value [0 = A4 ≤ 4 × ULN; 1 = A4 > 4 × ULN]) as a fixed effect in the mode1.

13.5.3 TART Lesion Volume CFB at Week 18

The Exploratory Efficacy Analysis 3.3 in Table 1 is the CFB in TART lesion volume after 12 weeks of randomized, double-blinded treatment for those who had detectable TARTs at Baseline in the Male ITT population. The analysis will be constrained to only those subjects with evaluable baseline and post baseline ultrasound results. Missing data will not be imputed.

13.5.3.1 Statistical Modeling and Testing

Statistical analysis will be achieved via an ANCOVA model. If non-Normality is evidenced via Normal scores plot and plots of predicted versus residual values, the analysis will be accompanied by a rank-based ANCOVA model. The change from baseline to study Week 18 (12 weeks of randomized treatment) on the log scale will be the dependent variable; randomized treatment group (0 = placebo, 50 = 50 mg QD tildacerfont, 100 = 100 mg QD tildacerfont, and 200 = 200 mg QD tildacerfont) as a fixed effect; baseline TART Volume and the baseline GC dose (in HCe) as continuous covariates; the randomization stratum (A4 ≤ 4 × ULN versus A4 > 4 × ULN) as categorical explanatory variables. The estimation method for the ANCOVA model will REML. A significance level of 0.05 2-sided will be applied.

The ranked value is obtained by ranking all non-missing change from baseline values at the specified analysis visit, if a rank-based model is used.

13.5.3.2 Reporting Results

Summary of Descriptive Statistics

Descriptive statistics (11-point) for TART lesion volume, change from baseline in TART lesion volume, and % change from baseline will be presented by scheduled time point (i.e., Baseline and study Week 18). Geometric means and geometric mean change from baseline, along with the associated log scale SDs will also be presented.

Presentation of Analysis Results

The log scale change from baseline (CFB) least squares (LS) means will be extracted from the model for each tildacerfont dose and placebo, along with the associated SEs and 95% CIs. Further, the treatment effect for each tildacerfont dose versus placebo, being the corresponding difference in log scale LS means, will also be extracted along with the associated SE, 95% CI and 2-sided p-values.

To facilitate ease of interpretation, the log scale CFB LS means for each tildacerfont dose and placebo and the difference between each tildacerfont dose and placebo in log scale CFB LS means (along with the associated 95% CI) will be back transformed as described for the Primary Efficacy Analysis.

13.5.4 Correlation of Tildacerfont Plasma Concentration vs Dose

Exploratory Efficacy Analysis 3.4 in Table 1 is the correlation of tildacerfont plasma concentration versus dose. This will be explored by presentation of PK descriptive statistics by dose and scheduled time point (i.e., study Week 10 and 18). Geometric means and geometric mean change from baseline, along with the associated log scale SDs, will be presented.

13.5.5 Correlation of A4 Level versus 17-OHP Level

The Exploratory Efficacy Analysis 3.5 in Table 1 is the within subject correlation of A4 Level versus 17-OHP Level. This will be assessed on the log scale by randomized treatment arm. Within a given treatment arm, the data will be organized within subject as follows: (i) dummy variable ‘vtype’= 1 for log A4 and =2 for log 17-OHP; (ii) dummy variable ‘replicate’ = 1 for

baseline, =2 for study Week 10, =3 for study Week 14 and =4 for study Week 18; and (iii) variable 'val' = value of log A4 or log 17-OHP. SAS code will be for the form,

```
PROC MIXED METHOD=ML DATA=X;
CLASS SUBJECT VTYPE REPLICATE;
MODEL VAL = VTYPE / SOLUTION DDFM=KR;
RANDOM VTYPE /TYPE=UN SUBJECT= SUBJECT V VCORR;
REPEATED VTYPE / TYPE=UN SUBJECT=REPLICATE(SUBJECT);
RUN;
```

The estimated correlation coefficient will be extracted from the VCORR matrix estimate and presented by randomized treatment arm.

13.5.6 SF-36 Change from Baseline at Week 18 Analysis

Exploratory Efficacy Analysis 3.6 in Table 1 is the change from baseline in SF-36 for each active arm treatment group (50 mg tildacerfont QD, 100 mg tildacerfont QD, 200 mg tildacerfont QD) compared against the matching placebo group after 12 weeks of randomized, double-blinded treatment. The efficacy of tildacerfont is evaluated by comparing the change from baseline in the tildacerfont treatment group to the placebo treatment group.

13.5.6.1 Statistical Modeling and Testing

If the distribution of residuals is judged non-normal, this analysis will be accompanied by a rank-based ANCOVA model. Statistical testing will be performed using a (rank-based, if non-normal residuals) ANCOVA model. In general, the model will comprise the (rank of, if non-normal residuals) change from baseline value to study Week 18 as the dependent variable; randomized treatment group (0 = placebo, 50 = 50 mg QD tildacerfont, 100 = 100 mg QD tildacerfont, and 200 = 200 mg QD tildacerfont) and sex (0=male, 1=female) as fixed effects; baseline SF-36 total score and baseline GC dose as continuous covariates; the randomization stratum (0 = A4 $\leq$ 4 $\times$ ULN, 1 = A4 > 4 $\times$ ULN) as categorical explanatory variables. The estimation method for the ANCOVA model will REML. A significance level of 0.05 2-sided will be applied.

The ranked value is obtained by ranking all non-missing change from baseline values at the specified analysis visit, if a rank-based model is used.

13.5.6.2 Reporting Results

Summary of Descriptive Statistics

Descriptive statistics (11-point) for absolute and change from baseline in SF-36 will be presented by scheduled time point (i.e., Baseline and Week 18).

Presentation of Analysis Results

The CFB least squares (LS) means will be extracted from the model for each tildacerfont dose and placebo, along with the associated SEs and 95% CIs. Further, the treatment effect for each tildacerfont dose versus placebo, being the corresponding difference in LS means, will also be extracted along with the associated SE, 95% CI and 2-sided p-values.

13.5.7 Change from Baseline Analyses through Week 76

Exploratory Efficacy Analyses 3.7 to 3.11 in Table 1 are the change from baseline at study Week 76 in patient- and clinician-reported outcomes (SF-36 [3.7], IGA acne score [3.8], mFG hirsutism score [3.9]); Exploratory Efficacy Analysis 3.10 is the percent of subjects with at least a 5 mg/day (HCe) reduction from baseline to study Week 76 in GC dose; and Exploratory Efficacy Analysis 3.11 is the change from baseline at study Week 76 in lesion volume of TART(s). For each of these study Week 76 exploratory endpoints, the efficacy of tildacerfont is evaluated by comparing the study Week 76 assessment against baseline.

For Exploratory Efficacy Analyses 3.7 and 3.11, an ANCOVA model will be applied, and results described, in the same manner as that for Exploratory Efficacy Analysis 3.6.

For Exploratory Efficacy Analysis 3.10, logistic regression modelling via SAS PROC GENMOD will be applied in a similar manner as for Exploratory Efficacy Analysis 3.1. The logistic regression model will include CG response (= 0 or 1) as the dependent variable; treatment group (0 = placebo, 50 = 50 mg QD tildacerfont, 100 = 100 mg QD tildacerfont, and 200 = 200 mg QD tildacerfont) and sex (0=male, 1=female as categorical covariates; log baseline A4 and baseline

GC dose (in HCe) as continuous covariates. The model will use a binomial distribution with the logit link function and a significance level of 0.05 2-sided.

The odds ratio for GC response at study Week 76 to each tildacerfont dose level versus placebo group, together with the associated 95% CI and 2-sided p-value will be extracted from the model. SAS code will be of the form,

```
PROC GENMOD DATA=x;
CLASS SUBJECT TREATMENT VISIT;
MODEL RESPONSE = TREATMENT VISIT / DIST=BINOMIAL LINK=LOGIT TYPE3
                    AGGREGATE=TREATMENT;
LSMESTIMATE TREATMENT 'OR 50MG V PCBO' -1 1 0 0 / EXP CL ALPHA=0.05;
LSMESTIMATE TREATMENT 'OR 100MG V PCBO' -1 0 1 0 / EXP CL ALPHA=0.05;
LSMESTIMATE TREATMENT 'OR 200MG V PCBO' -1 0 0 1 / EXP CL ALPHA=0.05;
RUN;
```

Since Exploratory Efficacy Analyses 3.8 and 3.9 relate to ordinal, logistic regression modelling via SAS PROC GENMOD will be applied as described in Section 13.5.1.2.

Since the change from baseline in IGA or mFG may have sparse values may have sparse values (i.e., < 5 distinctive values with counts ≥ 5), the change from baseline values will be further classified in fewer ordered categories (i.e., improvement, no change, worsening) prior to analysis.

13.5.8 Evaluation of ATCH

Exploratory Efficacy Analyses 3.12, 3.13 and 3.14 reflect the analysis of ATCH levels from baseline through 12 weeks of double blind randomized treatment. These Exploratory Efficacy Analyses will be achieved in exactly the same manner as Secondary Efficacy Analyses 2.3, 2.4 and 2.5 described in Sections 13.4.3, 13.4.4 and 13.4.5.

13.5.9 Within Subject Correlation of A4 and 17-OHP with tildacerfont Plasma Concentration

Exploratory Efficacy Analysis 3.15 will examine scatter plots of log A4 and log 17-OHP change from baseline values versus corresponding log tildacerfont plasma concentrations over time by tildacerfont dose group and overall. Further, a repeated measures random coefficients mixed-

effects model will be used to estimate the relationship between the change in log A4 (and log 17-OHP) versus the corresponding log tildacerfont plasma concentrations. The model will be a simplified form of the Primary Efficacy Analysis model,

$$y_{ijt} = (\beta_0 + \xi_i) + (\beta_1 + \eta_i)C_{ijt} + \beta_{4j}I_j + \beta_{5j}I_jC_{ijt} + e_{ijt}$$

where:

y_{ijt} is the log A4 change from baseline value for subject $i, i = 1, \ldots N$, at time t in treatment group j, $j = 1, 2, 3$ ($1 = 200$mg tildacerfont, $2 = 100$mg tildacerfont, $3 = 50$mg tildacerfont); I_t is the indicator for randomised treatment (I_t= 1 for tilacerfont 200mg, =2 for 100mg, =3 for 50mg); C_{ijt} is the corresponding log tildacerfont plasma concentration,

β_0 is the fixed intercept effect

β_1 is the fixed effect of time

β_{4j} is the fixed effect of treatment j parametrized such with $\beta_{43} = 0$

β_{5j} is the effect of treatment j on time parametrized such that $\beta_{53} = 0$

ξ_i is the random effect of subject I intercept

η_i is the random effect of subject i on the linear component of time

e_{ijt} is the random error for subject i on treatment j and time t

The intercept and slope values will be extracted from the model for each tildacerfont. The associated fitted regression line will be displayed for each tildacerfont dose, superimposed on top of the individual subject data over time. The y-axis will reflect log A4 (or 17-OHP) change from baseline and the x-axis will reflect log tildacerfont plasma concentration.

13.5.10 Additional Exploratory Analyses

13.5.10.1 Menstrual Cycle Length

Menstrual cycle length, which is defined in Section 9.3.27.1, will be summarized descriptively by treatment group and scheduled time point. The time between menstrual cycles, which is defined in Section 9.3.27.2, will also be summarized descriptively.

13.5.10.2 Number of Stress Dosing Events

A stress dosing event is derived from the daily GC diary. The number of stress dosing events (Section 9.3.26.4), number and proportions (expressed as percentages) of subjects who underwent stress dosing, along with their exact (Clopper-Pearson) 95% CIs, will be summarized by treatment period (Parts A, B, and C) and scheduled time point.

13.5.10.3 PGIC and CGI-I

Both patients' (PGIC) and clinician's (CGI-I) global impression of improvement scales will be summarized at Weeks 18 and 76. Summaries will include the count and proportion of subjects at each point on the Likert scale as well as an 8-point descriptive summary. At Week 18, a t-test will be used to derive a p-value on the mean score of the placebo and tildacerfont groups.

13.5.10.4 Key Laboratory and Vital data used in Clinical Outcome Components

Key laboratory and vitals parameters that are used to build key clinical outcomes will be summarized by treatment and post-baseline scheduled time point, based on randomized treatment received at each scheduled time point and will include p-values reporting change from baseline using paired-tests and difference from placebo using Wilcoxon rank sum tests.

- Fasting glucose, HbA1c, and fasting insulin
- Total cholesterol, LDL, HDL, triglycerides
- Blood pressure (systolic and diastolic)

14 SAFETY ANALYSES

Safety will be evaluated by AEs, SAEs, AESIs, clinical laboratory, vital signs (excluding the efficacy variables), ECG findings, monitoring for suicide risk and depression/anxiety, as well as physical examination. All analyses of the safety data will be performed using the safety analysis set, based on the actual treatment a subject received. Subjects who received different doses of study drug (placebo or tildacerfont) during the study will be assigned to the treatment group from the time when tildacerfont was received. All descriptive statistics (described in Section 9.1) will be presented by treatment group (including dose level) and treatment period.

14.1 Adverse Events

All AE verbatim terms reported on the eCRFs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA® version 23.0).

All reported AEs (including non-TEAEs) will be listed. Separate listings will be provided for SAEs, AESIs, and TEAEs leading to study drug discontinuation. For all TEAE tables, counting will be by subject and not by event. Additionally, if a subject reported the same TEAE on multiple occasions, the highest severity (severe > moderate > mild) or study drug relationship (related > probable > possible > unlikely > unrelated) recorded for the event will be summarized. All TEAE summary tables will present the number and percentages of subjects reporting TEAEs, unless otherwise specified. A summary of TEAEs by severity, seriousness, and relation to study drug will be tabulated.

AEs experienced during the Run-in Period will be reported separately. For Part A, AEs will be summarized by the actual randomized dose. For Part B, AEs will be summarized under the treatment at the onset of the AE. A subject who does-escalates in Part B will be counted within the denominator of each treatment group that s/he received. For example, a subject who started at 50 mg of tildacerfont, escalated to 100 mg of tildacerfont, and escalated again to 200 mg tildacerfont will be counted in all three dose levels; the AE that the subject experiences will be counted once under the dose level that the subject was taking at the onset of the AE. For Part C, all subjects will receive 200 mg tildacerfont.

An overall summary of AEs will be presented by study period (i.e., Part A, Part B, and Part C). The overall summary will include the number and percentage of subjects experiencing any AEs, study drug related AEs, AEs by maximum severity (highest toxicity grade), SAEs, study drug related SAEs, discontinued due to AE, and deaths.

The number and percentage of subjects experiencing each AE, study drug related AE, and SAEs will be summarized by treatment group according to system organ class and preferred term.

Additional table including only preferred term summarized by treatment for the number and percentage of subjects experiencing each AE will be presented as well

14.1.1 Overall Summary of TEAEs

Overall summary of TEAEs will be presented by study period (i.e., Part A, Part B, and Part C) as well as by treatment (including dose level) group. Subjects will be counted only once at the highest severity when summarizing TEAE by severity.

14.1.2 Summary of TEAEs by System Organ Class and Preferred Term

In addition, TEAEs will be summarized by MedDRA SOC and PT. Each summary table will be presented by study period (i.e., Part A, Part B, and Part C) and treatment (including dose level) group. A subject experiencing multiple occurrences of an adverse event will be counted only once for each PT. Subjects with multiple PTs within a SOC will be counted only once for that SOC. The summaries will be presented for:

- TEAEs by SOC and PT
- TEAEs by Baseline GC dose in HCe (≤ 30 mg/day, > 30 mg/day), SOC, and PT
- Moderate or Severe TEAEs by SOC and PT
- Study drug-related TEAEs by SOC and PT
- Study drug-related TEAEs by GC therapy, SOC, and PT
- AESIs by SOC and PT
- Serious TEAEs by SOC and PT
- Serious study drug-related TEAEs by SOC and PT
- TEAEs leading to study drug discontinuation by SOC and PT
- TEAEs leading to death by SOC and PT

14.2 Clinical Laboratory Evaluation

Laboratory parameters hematology, chemistry, lipid panel, thyroid panel, urinalysis, and other tests) will be summarized by treatment (including dose level) and scheduled time point from randomization, based on actual treatment received at each scheduled time point and will include both observed values and change from baseline values.

In addition, the number and percentage of subjects with potentially clinically significant abnormalities (PCSA, see Table 10) in selected chemistry, hematology, and hormone parameters

(i.e., above the specified upper limit or below the specified limit) will be summarized in shift from baseline tables by scheduled time point.

Table 10 Potentially Clinically Significant Abnormalities by Laboratory Parameter

Parameter	PCSA Low	PCSA High
Chemistry		
ALT		>3 x ULN
AST		>3 x ULN
ALP		>1.5 x ULN
Total Bilirubin		>1.5 x ULN
Creatinine		> 1.8 mg/dL or > 0.4 mg/dL increase from baseline (*> 159.16 µmol/L or > 35.37 µmol/L increase from baseline)*
Sodium	< 130 mEq/L (*< 130 mmol/L)*	>150 mEq/L (*> 150 mmol/L)*
Glucose	< 3 mmol/L *(< 54 mg/dL*)	
Potassium	< 3 mEq/L (*< 3 mmol/L)*	≥5.5 mEq/L (*≥ 5.5 mmol/L)*
Calcium		>11.4 mg/dL (*> 2.84 mmol/L)*
BUN		> 30 mg/dL and > 10 mg/dL increase from baseline (*> 10.71 mmol/L and > 3.57 mmol/L increase from baseline)*
INR		>1.5 x ULN
Total Bile Acids		>5 x ULN or >3 x ULN if ALT is >3 x ULN
FSH		>15 IU/L if baseline was <5 IU/L
Hematology		
Hemoglobin	≤10 g/dL and >2 g/dL reduction from baseline (*≤ 100 g/L and > 20 g/L)*	
Leukocytes	<500/µL (*< 0.5 x 10^9/L)*	
Neutrophils	<1800/µL and >500/µL reduction from baseline	
Platelets	<50,000/µL (*< 50 x 10^9/L)*	

All laboratory results will be listed with reference ranges and range indicator (Low, High). In addition, laboratory results that meet or exceed the pre-specified PCSA levels (Table 10) will be flagged.

14.3 Vital Signs

Descriptive statistics for pulse rate, respiratory rate, temperature, including baseline values and change from baseline values, will be summarized by treatment group and scheduled time point. In addition, the number and percentage of subjects with potentially clinically significant abnormalities (PCSA, see Table 11) in selected vital sign parameters (i.e., above the specified upper limit or below the specified limit) will be summarized in shift from baseline tables by scheduled time point. All vital signs parameters will be listed.

Table 11 Potentially Clinically Significant Abnormalities by Vital Sign Parameter

Vital Sign Parameters	PCSA Low	PCSA High
Pulse Rate (bpm)	< 40	> 110
Systolic Blood Pressure (mmHg)	< 85	> 160
Diastolic Blood Pressure (mmHg)	< 45	> 100

14.4 12-Lead Electrocardiogram

Electrocardiogram (ECG) data, such as clinical interpretation of ECGs, heart rate (HR) values, and interval assessments of QRS duration, QT interval, and the Fridericia's corrected value of the interval between the Q and T waves on the ECG tracing (QTcF) will be listed. Descriptive statistics for observed values and change from baseline at each scheduled time point will be presented for these 12-lead ECG interval and HR assessments.

In addition, the number and percentage of subjects with any abnormal values (i.e. outside a pre-specified threshold) will be summarized by treatment and scheduled time point. The pre-specified levels of ECG QTc thresholds are provided by Spruce (See Table 12 below).

Table 12 Pre-Specified Threshold Levels for ECG Parameters

ECG Parameter		Pre-Specified Level
Heart Rate (bpm)		< 40, > 120, > 130
Heart Rate Change from Baseline (bpm)		> 20, > 30
QRS Interval (msec)		> 120
QTcF (msec)	Normal	Males: ≤ 430, Females: ≤ 450
	Borderline	Male: > 430 to 450, Female > 450 to 470
	Prolonged	Males: > 450, Females: > 470
		Male and Female: > 500
QTcF change from Baseline (msec)		≤ 30, > 30 to 60, > 60

All ECG parameters will be listed with flags for the above pre-specified level. A separate listing of subjects with values of QTcF > 500 msec or an increase > 60 msec will be provided, as necessary.

14.5 Psychiatric Evaluations

14.5.1 The Hospital Anxiety and Depression Scale

The HADS is a self-assessment questionnaire that has been found to be a reliable instrument for detecting states of anxiety and depression in subjects in clinical trials. The total score of HADS, change from baseline in total score will be summarized by time point, treatment group (dose level), and Study Period (A, B/C).

14.5.2 Columbia–Suicide Severity Rating Scale

All responses from the Baseline/Screening Version of the C-SSRS assesses both lifetime history and history from the last 6 months will be listed by subject and by visit. Tables will include results from the Suicidal Ideation and Suicidal Behavior sections of the C-SSRS. Number and percentage of subjects with a response of "Yes" at any point on the Suicidal Ideation and Suicidal Behavior items will be summarized by treatment group and scheduled time point, if sufficient number (i.e., > 5 in any dose level) of subjects responded "Yes".

14.6 Quality of Life

14.6.1 Patient Global Impression of Change

The PGIC (Guy 1976) is a 1-question survey that asks subjects to evaluate whether there has been an improvement in overall subjective health status on a 7-point Likert scale at Week 18 and Week 76. Responses will be listed by subject and visit.

14.6.2 Clinical Global Impression – Improvement Scale

The CGI-I (Guy 1976) is a 1-question survey that requires the clinician to assess how much a patient's illness has improved or worsened relative to a baseline state on a 7-point Likert scale at Week 18 and Week 76. Responses will be listed by subject and visit.

15 CHANGES RELATIVE TO THE PROTOCOL-SPECIFIED ANALYSIS

Protocol Section 4.4.1 defines the completion of study treatment as completing "all visits through Week 76 of the Treatment Period." In the SAP Section 9.3.4, the end of study treatment is redefined to the completion of the Week 76 visit to remove the ambiguity surrounding the classification of subjects with missed or out-of-window visits.

Protocol Section 4.4.2 defines the end of the study as completing "all phases of the study, including the final follow-up visit." In the SAP Section 9.3.5, the end of study treatment is clarified as the completion of the three Treatment Period parts as well as the final Follow-up visit

Protocol Section 8.1.2.2 identifies the Patient Global Impression of Change and Protocol Section 8.1.2.2 identifies the Clinical Global Impression – Improvement Scale as efficacy assessments that evaluate the quality of life. Given that the protocol synopsis and Protocol Section 3 Objectives and Endpoints do not mention PGIC and CGI-I as efficacy assessments, these assessments will be listed as described in SAP Section 14.6.

The mITT Population was defined in Protocol Section 9.3 as all randomized subjects who receive at least 1 dose of study drug (tildacerfont or placebo), have at least 1 A4 assessment, and did not change background GC through Week 18. This definition has been updated to specify

that subjects must have at least one post-baseline A4 assessment in order to make this population relevant to the efficacy analyses.

The Per-Protocol Analysis Set for each part of the Treatment Period was defined in Protocol Section 9.3 as all eligible subjects who have no major protocol deviations that would affect the analysis of efficacy data. This definition has been updated to only pertain to Part A of the Treatment Period in order to align with the analyses of the primary and secondary endpoints.

Since Protocol Section 6.1 defines study drug as either tildacerfont or placebo, the Safety Analysis Set should include all subjects treated with study drug, which includes the single-blinded placebo Run-in Period. The SAP Section 8.4 therefore defines the Safety Analysis Set as subjects who receive at least one dose of study drug during the Run-in or Treatment Period, and the SAP Section 8.5 defines the Randomized Safety Analysis Set to serve the intention of the protocol-defined Safety population defined in Protocol Section 9.3.

Protocol Section 9.4.4.1 specified the analysis of changes IGA acne score, and mFG hirsutism score using a repeated measure mixed effects model if >2 time points are assessed or an ANCOVA for a single post-baseline time point. Given that these scores are ordinal and not continuous variables, the cumulative logit model was deemed more appropriate for the analyses of these two exploratory endpoints.

16 REFERENCE

Benjamini, Y. a. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

FDA. (2007). *Guidance for Industry Developing Products for Weight Management.* FDA.

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. (1996). *Guideline for Industry: Structure and Content of Clinical Study Reports (ICH E3).* Rockville, MD: Federal Register.

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. (1998). *Guideline for Industry: Statistical Principles for Clinical Trials (ICH E9).* Rockville, MD: Federal Register.

Ouyang, J., Carroll, K., Koch, G., & Li, J. (2017, March 7). Coping with missing data in phase III pivotal registration trials: Tolvaptan in subjects with kidney disease, a case study. *Pharmaceutical Statistics*. Retrieved from https://doi.org/10.1002/pst.1808

Tandon, P. a. (2021, April 19). The multi-domain responder index: a novel analysis tool to capture a broader assessment of clinical benefit in heterogeneous complex rare diseases. *Orphanet Journal of Rare Diseases, 16*(183).

Use, I. C. (1996). *Guideline for Industry: Structure and Content of Clniical Stduy Reports (ICH E3).* Rockville, MD: Federal Register.

Use, I. C. (1998). *Guideline for Industry: Statistical Principles for Clnical Trials.* Rockville, MD: Federal Register.

WHO. (2008). *Waist circumference and waist-hip ratio: report of a WHO expert consultation.* WHO.

17 Appendix A: Schedule of Assessments

Table 13 Study SPR001-203 Protocol Schedule of Events at Screening (Protocol Version 7.0, March 14, 2022

	Screening Period	
STUDY VISIT NUMBER	**1a Optional A4 Screening**	**1b Screening**
	In-Clinic Screening Visit	**In-Clinic Screening Visit**
STUDY DAY	**≤45 days before Day 1**	**≤45 days before Day 1 [1]**
Informed consent	X	X
Inclusion/exclusion criteria		X
Demography		X
Medical history		X
Prior medications from past year		X
Concomitant medications		X
Prior and current GC regimens [2]		X
Vital signs [3], body weight		X
Height		X
Full physical examination		X
C-SSRS		X
HADS		X
Hepatitis B & C and HIV tests		X[4]
Urine drug screen		X[4]
Serum pregnancy test for FCP		X[4]

[1] Screening information captured within 45 days of the start of Day 1 in this study (particularly screening information transferred from Spruce Biosciences Study SPR001-204) will be used to determine eligibility and fulfill screening requirements for this study.
[2] Subjects must be on a stable, supraphysiologic dose of GC replacement (defined as ≥ 15 and ≤ 60 mg HCe per day) for ≥ 1 month before the Screening Visit. Information to be collected at screening about a subject's current and historical GC therapy during the past year include the type(s) of GC, the regimen(s), reason(s) that the subject is/was on a particular GC regimen, and any GC stress dosing during the past year.
[3] Vital signs consist of systolic and diastolic blood pressure, pulse rate, body temperature, and respiration rate and will be measured after a 5-min rest period.
[4] These clinic activities of the Screening Visit may be conducted remotely (at the subject's home) if approved by the site's IRB/EC.

	Screening Period	
STUDY VISIT NUMBER	**1a Optional A4 Screening**	**1b Screening**
	In-Clinic Screening Visit	**In-Clinic Screening Visit**
STUDY DAY	**≤45 days before Day 1**	**≤45 days before Day 1 [1]**
Hormones from blood	X[5]	X[4, 6]
Clinical laboratory [7]		X[4]
Urinalysis		X [4]
HbA1c		X [4]
12-lead ECG		X [4]
Report subject's status to site		X [4]

17-OHP = 17 hydroxyprogesterone; A4 = androstenedione; ACTH = adrenocorticotropic hormone; ALP = alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase; C SSRS = Columbia–Suicide Severity Rating Scale; ECG = electrocardiogram; eGFR = estimated glomerular filtration rate; FCP = female of childbearing potential; FSH = follicle-stimulating hormone; GC = glucocorticoid; GGT = gamma-glutamyl transferase; HADS = Hospital Anxiety and Depression Scale; HbA1c = hemoglobin A1c; HCe = hydrocortisone equivalents; HIV = human immunodeficiency virus; INR = international normalized ratio; LH = luteinizing hormone; PT = prothrombin time; PTT = partial thromboplastin time; SHBG = sex hormone–binding globulin

[5] A4 and 17-OHP will be measured. The blood draw for the A4 and 17-OHP sample will be obtained at 8 AM ± 1 hour prior to a morning GC dose.

[6] A4, 17-OHP, ACTH, and testosterone will be measured prior to a morning GC dose.

[7] Clinical laboratory assessments include hematology, clinical chemistry (including liver function tests such as ALT, AST, ALP, GGT, bilirubin, and total bile acids), coagulation (PT/INR and PTT), lipid panel, thyroid panel, LH, FSH, SHBG, renin, aldosterone, inhibin B for males only, and estradiol, prolactin, and progesterone for females only. eGFR for screening will be calculated from blood creatinine measured as part of screening clinical chemistry.

Table 14 Study SPR001-203 Protocol Schedule of Post-Screening Events (Protocol Version 7.0 March 14, 2022)

	Placebo Run-in Period			Treatment Period														Follow-up[8]	ET
				Part A					Part B			Part C							
STUDY VISIT NUMBER	**2**		**3**	**4 Base line**		**5**	**6**	**7**		**8**	**9**	**10**	**11**		**12**	**13**	**14**	**15**	
STUDY WEEK	**0**	**2**	**4**	**6**	**8**	**10**	**14**	**18**	**21**	**24**	**30**	**38**	**46**	**52**	**58**	**70**	**76**	**80**	
STUDY DAY [9]	**1**	**15**	**29**	**43**	**57**	**71**	**99**	**127**	**148**	**169**	**211**	**267**	**323**	**365**	**407**	**491**	**533**	**Last dose +30 days**	
Study Visit (V)/ Telephone Contact (T) [10]	**V**	**T**	**V**	**V**	**T**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**V**
Urine pregnancy test for FCP	X		X	X		X	X	X		X	X	X	X		X	X	X	X	X
Hormones from blood [11,12]	X		X	X		X	X	X		X	X	X	X		X	X	X	X	X

[8] Subjects who do not continue to the optional Open-Label Extension Period upon completion of Treatment Period Part C will enter a 30-day Follow-up Period.

[9] All study visits and telephone contacts should be performed on the indicated study days. In cases where adherence to the foregoing schedule is not possible, all activities for study visits and telephone contacts after the Week 0 visit must be completed within a +6-day window after the indicated study days.

[10] During scheduled telephone contacts, sites will record any AEs and concomitant medications. After receiving A4 results, sites will make telephone contacts to applicable subjects concerning study drug dose escalation and GC reduction (detailed later in this Schedule of Post-Screening Activities). Subjects should be instructed to telephone sites if they have any concerns about their health. PRN telephone contacts initiated by sites and telephone contacts initiated by subjects should be captured in the EDC system as "unscheduled" telephone contacts.

[11] Samples for these lab assessments will be obtained at 8 AM ± 1 hour, after an overnight fast (nothing to eat after midnight), and before any morning dose of GC medication. On the evenings before scheduled laboratory assessments, subjects should take their evening GC dose before 10 PM.

[12] A4, 17-OHP, ACTH, and testosterone will be measured.

	Placebo Run-in Period			Treatment Period														Follow-up[8]	ET
				Part A					Part B			Part C							
STUDY VISIT NUMBER	**2**		**3**	**4 Base line**		**5**	**6**	**7**		**8**	**9**	**10**	**11**		**12**	**13**	**14**	**15**	
STUDY WEEK	**0**	**2**	**4**	**6**	**8**	**10**	**14**	**18**	**21**	**24**	**30**	**38**	**46**	**52**	**58**	**70**	**76**	**80**	
STUDY DAY [9]	**1**	**15**	**29**	**43**	**57**	**71**	**99**	**127**	**148**	**169**	**211**	**267**	**323**	**365**	**407**	**491**	**533**	**Last dose +30 days**	
Study Visit (V)/ Telephone Contact (T) [10]	**V**	**T**	**V**	**V**	**T**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**V**
Clinical laboratory [11,13]				X		X	X	X		X	X	X	X		X	X	X	X	X
Urinalysis [11]				X				X									X		X
HbA1c, fasting glucose and insulin, HOMA-IR [11]				X				X									X	X	X
PK [11,14]				X		X		X					X				X	X	X
Genetic sample [15]				X															
Vital signs [16], body weight	X		X	X		X	X	X		X	X	X	X		X	X	X	X	X
Waist circumference				X				X									X		X

[13] Clinical laboratory assessments include hematology, clinical chemistry (including liver function tests such as ALT, AST, ALP, GGT, bilirubin, and total bile acids), coagulation (PT/INR, and PTT), lipid panel, thyroid panel, LH, FSH, SHBG, renin, aldosterone, inhibin B for males only, and estradiol, prolactin, and progesterone for females only.

[14] A single blood sample will be drawn for PK measurement at each specified study visit.

[15] Where local regulations permit and subject to discretionary approval from each site's IRB/EC and to subject consent, a voluntary blood sample may be collected for DNA analysis. See Protocol Section 13.4 for specifics on genetic testing that may be performed.

[16] Vital signs consist of systolic and diastolic blood pressure, pulse rate, body temperature, and respiration rate and will be measured after a 5-min rest period.

	Placebo Run-in Period			Treatment Period														Follow-up[8]	ET
				Part A					Part B			Part C							
STUDY VISIT NUMBER	**2**		**3**	**4 Baseline**		**5**	**6**	**7**		**8**	**9**	**10**	**11**		**12**	**13**	**14**	**15**	
STUDY WEEK	**0**	**2**	**4**	**6**	**8**	**10**	**14**	**18**	**21**	**24**	**30**	**38**	**46**	**52**	**58**	**70**	**76**	**80**	
STUDY DAY [9]	**1**	**15**	**29**	**43**	**57**	**71**	**99**	**127**	**148**	**169**	**211**	**267**	**323**	**365**	**407**	**491**	**533**	**Last dose +30 days**	
Study Visit (V)/ Telephone Contact (T) [10]	**V**	**T**	**V**	**V**	**T**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**V**
12-lead ECG				X				X			X						X	X	X
HADS				X				X			X				X	X	X		X
SF-36	X			X				X									X		X
PGIC								X									X		X
Scrotal ultrasound for males			X [17]					X [18]									X [18]		X [18]
Inclusion/exclusion criteria [19]				X															
Concomitant medications	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

[17] The initial scrotal ultrasound can be performed anytime during screening up to Week 4 of run-in and will be considered the baseline measurement for TART.

[18] Scrotal ultrasounds will only be conducted in subjects who had TART(s) at baseline.

[19] If a subject fails to meet eligibility criteria during or at the end of the Run-in Period, the investigational site will schedule an abbreviated early termination visit to review AEs and concomitant medications and perform an abbreviated physical examination (preferably in the clinic, but also permissible via home visit).

	Placebo Run-in Period			Treatment Period														Follow-up[8]	ET
				Part A					Part B			Part C							
STUDY VISIT NUMBER	**2**		**3**	**4 Baseline**		**5**	**6**	**7**		**8**	**9**	**10**	**11**		**12**	**13**	**14**	**15**	
STUDY WEEK	**0**	**2**	**4**	**6**	**8**	**10**	**14**	**18**	**21**	**24**	**30**	**38**	**46**	**52**	**58**	**70**	**76**	**80**	
STUDY DAY [9]	**1**	**15**	**29**	**43**	**57**	**71**	**99**	**127**	**148**	**169**	**211**	**267**	**323**	**365**	**407**	**491**	**533**	**Last dose +30 days**	
Study Visit (V)/ Telephone Contact (T) [10]	**V**	**T**	**V**	**V**	**T**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**V**
GC monitoring [20]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
C-SSRS				X				X			X						X		X
CGI-I								X									X		X
Randomization to treatment in Part A[21]/Determination of tildacerfont dose in Parts B and C [22]				X				X		X	X								

[20] Subjects must be on a stable regimen of GC replacement ≥15 and ≤60 mg HCe per day throughout the Run-in Period, Part A, and Part B. Subjects will confirm whether their GC regimen was followed each day in a study diary. On the evenings before scheduled laboratory assessments, subjects should take their evening GC dose before 10 PM. On the mornings of scheduled laboratory assessments, subjects should delay taking any morning dose of GC medication until after the laboratory assessments have been completed.

[21] In Part A, subjects will be randomized to tildacerfont 50 mg QD, tildacerfont 100 mg QD, tildacerfont 200 mg QD, or placebo.

[22] In Parts B and C, all subjects will receive tildacerfont. Subjects who were randomized to placebo in Part A will receive tildacerfont 100 mg QD at the beginning of Part B. Subjects who were randomized to tildacerfont 200 mg QD in Part A will continue to receive tildacerfont 200 mg QD throughout Parts B and C. Based on A4 results from the Weeks 18 and 24 study visits, eligible subjects may undergo a tildacerfont dose escalation after each of these study visits. Specifically, subjects taking tildacerfont 50 mg QD will either continue to receive the same dose (if A4 ≤ULN) or

	Placebo Run-in Period			Treatment Period														Follow-up[8]	ET
				Part A					Part B			Part C							
STUDY VISIT NUMBER	**2**		**3**	**4 Base line**		**5**	**6**	**7**		**8**	**9**	**10**	**11**		**12**	**13**	**14**	**15**	
STUDY WEEK	**0**	**2**	**4**	**6**	**8**	**10**	**14**	**18**	**21**	**24**	**30**	**38**	**46**	**52**	**58**	**70**	**76**	**80**	
STUDY DAY [9]	**1**	**15**	**29**	**43**	**57**	**71**	**99**	**127**	**148**	**169**	**211**	**267**	**323**	**365**	**407**	**491**	**533**	**Last dose +30 days**	
Study Visit (V)/ Telephone Contact (T) [10]	**V**	**T**	**V**	**V**	**T**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**V**
Dispense study drug [23]	X			X		X	X	X [24]		X [24]	X [24]	X [24]	X [24]		X	X			
Study drug accountability			X	X		X	X	X		X	X	X	X		X	X	X		X

be escalated to tildacerfont 100 mg QD (if A4 >ULN and the subject is tolerating his/her current tildacerfont dose); subjects taking tildacerfont 100 mg QD will either continue to receive the same dose (if A4 ≤ULN) or be escalated to tildacerfont 200 mg QD (if A4 >ULN and the subject is tolerating his/her current tildacerfont dose). At Week 30, all subjects will be escalated to 200 mg QD.

[23] Either the subject will pick up study drug in the clinic or the study drug will be shipped directly to the subject. Study drug will be taken daily between 6 PM and midnight, with an evening meal. The evening meal should contain <50% fat content. Study drug may be consumed up to 30 minutes after completing the evening meal, if necessary.

[24] Immediately following each indicated study visit, subjects will continue to take their previously assigned study drug until A4 laboratory results are received, the RTSM system determines the appropriate tildacerfont doses for the next portion of the study, and subjects receive their tildacerfont doses for the next portion of the study. Subjects will receive enough previously assigned study drug to last until they receive their appropriate tildacerfont dose for the next portion of the study. Subjects will either return to the clinic to pick up their study drug or will be shipped their study drug.

	Placebo Run-in Period			Treatment Period														Follow-up[8]	ET
				Part A					Part B			Part C							
STUDY VISIT NUMBER	**2**		**3**	**4 Baseline**		**5**	**6**	**7**		**8**	**9**	**10**	**11**		**12**	**13**	**14**	**15**	
STUDY WEEK	**0**	**2**	**4**	**6**	**8**	**10**	**14**	**18**	**21**	**24**	**30**	**38**	**46**	**52**	**58**	**70**	**76**	**80**	
STUDY DAY [9]	**1**	**15**	**29**	**43**	**57**	**71**	**99**	**127**	**148**	**169**	**211**	**267**	**323**	**365**	**407**	**491**	**533**	**Last dose +30 days**	
Study Visit (V)/ Telephone Contact (T) [10]	**V**	**T**	**V**	**V**	**T**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**V**
Dispense HC for stress dosing and perform accountability [25]	X		X	X		X	X	X		X	X	X	X		X	X	X	X	
PRN telephone contact to follow up after possible tildacerfont dose escalation [26]										X	X								
PRN telephone contacts regarding GC reduction [27]												X	X						

[25] Bottles of HC for stress dosing (see Protocol Section 6.5.1.5) will be dispensed starting at Day 1 and thereafter as needed to replace opened bottles. Accountability for HC will be performed at every study visit.

[26] For subjects who are eligible for blinded tildacerfont dose escalation following the indicated study visits (ie, subjects with A4 >ULN who have not reduced GC), the site will telephone the subject again 7 days after the dose adjustment to assess for the occurrence of AEs and the initiation of new concomitant medications. Subjects who were not eligible for study drug dose escalation will not be telephoned.

[27] In Part C, subjects with A4 ≤ULN at Week 38 or Week 46 will begin to reduce their daily GC dose level by increments of no more than 5 mg/day HCe each time, down to a minimum of 15 mg/day HCe. Sites will contact subjects who are eligible for a GC reduction by telephone within 2 weeks after each applicable study visit. Subjects who are not eligible for a GC reduction will not be telephoned. The site will direct the subject to reduce his/her GC dose at that time if the subject has not experienced any change in clinical status since the previous study visit. If a

	Placebo Run-in Period			Treatment Period														Follow-up[8]	ET
				Part A					Part B			Part C							
STUDY VISIT NUMBER	**2**		**3**	**4 Baseline**		**5**	**6**	**7**		**8**	**9**	**10**	**11**		**12**	**13**	**14**	**15**	
STUDY WEEK	**0**	**2**	**4**	**6**	**8**	**10**	**14**	**18**	**21**	**24**	**30**	**38**	**46**	**52**	**58**	**70**	**76**	**80**	
STUDY DAY [9]	**1**	**15**	**29**	**43**	**57**	**71**	**99**	**127**	**148**	**169**	**211**	**267**	**323**	**365**	**407**	**491**	**533**	**Last dose +30 days**	
Study Visit (V)/ Telephone Contact (T) [10]	**V**	**T**	**V**	**V**	**T**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**T**	**V**	**V**	**V**	**V**	**V**
Review study diary/drug adherence [28]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Review AEs	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Physical examination [29]	X		X	X		X	X	X		X	X	X	X		X	X	X	X	X
Acne IGA score [30]				X				X									X		X
Hirsutism mFG score for females [30]				X				X									X		X

subject's GC dose is adjusted, the site will telephone the subject again 7 days after the dose adjustment to assess for the occurrence of AEs and the initiation of new concomitant medications.

[28] Subjects will use an electronic study diary to document study drug and background GC adherence through Week 18. Female subjects will also use the study diary to record menstrual information (including start and stop dates of menses) across the full treatment period up to Week 76.

[29] A full physical examination will be conducted at Day 1, Week 18, and Week 76. The full physical examination may exclude rectal, genitourinary, and breast exams. An abbreviated physical examination will be conducted at all other visits indicated. If the physical exam is performed at the subject's home, the qualified medical professional will report any changes to the subject's health to the site to determine whether further evaluation is needed via an unscheduled visit.

[30] At Week 6, acne will be evaluated in all subjects and hirsutism in all female subjects. After Week 6, acne/hirsutism will only be evaluated in subjects who had acne/hirsutism at Week 6.

17-OHP = 17-hydroxyprogesterone; A4 = androstenedione; ACTH = adrenocorticotropic hormone; AE = adverse event; ALP = alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase; CGI-I = Clinical Global Impression – Improvement Scale; C-SSRS = Columbia–Suicide Severity Rating Scale; EC = Ethics Committee; ECG = electrocardiogram; EDC = electronic data capture; ET = early termination; FCP = female of childbearing potential; FSH = follicle-stimulating hormone; GC = glucocorticoid; GGT = gamma-glutamyl transferase; HADS = Hospital Anxiety and Depression Scale; HbA1c = hemoglobin A1c; HC(e) = hydrocortisone (equivalents); HIV = human immunodeficiency virus; HOMA-IR = homeostatic model assessment of insulin resistance; IGA = Investigator's Global Assessment (score for acne); INR = international normalized ratio; IRB = Institutional Review Board; LH = luteinizing hormone; mFG = modified Ferriman-Gallwey (score for hirsutism); PGIC = Patient Global Impression of Change; PK = pharmacokinetics; PRN = as needed; PROs = patient-reported outcome measures; PT = prothrombin time; PTT = partial thromboplastin time; RTSM = randomization and trial supply management (system); SF-36 = Short Form 36; SHBG = sex hormone–binding globulin; T = telephone contact; TART = testicular adrenal rest tumor; V = study visit.

Table 15 Study SPR001-203 Protocol Schedule of Post-Screening Events (Protocol Version 7.0 March 14, 2022)

	Extension Treatment Period												Follow-up	ET
STUDY VISIT NUMBER	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	
STUDY WEEK	**76**	**88**	**100**	**124**	**148**	**172**	**196**	**220**	**244**	**268**	**292**	**316**	**320**	
STUDY DAY[31]	**533**	**617**	**701**	**869**	**1037**	**1205**	**1373**	**1541**	**1709**	**1877**	**2046**	**2213**	**Last dose +30 days**	
Study Visit (V)	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**
Urine pregnancy test for FCP	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Hormones from blood[32, 33]	X	X	X	X	X	X	X	X	X	X	X	X		X
Clinical laboratory[32]	X	X	X	X	X	X	X	X	X	X	X	X		X
Vital signs[34], body weight, waist circumference	X	X	X	X	X	X	X	X	X	X	X	X		X
SF-36				X		X		X		X		X		X
C-SSRS				X		X		X		X		X		X
Scrotal ultrasound for males[35]				X		X		X		X		X		X[35]
Concomitant medications	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Review adverse events	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Abbreviated physical examination	X			X		X		X		X		X		X
Dispense study drug[36]	X	X	X	X	X	X	X	X	X	X	X	X		

[31] All study visits should be performed within a +30-day window after the indicated study days.
[32] Samples for these lab assessments will be obtained at 8 AM ± 1 hour, after an overnight fast (nothing to eat after midnight) and before any morning dose of GC medication. On the evenings before scheduled laboratory assessments, subjects should take their evening GC dose before 10 PM.
[33] A4 and 17-OHP and testosterone will be measured.
[34] Vital signs consist of systolic and diastolic blood pressure, pulse rate, body temperature, and respiration rate and will be measured after a 5-min rest period.
[35] Scrotal ultrasounds will only be conducted in subjects who had TART(s) at baseline.
[36] Subject will pick up study drug in the clinic. Study drug will be taken daily between 6 PM and midnight, with an evening meal. The evening meal should contain <50% fat content. Study drug may be consumed up to 30 minutes after completing the evening meal, if necessary.

	Extension Treatment Period												Follow-up	ET
STUDY VISIT NUMBER	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	
STUDY WEEK	**76**	**88**	**100**	**124**	**148**	**172**	**196**	**220**	**244**	**268**	**292**	**316**	**320**	
STUDY DAY[31]	**533**	**617**	**701**	**869**	**1037**	**1205**	**1373**	**1541**	**1709**	**1877**	**2046**	**2213**	**Last dose +30 days**	
Study Visit (V)	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**
Study drug accountability	X	X	X	X	X	X	X	X	X	X	X	X		X
GC dose adjustment telephone contacts[37]	X	X	X	X	X	X	X	X	X	X	X			
Dispense HC for stress dosing and perform accountability[38]	X	X	X	X	X	X	X	X	X	X	X			X

17-OHP = 17 hydroxyprogesterone; A4 = androstenedione; C-SSRS = Columbia–Suicide Severity Rating Scale; ET = early termination; FCP = female of childbearing potential; GC = glucocorticoid; HC(e) = hydrocortisone (equivalents); SF-36 = Short Form 36; V = study visit

[37] During the Treatment Period, sites will make telephone contacts to subjects who are eligible for a GC dose adjustment within 2 weeks after each applicable study visit and based on their A4 level at that visit. Subjects who are not eligible for a GC change will not be telephoned. The site will direct the subject to adjust his/her GC dose by no more than 5 mg/day HCe increments if the subject has not experienced any change in clinical status since the previous study visit. If a subject's GC dose is adjusted, the site will telephone the subject again 7 days after the GC dose adjustment to assess for the occurrence of AEs and the initiation of new concomitant medications. GC dose may be reduced to a minimum of 15 mg HCe per day (approximately physiologic replacement level).

[38] Bottles of HC for stress dosing (see Section Protocol 6.5.1.5) will be dispensed as needed. Accountability for bottles of HC will be performed at every study visit.

18 Appendix B: Clinical Laboratory Parameters

Table 16 Clinical Laboratory Parameters

Chemistry	Hematology	Urinalysis	Thyroid
Glucose, fasting	Platelet count	Specific gravity	T3
Potassium	RBC count	pH	T4
Calcium	MCV	Glucose	TSH
Sodium	MCH	Protein	
BUN	% reticulocytes	Blood	**Other**
Creatinine	Hemoglobin	Ketones	LH
Total protein	Hematocrit	Bilirubin	FSH
ALP	WBC count	Urobilinogen	SHBG
ALT/SGPT	Neutrophils	Nitrite	Renin
AST/SGOT	Lymphocytes	Microscopic examination	Aldosterone
GGT	Monocytes		*Male:* Inhibin B
Total bilirubin	Eosinophils	**Lipid**	*Female:* Estradiol
Direct bilirubin	Basophils	Total cholesterol	*Female:* Prolactin
Total bile acids		LDL	*Female:* Progesterone
PT/INR		HDL	
PTT		Triglycerides	

19 Appendix C: Calculating Total Score for Short-From 36

20 Appendix D: A4 Dose Response Over Time by Randomized Treatment arm

Example SAS codes to execute the dose response random coefficients analysis model.

```
DATA STUDY2;
  INPUT ID EXERTYPE DIET PULSE TIME;
CARDS;
1 1 1 90 0
1 1 1 92 228
1 1 1 93 296
1 1 1 93 639
2 1 1 90 0
2 1 1 92 56
2 1 1 93 434
2 1 1 93 538
3 1 1 97 0
3 1 1 97 150
3 1 1 94 295
3 1 1 94 541
4 1 1 80 0
4 1 1 82 121
4 1 1 83 256
4 1 1 83 575
5 1 1 91 0
```

```
5 1 1 92 161
5 1 1 91 252
5 1 1 91 526
6 1 2 83 0
6 1 2 83 73
6 1 2 84 320
6 1 2 84 570
7 1 2 87 0
7 1 2 88 40
7 1 2 90 325
7 1 2 90 730
8 1 2 92 0
8 1 2 94 205
8 1 2 95 276
8 1 2 95 761
9 1 2 97 0
9 1 2 99 57
9 1 2 96 244
9 1 2 96 695
10 1 2 100 0
10 1 2 97 143
10 1 2 100 296
10 1 2 100 722
11 2 1 86 0
11 2 1 86 83
11 2 1 84 262
11 2 1 84 566
12 2 1 93 0
12 2 1 103 116
12 2 1 104 357
12 2 1 104 479
13 2 1 90 0
13 2 1 92 191
13 2 1 93 280
13 2 1 93 709
14 2 1 95 0
14 2 1 96 112
14 2 1 100 219
14 2 1 100 367
15 2 1 89 0
15 2 1 96 96
15 2 1 95 339
15 2 1 95 639
16 2 2 84 0
16 2 2 86 92
16 2 2 89 351
16 2 2 89 508
17 2 2 103 0
17 2 2 109 196
17 2 2 114 213
17 2 2 120 634
18 2 2 92 0
18 2 2 96 117
18 2 2 101 227
18 2 2 101 614
19 2 2 97 0
19 2 2 98 70
19 2 2 100 295
19 2 2 100 515
20 2 2 102 0
20 2 2 104 165
20 2 2 103 302
20 2 2 103 792
21 3 1 93 0
21 3 1 98 100
21 3 1 110 396
21 3 1 115 498
22 3 1 98 0
22 3 1 104 104
22 3 1 112 310
```

```
22 3 1 117 518
23 3 1 98 0
23 3 1 105 148
23 3 1 118 208
23 3 1 121 677
24 3 1 87 0
24 3 1 122 171
24 3 1 127 320
24 3 1 133 633
25 3 1 94 0
25 3 1 110 57
25 3 1 116 268
25 3 1 119 657
26 3 2 95 0
26 3 2 126 163
26 3 2 143 382
26 3 2 147 501
27 3 2 100 0
27 3 2 126 70
27 3 2 140 347
27 3 2 148 737
28 3 2 103 0
28 3 2 124 61
28 3 2 140 263
28 3 2 143 588
29 3 2 94 0
29 3 2 135 164
29 3 2 130 353
29 3 2 137 560
30 3 2 99 0
30 3 2 111 114
30 3 2 140 362
30 3 2 148 501
;
RUN;

DATA STUDY2;
 SET STUDY2;
 G=EXERTYPE;
RUN;

PROC SORT DATA=STUDY2; BY ID TIME; RUN;

DATA BASE;
 SET STUDY2(RENAME=(PULSE=BASE));
 IF TIME=0;
KEEP ID BASE;
RUN;

PROC SORT DATA = BASE; BY ID; RUN;

DATA STUDY2;
 MERGE STUDY2 BASE;
 BY ID;
 IF TIME >0;
 DUM=1;
RUN;

PROC SORT DATA = STUDY2; BY ID G TIME; RUN;

DATA STUDY2;
 SET STUDY2;
 BY ID G TIME;
  IF FIRST.ID THEN DO;
    T=0;
  END;
  T+DUM;
RUN;

DATA STUDY2;
```

```
 SET STUDY2;
 CFB=PULSE-BASE;
 LPULSE=LOG(PULSE);
 LBASE=LOG(BASE);
 LCFB=LPULSE-LBASE;
RUN;
DATA STUDY2;
 SET STUDY2;
 G=EXERTYPE;
RUN;

PROC SORT DATA=STUDY2;
  BY ID TIME;
RUN;

** FULL MODEL WITH TIME*G AND TIME*TIME*G = DOSE RESPONSE PARAMETERS **;

PROC MIXED DATA=STUDY2 COVTEST NOCLPRINT METHOD=ML  ;
  CLASS ID G ;
  MODEL  PULSE =  BASE TIME G TIME*G TIME*TIME TIME*TIME*G /
  SOLUTION OUTP=PRED1RX OUTPM = PRED1FX;
  RANDOM INTERCEPT TIME TIME*TIME  / SUBJECT = ID SOLUTION;

 ** ESTIMATES WHEN BASELINE = 90 BPM, I.E. LOG BASE = 4.5 **;
 ESTIMATE "INTERCEPT: 1 " INTERCEPT 1 BASE 90 G   1 0 0 ;
 ESTIMATE "INTERCEPT: 2 " INTERCEPT 1 BASE 90 G   0 1 0 ;
 ESTIMATE "INTERCEPT: 3 " INTERCEPT 1 BASE 90 G   0 0 1 ;

 ** TRT EFFECTS AND CONTRASTSAT 600 DAYS **;

ESTIMATE "VALUE AT 600 1"  INTERCEPT 1 BASE 90   G   1 0 0  TIME    600 G*TIME 600 0 0
                           TIME*TIME  360000  G*TIME*TIME  360000 0 0 ;

ESTIMATE "VALUE AT 600 2"  INTERCEPT 1  BASE 90  G   0 1 0  TIME    600 G*TIME 0 600 0
                           TIME*TIME  360000  G*TIME*TIME  0 360000 0 ;

ESTIMATE "VALUE AT 600 3"  INTERCEPT 1 BASE 90   G   0 0 1  TIME    600 G*TIME 0 0 600
                           TIME*TIME  360000  G*TIME*TIME  0 0 0      ;

ESTIMATE "VALUE AT 600 1-3"  G 1 0 -1 TIME    0   G*TIME 600 0 -600 TIME*TIME  0 G*TIME*TIME  360000 0 0;

ESTIMATE "VALUE AT 600 2-3"  G 0 1 -1 TIME    0   G*TIME 0 600 -600 TIME*TIME  0 G*TIME*TIME  0 360000 0 ;

ESTIMATE "LINEAR CONTRAST @ 600"    G   -1 0 1 TIME    0   G*TIME -600 0 600
                                    TIME*TIME  0  G*TIME*TIME  -360000 0 0;

ESTIMATE "QUADRATIC CONTRAST @ 600" G   1 -2 1 TIME    0   G*TIME 600 -1200 600
                                    TIME*TIME  0  G*TIME*TIME  360000 -720000 0;
ODS OUTPUT FITSTATISTICS=FIT1;
ODS OUTPUT SOLUTIONF=F1;
ODS OUTPUT COVPARMS =C1;

TITLE 'RC MODEL TIME G TIME*G TIME*TIME TIME*TIME*G';
RUN;
TITLE;

** REDUCUED MODEL WITH TIME*G AND TIME*TIME*G = DOSE RESPONSE PARAMETERS **;
PROC MIXED DATA=STUDY2 COVTEST NOCLPRINT METHOD=ML  ;
  CLASS ID G ;
  MODEL  PULSE =  BASE TIME G TIME*TIME /
                SOLUTION;
  RANDOM INTERCEPT TIME TIME*TIME / SUBJECT = ID SOLUTION;

 ** ESTIMATES WHEN BASELINE = 90 BPM, I.E. LOG BASE = 4.5 **;
 ESTIMATE "INTERCEPT: 1 " INTERCEPT 1 BASE 90 G   1 0 0 ;
 ESTIMATE "INTERCEPT: 2 " INTERCEPT 1 BASE 90 G   0 1 0 ;
 ESTIMATE "INTERCEPT: 3 " INTERCEPT 1 BASE 90 G   0 0 1 ;

 ** TRT EFFECTS AND CONTRASTSAT 600 DAYS **;
```

```
ESTIMATE "VALUE AT 600 1" INTERCEPT 1 BASE 90   G   1 0 0  TIME  600 TIME*TIME  360000  ;

ESTIMATE "VALUE AT 600 2" INTERCEPT 1  BASE 90  G   0 1 0  TIME  600 TIME*TIME  360000  ;

ESTIMATE "VALUE AT 600 3" INTERCEPT 1 BASE 90   G   0 0 1  TIME  600 TIME*TIME  360000     ;

ESTIMATE "VALUE AT 600 1-3" G   1 0 -1 TIME    0    TIME*TIME  0     ;

ESTIMATE "VALUE AT 600 2-3" G   0 1 -1 TIME    0    TIME*TIME  0      ;

ESTIMATE "LINEAR CONTRAST @ 600"      G   -1 0 1 TIME  0  TIME*TIME  0 ;

ESTIMATE "QUADRATIC CONTRAST @ 600"   G   1 -2 1 TIME  0  TIME*TIME  0   ;

ODS OUTPUT FITSTATISTICS=FIT2;
ODS OUTPUT SOLUTIONF=F2;
ODS OUTPUT COVPARMS =C2;

TITLE 'RC MODEL TIME G TIME*TIME';
RUN;
TITLE;

DATA C1;
 SET C1;
 TP='COVPARM';
RUN;
DATA F1;
 SET F1;
 TP='FIXED  ';
RUN;

DATA DF1;
 LENGTH P $24;
 LENGTH MODEL $12;
 SET F1(RENAME=(EFFECT=P) ) C1(RENAME=(COVPARM=P));
 IF ESTIMATE NE 0;
 DUM=1;
 MODEL='FULL';
KEEP MODEL TP ESTIMATE STDERR P DUM;
RUN;

/*
PROC PRINT DATA = DF1;
RUN;
*/

DATA C2;
 SET C2;
 TP='COVPARM';
RUN;
DATA F2;
 SET F2;
 TP='FIXED  ';
RUN;

DATA DF2;
 LENGTH P $24;
 LENGTH MODEL $12;
 SET F2(RENAME=(EFFECT=P) ) C2(RENAME=(COVPARM=P));
 IF ESTIMATE NE 0;
 DUM=1;
 MODEL='REDUCED';
KEEP MODEL TP ESTIMATE STDERR P DUM;
RUN;

/*
PROC PRINT DATA = DF2;
RUN;
*/
```

```
DATA DF;
 SET DF1 DF2;
RUN;

/*
PROC PRINT DATA = DF;
RUN;
*/

PROC SORT DATA = DF;BY MODEL;RUN;
PROC UNIVARIATE DATA = DF NOPRINT;
BY MODEL;
VAR DUM;
OUTPUT OUT=DF_ SUM=DF;
RUN;

/*
PROC PRINT DATA = DF_;
RUN;
*/

PROC TRANSPOSE DATA = DF_ OUT=DFT(DROP=_NAME_ _LABEL_);
VAR DF;
ID MODEL;
RUN;

DATA DFT;
 SET DFT;
 DUM=1;
RUN;

/*
PROC PRINT DATA = DFT;
RUN;
*/

DATA LOGL1;
 SET FIT1(RENAME=(VALUE=NEG2LOGLFULL));
 IF SUBSTR(DESCR,1,2)='-2';
 DUM=1;
KEEP DUM NEG2LOGLFULL;
RUN;

DATA LOGL2;
 SET FIT2(RENAME=(VALUE=NEG2LOGLREDUCED));
 IF SUBSTR(DESCR,1,2)='-2';
 DUM=1;
KEEP DUM NEG2LOGLREDUCED;
RUN;

PROC SORT DATA = DFT;BY DUM;RUN;
PROC SORT DATA = LOGL1;BY DUM;RUN;
PROC SORT DATA = LOGL2;BY DUM;RUN;

DATA LOGL;
 MERGE LOGL1 LOGL2 DFT;
 BY DUM;
 CHI=NEG2LOGLREDUCED - NEG2LOGLFULL;
 DF=FULL-REDUCED;
 PDOSEREPONSE=PUT(1-CDF('CHISQUARE',CHI,DF),PVALUE6.4);
RUN;

PROC PRINT DATA = LOGL;
TITLE 'PRIMARY ANALYSIS DOSE RESPONSE UNION INTERSECTION P-VALUE';
RUN;
TITLE;
QUIT;
```

21 Appendix E: Calculating Absolute CFB and Absolute Treatment Effects for A4 at Based on Analyses Conducted on the Log Scale.

The A4 Primary Efficacy Analysis is conducted on the log scale due to the inherent skewness and non-Normality of untransformed A4 values. While back transformation of log scale analysis results leads to interpretable quantities on the ratio scale, interest still resides in appreciation of the results on the absolute scale in ng/dL units. Given the non-Normality of A4 values, a simplistic analysis of CFB = (Week 18 A4 value – baseline A4 value) is inappropriate as this simple CFB will also be non-Normal, resulting in upwardly biased estimates of mean A4 level and inflated SDs and SEs.

Thus, to address the Second Secondary Efficacy Analysis (Table 1, Analysis 2.2), the Primary Analysis repeated measures random coefficients mixed-effects model will be re-run exactly as described in Section 13.3.3 but with log A4 as the dependent variable as opposed to the change (i.e. as opposed to log A4 value minus log baseline value). Note, this re-run of the model with log A4 as the dependent variable will give precisely the same results, parameter estimates and variance-covariance matrix estimate as the analysis with log A4 minus log baseline as the dependent variable. The resulting estimates of mean log A4 and mean change from baseline in log A4 will be used via the Delta Theorem to provide consistent estimate of A4 treatment effects on the absolute scale.

Delta Theorem

Suppose x_{ij} are independent, identically distributed Normal random variables with $i = E$ and C representing experimental and control treatments, and $j = 1 \ldots n_i$ representing subject number. Then $x_{ij} \sim N(\mu_i, \sigma^2)$ and $\overline{x}_i \sim N(\mu_i\, \sigma^2/n_i)$.

Suppose there is interest in the function of $\overline{x}_i$, i.e., $u_i = f(\overline{x}_i)$. Then, the delta theorem (= first order Taylors expansion of u_i) as described by Armitage et al (Armitage 2010) states that the variance of u_i is given by

$$Var(u_i) = \left[\frac{\partial f}{\partial \overline{x}_i}\right]^2 Var(\overline{x}_i)$$

And, further,

$$Var(u_E - u_C) = \left[\frac{\partial f}{\partial \overline{x}_E}\right]^2 Var(\overline{x}_E) + \left[\frac{\partial f}{\partial \overline{x}_C}\right]^2 Var(\overline{x}_C) + 2\left[\frac{\partial f}{\partial \overline{x}_E}\right]\left[\frac{\partial f}{\partial \overline{x}_C}\right] Cov(\overline{x}_E, \overline{x}_c)$$

In the case of the Primary Efficacy Analysis, let $\overline{x}_i$ represent the estimated A4 CFB o the log scale, L_i, with $Var(L_i) = [SE(\overline{x}_i)]^2$ and $Cov(L_E, L_C) = Cov(\overline{x}_E, \overline{x}_C)$ where $Cov(.,.)$ denotes covariance.

Interest lies in A4 on the absolute scale, i.e. in the function $f(y) = e^{-y}$, and thus in the function $u_i = e^{-L_i}$. Hence,

$$Var(u_i) = [-e^{-L_i}]^2 Var(L_i) = e^{-2L_i} \cdot Var(L_i)$$

and

$$Var(u_E - u_C) = e^{-2L_E} \cdot Var(L_E) + e^{-2L_C} \cdot Var(L_C) + 2\, Cov(L_E, L_C) \cdot e^{-L_E} \cdot e^{-L_C}$$

Hence, the estimate of the absolute difference in mean A4 values, $E - C$, is given by $u_E - u_C = e^{-L_E} - e^{-L_C}$ with $SE(u_E - u_C) = \sqrt{Var(u_E - u_C)}$ such that the $100(1-\alpha)\%$ CI is given by

$$(u_E - u_C) \pm t_{df,\alpha/2} \cdot \sqrt{Var(u_E - u_C)}$$

where $t_{df,\alpha/2}$ represents the $(1 - \alpha/2)$ precentile of the studentized t cumulative distribution function on df, where df is the error degrees of freedom associated with the corresponding log scale A4 treatment effect estimate at 18 Weeks from the Primary Efficacy Analysis model.

Example SAS Code

The following SAS codes provides an illustrative application of the Delta Theorem.

```
OPTIONS PAGESIZE=58 LINESIZE=180 NONUMBER NODATE;


** DEPENDENT VARIABLE IS LOG PULSE ** ;
PROC MIXED DATA=STUDY2  COVTEST NOCLPRINT METHOD=ML;
  CLASS ID G T;
  MODEL LPULSE = LBASE T G T*G / SOLUTION OUTP=PRED1R OUTPM = PRED1F;
  REPEATED T / TYPE=UN SUBJECT = ID;
  LSMEANS T*G / PDIFF DIFF ALPHA=0.05 CL COV;
  ODS OUTPUT LSMEANS=LSM;
  ODS OUTPUT DIFFS=LSMDIFF;
TITLE 'LOG ABS MMRM';
RUN;
TITLE;

*** TRT EFFECT BASE ON ABSOLUTE SCALE  **;
%MACRO ABSTRTEFF(DRUG,PCBO,T,CVPOS,DAT);
DATA DRUG;
 SET LSM;
 LENGTH LABEL $12.;
 LABEL=LEFT(TRIM(PUT(&DRUG,1.0)))||' V '||LEFT(TRIM(PUT(&PCBO,1.0)));
 IF T=&T AND G=&DRUG;
 CV=COV&CVPOS;
 TIME=&T;
 DRUG=ESTIMATE;
 SEDRUG=STDERR;
 DF=DF;
KEEP LABEL TIME DF DRUG SEDRUG CV;
RUN;
DATA PCBO;
 SET LSM;
 LENGTH LABEL $12.;
 LABEL=LEFT(TRIM(PUT(&DRUG,1.0)))||' V '||LEFT(TRIM(PUT(&PCBO,1.0)));
 IF T=&T AND G=&PCBO;
 TIME=&T;
 PCBO=ESTIMATE;
 SEPCBO=STDERR;
KEEP LABEL TIME PCBO SEPCBO;
RUN;
PROC SORT DATA = DRUG;BY LABEL TIME;RUN;
PROC SORT DATA = PCBO;BY LABEL TIME;RUN;
DATA COMB&DAT;
 MERGE DRUG PCBO;
 BY LABEL TIME;
 ABSDRUG=EXP(DRUG);
 ABSPCBO=EXP(PCBO);
 ABSEFF= ABSDRUG-ABSPCBO;
 ABSDRUGV= SEDRUG**2*ABSDRUG**2;
 ABSPCBOV= SEPCBO**2*ABSPCBO**2;
 V = SEDRUG**2*ABSDRUG**2 + SEPCBO**2*ABSPCBO**2 - 2*CV*ABSDRUG*ABSPCBO;
 ABSLOWER = ABSDRUG-ABSPCBO - SQRT(V)*QUANTILE('T',0.975,DF);
 ABSUPPER = ABSDRUG-ABSPCBO + SQRT(V)*QUANTILE('T',0.975,DF);
 ABSP = 2*(1-CDF('T',ABS(ABSDRUG-ABSPCBO)/SQRT(V),DF));
RUN;
```

```
%MEND ABSTRTEFF;

%ABSTRTEFF(1,3,1,7,1);
%ABSTRTEFF(1,3,2,7,2);
%ABSTRTEFF(1,3,3,7,3);
%ABSTRTEFF(2,3,1,7,4);
%ABSTRTEFF(2,3,2,7,5);
%ABSTRTEFF(2,3,3,7,6);

DATA COMB;
 SET COMB1 COMB2 COMB3 COMB4 COMB5 COMB6;
RUN;

PROC PRINT DATA = COMB;
TITLE 'TREATMENT EFFECT ON ABSOLUTE SCALE';
RUN;
TITLE;

** DEPENDENT VARIABLE IS NOW LOG CHANGE =  LOG PULSE - LOG BASE ** ;
PROC MIXED DATA=STUDY2  COVTEST NOCLPRINT METHOD=ML;
  CLASS ID G T;
  MODEL LCFB = LBASE T G T*G / SOLUTION OUTP=PRED1R OUTPM = PRED1F;
  REPEATED T / TYPE=UN SUBJECT = ID;
  LSMEANS T*G / PDIFF DIFF ALPHA=0.05 CL COV;
  ODS OUTPUT LSMEANS=LSM_;
  ODS OUTPUT DIFFS=LSMDIFF_;
TITLE 'LOG CFB MMRM';
RUN;
TITLE;

DATA LSMDIFF_;
 SET LSMDIFF_;
IF T = _T AND G NE _G ;
IF _G=3;
TIME=T;
LABEL=LEFT(TRIM(PUT(G,1.0)))||' V '||LEFT(TRIM(PUT(_G,1.0)));
KEEP ESTIMATE STDERR LOWER UPPER PROBT TIME LABEL;
RUN;

*** TRT EFFECT ON LOG SCALE **;
%MACRO ABSTRTEFF2(DRUG,PCBO,T,DAT);
DATA DRUG;
 SET LSM_;
 LENGTH LABEL $12.;
 LABEL=LEFT(TRIM(PUT(&DRUG,1.0)))||' V '||LEFT(TRIM(PUT(&PCBO,1.0)));
 IF T=&T AND G=&DRUG;
 TIME=&T;
 DRUG_=ESTIMATE;
KEEP LABEL TIME DRUG_;
RUN;
DATA PCBO;
 SET LSM_;
 LENGTH LABEL $12.;
 LABEL=LEFT(TRIM(PUT(&DRUG,1.0)))||' V '||LEFT(TRIM(PUT(&PCBO,1.0)));
 IF T=&T AND G=&PCBO;
 TIME=&T;
 PCBO_=ESTIMATE;
KEEP LABEL TIME PCBO_;
RUN;
PROC SORT DATA = DRUG;BY LABEL TIME;RUN;
PROC SORT DATA = PCBO;BY LABEL TIME;RUN;
```

```
DATA COMB_&DAT;
 MERGE DRUG PCBO;
 BY LABEL TIME;
RUN;

%MEND ABSTRTEFF2;

%ABSTRTEFF2(1,3,1,1);
%ABSTRTEFF2(1,3,2,2);
%ABSTRTEFF2(1,3,3,3);
%ABSTRTEFF2(2,3,1,4);
%ABSTRTEFF2(2,3,2,5);
%ABSTRTEFF2(2,3,3,6);

DATA COMB_;
 SET COMB_1 COMB_2 COMB_3 COMB_4 COMB_5 COMB_6;
RUN;

PROC PRINT DATA = COMB_;
TITLE 'TREATMENT EFFECT ON LOG SCALE';
RUN;
TITLE;

PROC SORT DATA = COMB;BY LABEL TIME;RUN;
PROC SORT DATA = COMB_;BY LABEL TIME;RUN;
PROC SORT DATA = LSMDIFF_;BY LABEL TIME;RUN;

DATA ALL;
 MERGE COMB COMB_ LSMDIFF_;
 BY LABEL TIME;
 BASED = DRUG - DRUG_;
 BASEP = PCBO - PCBO_;
 BASEC = LEFT(TRIM(PUT(ROUND(EXP(BASED),0.1),5.1)));
 DRUGC = LEFT(TRIM(PUT(ROUND(EXP(DRUG ),0.1),5.1)))||'
['||LEFT(TRIM(PUT(ROUND(SQRT(ABSDRUGV),0.01),6.2)))||']';
 PCBOC = LEFT(TRIM(PUT(ROUND(EXP(PCBO ),0.1),5.1)))||'
['||LEFT(TRIM(PUT(ROUND(SQRT(ABSPCBOV),0.01),6.2)))||']';         ;
 PCTCHGDRUG = LEFT(TRIM(PUT(ROUND(100*(EXP(DRUG_)-1),0.1),6.1)))||'%
['||LEFT(TRIM(PUT(ROUND(SEDRUG,0.0001),6.4)))||']';
 PCTCHGPCBO = LEFT(TRIM(PUT(ROUND(100*(EXP(PCBO_)-1),0.1),6.1)))||'%
['||LEFT(TRIM(PUT(ROUND(SEPCBO,0.0001),6.4)))||']';
 EFFECTPCTCHG = LEFT(TRIM(PUT(ROUND(100*(EXP(ESTIMATE)-1),0.1),6.1)))||'%
('||LEFT(TRIM(PUT(ROUND(100*(EXP(LOWER)-1),0.1),6.1)))
                ||'%, '||LEFT(TRIM(PUT(ROUND(100*(EXP(UPPER)-1),0.1),6.1)))||'%)';
 EFFECTABSCHG = LEFT(TRIM(PUT(ROUND(ABSEFF,0.1),6.1)))||'
('||LEFT(TRIM(PUT(ROUND(ABSLOWER,0.1),6.1)))
                ||', '||LEFT(TRIM(PUT(ROUND(ABSUPPER,0.1),6.1)))||')';
 P_LOGEFFECT = PUT(PROBT,PVALUE7.4);
 P_ABSEFFECT = PUT(ABSP,PVALUE7.4);
RUN;

PROC SORT DATA = ALL;BY LABEL TIME;RUN;
PROC PRINT DATA = ALL SPLIT='*' ;
BY LABEL;
ID LABEL;
LABEL TIME= 'TIME*POINT'
    LABEL= 'TRT*COMP'
    BASEC= 'BASELINE*(BPM)'
    DRUGC= 'DRUG ABS*MEAN [SE]*(BPM)'
    PCBOC= 'PCBO ABS*MEAN [SE]*(BPM)'
    PCTCHGDRUG = 'DRUG PCT CFB*MEAN [SE]'
```

```
    PCTCHGPCBO = 'PCBO PCT CFB*MEAN [SE]'
    EFFECTPCTCHG= 'DRUG-PCBO PCT EFFECT*(95 %CI)'
    EFFECTABSCHG= 'DRUG-PCBO ABS EFFECT*(95 %CI) (BPM)'
    P_LOGEFFECT='DRUG-PCBO P-VALUE' ;
VAR TIME BASEC DRUGC PCBOC PCTCHGDRUG PCTCHGPCBO EFFECTPCTCHG EFFECTABSCHG P_LOGEFFECT ;
RUN;
```