STATISTICAL ANALYSIS PLAN

Study Title:	A Randomized, Placebo-Controlled, Dose Escalation Study to Evaluate the Safety and Efficacy of Tildacerfont in Adult Subjects with Polycystic Ovary Syndrome (PCOS) and Elevated Adrenal Androgens
Study Number:	Study SPR001-210
Investigational Drug:	Tildacerfont (SPR001)
Indication:	Treatment of Polycystic Ovary Syndrome
Investigators:	Multicenter
IND Number:	
Sponsor:	Spruce Biosciences, Inc. 611 Gateway Drive, Suite 740 South San Francisco, CA 94080
Plan Version:	2.0, 17 May 2023
Plan Prepared by:	Inka Leprince, MS Statistical Consultant PharmaStat, LLC

STATISTICAL ANALYSIS PLAN

Study SPR001-210

A Randomized, Placebo-Controlled, Dose Escalation Study to Evaluate the Safety and Efficacy of Tildacerfont in Adult Subjects with Polycystic Ovary Syndrome (PCOS) and Elevated Adrenal Androgens

Plan Version: 2.0, 17 May 2023

Plan version reviewed: 17 May 2023

Signature Page

Approver: DocuSigned by: Mo Noursalehi 19-May-2023

Mo Noursalehi, PhD
SVP, Biometrics
Spruce Biosciences, Inc.

Date

Reviewer: DocuSigned by: Saba Sile 19-May-2023

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

Date

Approver: DocuSigned by: Will Charlton 19-May-2023

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

Date

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Definition of Terms

Abbreviation	Description
11KA4	11-ketoandrostenedione
11KT	11-ketotestosterone
11OHA4	11β-hydroxyandrostenedione
11OHT	11β-hydroxytestosterone
17-OHP	17-hydroxyprogesterone
A4	androstenedione
ACTH	adrenocorticotropic hormone, corticotropin
ADaM	analysis data model
AE	adverse event
AESI	adverse event of special interest
BMI	body mass index
C-SSRS	Columbia–Suicide Severity Rating Scale
CDISC	Clinical Data Interchange Standards Consortium
CI	confidence interval
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
CV	coefficient of variation
DHEA	dehydroepiandrosterone
DHEAS	dehydroepiandrosterone sulfate
ECG	electrocardiogram
eCRF	electronic case report form
FDA	Food and Drug Administration
FSH	follicle-stimulating hormone
GM	geometric mean
GMR	geometric mean ratio
HR	heart rate
ITT	Intent to Treat (Population)
LH	leutenizing hormone
log	logarithmic function
LS	least squares
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified Intent to Treat (Population)
MMRM	mixed-model for repeated measures
PCOS	polycystic ovary syndrome
PK	pharmacokinetic

Abbreviation	Description
PP	per-protocol
PT	preferred term
QD	once daily
QTcF	Fridericia-corrected QT interval
REML	restricted maximum likelihood
RTSM	randomization and trial supply management
SAE	serious adverse event
SAP	statistical analysis plan
SDTM	standard data tabulation model
SHBG	sex hormone binding globulin
SUSAR	suspected unexpected serious adverse reaction
SOC	system organ class
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	20MAR2023	Inka Leprince	First version
2.0	17MAY2023	Inka Leprince	Addition of subgroup analysis for the primary and secondary endpoints based on Inclusion Criteria #3; Added additional analysis of the primary endpoint using an MMRM model; Added supportive analyses of the secondary endpoint

2 RELATED DOCUMENTS: PROTOCOL AND CASE REPORT FORMS

Version	Date
Protocol Version 1.0	21JUN2021
Protocol Version 2.0	03SEPT2021
Protocol Version 3.0	04OCT2021
Protocol Version 4.0	18FEB2022
Protocol Version 5.0	17OCT2022
Protocol Version 6.0	09NOV2022
Protocol Version 6.1	28NOV2022
Case Report Forms	15FEB2022

3 COMMITMENT TO GOOD STATISTICAL PRACTICE

3.1 Definition of Good Statistical Practice

The International Council for Harmonisation (ICH) 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 and 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 topline results and an abbreviated clinical study report (CSR) of Study SPR001-210 for tildacerfont, an investigational drug candidate designed to treat polycystic ovary syndrome (PCOS). This SAP will be used to analyze the safety and efficacy data collected during the study. 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 freeze and treatment un-blinding. 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.

5 STUDY DESIGN

This is a multi-center, randomized, placebo-controlled, dose escalation clinical study that will evaluate the preliminary safety and efficacy of up to 12 weeks of double-blinded treatment with tildacerfont (50 mg, 100 mg, and 200 mg once daily [QD]) in adult women subjects with PCOS who have elevated adrenal androgens at baseline. Up to approximately 40 adult subjects will be randomized into the study.

Subjects who meet all eligibility criteria at Screening will be randomized into the Treatment Period in a 2:1 ratio to receive either tildacerfont or matching placebo. Subjects randomized to the tildacerfont group will receive 3 escalating dose levels of tildacerfont with each dose level administered for 4 weeks. Subjects randomized to placebo will remain on placebo over the full 12-week Treatment Period. Upon completion of the Treatment Period, subjects will return to the clinic within 30 days following their last dose for a Follow-up visit.

Clinical visits during the Treatment Period and Follow-up Period will include safety, efficacy, and PK assessments. Figure 1 depicts study visits from Screening Period to Follow-up Period.

The maximum duration of Study SPR001-210 per subject is expected to be up to 204 days (i.e., up to 30 days of optional DHEAS screening, a maximum of 30-day Screening Period, a 12-week Treatment Period, and a 30-day safety Follow-up Period). Subjects randomized to tildacerfont will receive up to 12 weeks of treatment with tildacerfont. Assessments and procedures for evaluation of safety, efficacy, and PK will be conducted per the protocol-specified schedule (see Appendix A: Table 6 and Table 7).

5.1 Randomization and Blinding

Eligible subjects will be centrally randomized to study intervention prior to entering the Treatment Period using a randomization and trial supply management (RTSM) system. Subjects will be randomized to study treatment in a 2:1 ratio to receive either tildacerfont (starting at 50 mg QD) or matching placebo, stratified by screening dehydroepiandrosterone sulfate (DHEAS: $\leq 1.2 \times$ ULN; $> 1.2 \times$ ULN). At Week 4, subjects receiving tildacerfont will escalate their dose to 100 mg QD. At Week 8, subjects receiving tildacerfont will escalate their dose to 200 mg QD.

Investigators, site personnel, subjects, and the Sponsor's study team will be blinded to study treatment assignment. To maintain the blind, all subjects will receive the same number of study drug tablets, with the appropriate number of tildacerfont and/or placebo tablets for the assigned study intervention.

5.2 Study Treatment

5.2.1 Study Drug Administration

Subjects randomized to the tildacerfont group will initiate treatment with 4 weeks of 50 mg QD, then proceed to 4 weeks at 100 mg QD, and complete the study with 4 weeks at 200 mg QD. All placebo subjects will remain on placebo over the full 12-week Treatment Period.

5.2.2 Assessments

Appendix A: Table 6 and Table 7 show the schedule of events for the study.

5.2.2.1 Efficacy Hormone Assessments

The primary and secondary efficacy assessments are serum DHEAS. Exploratory efficacy assessments include adrenocorticotropic hormone (ACTH), dehydroepiandrosterone (DHEA), testosterone (T), androstenedione (A4), 17-hydroxyprogesterone (17-OHP), 11-ketoandrostenedione (11OHA4), 11β-hydroxytestosterone (11OHT), 11-ketoandrostenedione (11KA4), and 11-ketotestosterone (11KT) will be measured. Efficacy hormone assessments will be collected Day 1, Week 4, Week 8, and Week 12 as shown in Table 7.

5.2.3 Safety Assessments

Safety will be assessed by repeated clinical evaluations including adverse events (AEs), serious AEs (SAEs), AEs leading to discontinuation/withdrawal, dose limiting toxicities (DLTs), AEs of special interest (AESIs), physical examination, vital signs, electrocardiograms (ECGs), clinical laboratory tests (serum chemistry, hematology, coagulation, lipid panel, estimated glomerular filtration rate (eGFR), thyroid panel, cortisol, and urinalysis), and psychiatric evaluations for suicide risk by Columbia–Suicide Severity Rating Scale (C-SSRS).

5.2.3.1 Adverse Events

Investigators will collect information related to AEs 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 treatment (Follow-up Visit).

5.2.3.2 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 Day 1 and Week 12 visits. Abbreviated physical examinations will be performed at Week 4 and Week 8.

5.2.3.3 Vital Signs

The following vital signs will be assessed: blood pressure (systolic and diastolic; mmHg); pulse rate (beats per minute); respiration rate (breaths per minute); body temperature (°C); and weight (kg). Vital signs will be obtained at Screening, Day 1, Week 4, Week 8, and Week 12.

5.2.3.4 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, Day 1, Week 4, Week 8, and Week 12 as shown in Table 7.

Clinically significant adverse findings from the ECG will be recorded on the AE CRF.

5.2.3.5 Clinical Laboratory and Urinalysis

Clinical laboratory assessments include serum chemistry, hematology, coagulation, lipid panel, thyroid panel, cortisol, and urinalysis. eGFR will be calculated from blood creatinine measured as part of screening clinical chemistry. Clinical lab assessments will be performed at Screening, Day 1, Week 4, Week 8, and Week 12.

Clinically significant adverse findings from laboratory or urinalysis assessments will be recorded on the AE CRF.

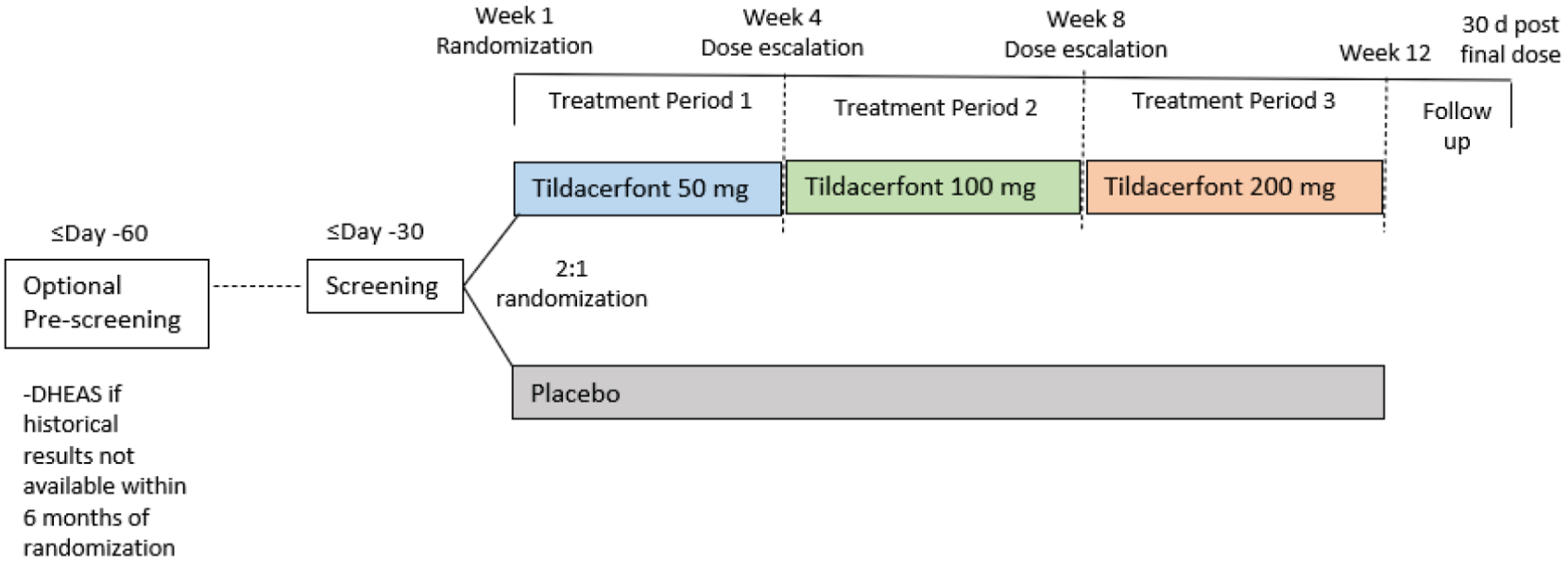
5.2.3.6 Psychiatric Evaluations

The Columbia–Suicide Severity Rating Scale 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 the Screening Visit to determine subject eligibility. The Since Last Visit Version of the C-SSRS will be used at all subsequent visits (Day 1, Week 4, Week 8, and Week 12) specified in Table 7.

5.2.4 Pharmacokinetic Measurements

A plasma samples will be collected for measurement of tildacerfont concentration at each on-treatment clinic visit (Day 1, Week 4, Week 8, Week 12) as indicated in Table 7.

Figure 1 Schema of Study SPR001-210 Study Visits



Abbreviations: d = day; DHEAS = dehydroepiandrosterone sulfate

6 STUDY OBJECTIVES AND ENDPOINTS

This study will evaluate the potential of tildacerfont to reduce and control DHEAS (change from baseline in DHEAS; ≥ 30% (as well as 25% and 20%) reduction from baseline in DHEAS; DHEAS ≤ ULN) in adult subjects with PCOS who have elevated baseline androgens. Exploratory endpoints further explore the effect of tildacerfont on biomarkers (ACTH, DHEA, T, A4, 17-OHP, 11OHA4, 11OHT, 11KA4, and 11KT).

Study SPR001-210 will characterize biomarker outcomes up to 12 weeks of double-blind treatment (see Table 1 for a summary of study objectives).

Table 1 Study 210 Objectives and Endpoints

Objective		Efficacy Variable	Section Number
1.	***Primary Efficacy***		
1.1.	To evaluate the effect of tildacerfont in reducing DHEAS in subjects with PCOS and elevated adrenal androgens	Change from baseline in DHEAS at 12 weeks, 8 weeks, and 4 weeks on treatment	13.1
2.	***Secondary Efficacy***		
2.1.	To evaluate the effect of tildacerfont in achieving target reductions in DHEAS in subjects with PCOS and elevated adrenal androgens	Proportion of subjects who achieve ≥ 30% reduction from baseline in DHEAS at 12 weeks, 8 weeks, and 4 weeks on treatment [1]	13.2
2.2.		Proportion of subjects who achieve DHEAS ≤ ULN at 12 weeks, 8 weeks, and 4 weeks on treatment [2]	
3.	***Exploratory Efficacy***		
3.1.	To evaluate the effect of tildacerfont in reducing key serum hormones and androgens in subjects with PCOS and elevated adrenal androgens	Change from baseline at Week 12, Week 8, and Week 4 in ACTH	13.3
3.2.		Change from baseline at Week 12, Week 8, and Week 4 in DHEA	
3.3.		Change from baseline at Week 12, Week 8, and Week 4 in T	
3.4.		Change from baseline at Week 12, Week 8, and Week 4 in A4	
3.5.		Change from baseline at Week 12, Week 8, and Week 4 in 17-OHP	
3.6.		Change from baseline at Week 12, Week 8, and Week 4 in 11OHA4	

Objective		Efficacy Variable	Section Number
3.7.		Change from baseline at Week 12, Week 8, and Week 4 in 11OHT	
3.8.		Change from baseline at Week 12, Week 8, and Week 4 in 11KA4	
3.9.		Change from baseline at Week 12, Week 8, and Week 4 in 11KT	
4.	***Safety***		
4.1.	To evaluate the safety of tildacerfont in subjects with CAH	AEs (SAE, AEs leading to treatment discontinuation, DLTs, AESIs), physical examination, clinical laboratory assessments (serum chemistry, hematology, coagulation, lipid panel, thyroid panel, cortisol, and urinalysis)	14
[1] Similar endpoints added in the analyses section for 25% and 20% reduction from baseline [2] For subjects who enter the study with DHEAS values higher than ULN			

6.1 Efficacy Objectives and Endpoints

6.1.1 Primary Efficacy

The primary objective of the study is to evaluate the effect of tildacerfont in reducing DHEAS levels in subjects with PCOS and elevated baseline adrenal androgens over the 12 weeks of double-blind, placebo-controlled, dose escalating treatment. The primary efficacy endpoint is the change in DHEAS from baseline (Section 9.3.7) at 12 weeks, 8 weeks, and 4 weeks on treatment.

6.1.2 Secondary Efficacy

The secondary objective evaluates the effect of tildacerfont in achieving adequate control of DHEAS in adult subjects with PCOS and elevated baseline adrenal androgens over the 12 weeks of double-blind, placebo-controlled, dose escalating treatment. The secondary efficacy endpoints are the proportion of subjects with a ≥ 30%, 25% and 20% reduction from baseline at Week 12, Week 8, and Week 4; and the proportion of subjects with DHEAS ≤ ULN (Section 9.3.11) at Week 12, Week 8, and Week 4.

6.1.3 Exploratory Efficacy

The exploratory objectives of the study are to assess the effect of tildacerfont on biomarker control (ACTH, DHEA, T, A4, 17-OHP, 11OHA4, 11OHT, 11KA4, and 11KT) over the 12 weeks of double-blind, placebo-controlled, dose escalating treatment.

6.2 Safety Objective

The safety objective of the study is to evaluate the safety of tildacerfont in subjects with PCOS and elevated baseline adrenal androgens using safety assessments described in Section 5.2.3.

7 SAMPLE SIZE AND POWER

A sample size of n = 33 subjects with 2:1 randomization will provide at least 80% power to detect a difference of 164 µg/dL in mean DHEAS between at least one dose of tildacerfont and placebo, assuming a standard deviation of 154 µg/dL and a two-sided alpha of 0.05. Accounting for a 15% drop out rate, the projected final sample size is N = 39 (n = 26 for tildacerfont, n = 13 for placebo).

8 ANALYSIS SETS

8.1 Intent-To-Treat Analysis Set

The Intent-To-Treat (ITT) Analysis Set will include all randomized subjects. The ITT Analysis Set will be the basis for demographics and baseline characteristics.

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 a baseline DHEAS assessment, and have at least one post-baseline DHEAS assessment. The mITT Analysis Set will be the basis for efficacy analyses.

8.3 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. The safety analysis set, which will be summarized based on the actual randomized treatment administered, will be used for evaluation of study drug exposure, concomitant medication, and safety.

9 GENERAL CONSIDERATIONS

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

9.1 Presentation of Summary Statistics

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

Unless otherwise specified, continuous efficacy hormones (Section 5.2.2.1) 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 categorical/qualitative data will be presented using the frequency of events and percentages. All percentages will be presented to 1 decimal place, unless otherwise specified. Percentages equal to 100 will be presented as 100% and percentages will not be presented for zero frequencies, with the exception of binary variables. For summaries of AEs and concomitant medications (CMs), the percentages will be based on the number of subjects who received study drug.

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

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 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 are randomized to receive study drug treatment are considered enrolled subjects.

9.3.2 Study Day

Study Day, which follows the CDISC SDTM standard, is defined as (Assessment 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 End of Study Treatment Definition

A subject is considered to have completed study treatment if the subject has completed the Week 12 Visit.

9.3.4 End of Study Definition

A subject is considered to have completed the study if the subject has completed the 30 day Follow-up Visit.

9.3.5 Age

Age (years) will be calculated as the number of years between year of birth and year of informed consent, rounded down to the largest to a whole number less than or equal to the age at measurement.

9.3.6 Body Mass Index

Body mass index (BMI, kg/m^2) is derived as weight (kg) / [height (m) × height (m)].

9.3.7 Baseline Values

Baseline values are defined as the last non-missing assessment prior to the first dose of randomized, double-blind study drug in the study.

9.3.8 Change from Baseline

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

9.3.9 Percent Change from Baseline

Percent change from baseline is calculated as the change from baseline divided by the baseline assessment and multiplied by 100%. This value will be rounded to two decimal places unless otherwise stated.

9.3.10 Log-transformed Change from Baseline

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

$$\begin{aligned} &log\text{-}tranformed\ change\ from\ baseline \\ &\quad = \log(post\text{-}\ baseline\ assessment) - \log(baseline) \\ &\quad = \log\left(\frac{post\text{-}\ baseline\ assessment}{baseline\ assessment}\right) \end{aligned}$$

9.3.11 Upper Limit of Normal or Target by Efficacy Biomarker

The key biomarker upper limit of normal (ULN) and Target values that are displayed in Table 2 will be used in data analysis. The age at screening will be used to determine those age-specific ULN values.

Table 2 Upper Limit of Normal Values by Subgroup and Key Biomarker

Baseline Characteristic	DHEAS ULN	ACTH ULN	17-OHP Target	A4 ULN
≥ 18 - ≤ 30 years	377 µg/dL	63.0 ng/dL	200 ng/dL	200 ng/dL
≥ 31 - ≤ 40 years	295 µg/dL	63.0 ng/dL	200 ng/dL	200 ng/dL
≥ 41 - ≤ 50 years	240 µg/dL	63.0 ng/dL	200 ng/dL	200 ng/dL

9.3.12 Responder Definitions

The second secondary efficacy and exploratory variables are proportions of responders based on serum DHEAS assessments. Indicator variables will be created for the responder criteria as follows:

Table 3 Responder Definition by Endpoint

Endpoint	Definition
Proportion of subjects with at least a 30% reduction from baseline in DHEAS dose [1]	$\begin{cases} 1 & \geq 30\%\ reduction\ in\ DHEAS \\ 0 & < 30\%\ reduction\ in\ DHEAS\ OR\ missing \end{cases}$
Proportion of subjects with DHEAS ≤ ULN	$\begin{cases} 1 & DHEAS \leq ULN \\ 0 & DHEAS > ULN\ OR\ missing \end{cases}$
[1] repeat same definition for 25% and 20% reduction	

9.3.13 Geometric Mean Ratio and Percent Change

For efficacy hormone variables (Section 5.2.2.1), 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).

9.3.14 Study Drug Exposure Variables

9.3.14.1 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*.

9.3.14.2 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 is 0 mg if placebo or 50 mg, 100 mg, 200 mg tildacerfont if active drug.

9.3.14.3 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.

The date of the first dose is captured by the eCRF question, "Date of first dose?" The first dose date at the 100 mg tildacerfont QD and 200 mg tildacerfont QD is presumed to be the same as the dispensed date at the Week 4 and Week 8 visits, respectively.

The last dose date can be the last dose date of a treatment dose level or treatment end date.

9.3.14.4 Mean Daily Tildacerfont Dose (mg/day)

Mean Daily tildacerfont Dose (mg/day) is calculated as Total tildacerfont Dose (mg) divided by Treatment Duration (days) for each treatment period and over the entire study.

9.3.15 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 that is continued after the first study drug dosing (i.e., a prior medication can also be a CM if it continued after the first dose of study drug) or 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 study drug dosing, or with start dates or stop dates within the first dose date through last dose date + 24 hours, missing CM end date, or ongoing with the study drug administration.

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

9.3.15.1 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 5.6.1 and 12.3.

9.3.16 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 study drug until 30 days after the final dose of study drug (Follow-up visit). For AEs that occur on the date of the first dose of study drug, the time of onset (before or after intake of study drug) will be specified and used for comparison.

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 do not meet SAE criteria but must be monitored on an ongoing basis as defined in Protocol Section 7.2.8 and identified in the AE eCRF page.

Serious AEs (SAEs) are defined in Protocol Section 7.2.6 and identified in the AE eCRF page.

Dose-limiting toxicities are defined as a Common Terminology Criteria for Adverse Events (CTCAE) Grade 3 or higher TEAE that is also considered at least possibly related to study drug.

Suspected unexpected serious adverse reactions (SUSARs) is an AE that is serious, related to study drug, and is not expected in the targeted disease (PCOS).

9.3.17 Subject-Level Study Treatment

This study consists of Screening, Treatment, and Follow-up periods. For a data record, the planned treatment and dose level will be based on a specific study period as follows:

Study Period	Planned Treatment
Screening	Placebo or tildacerfont (50 mg QD)
Treatment Period (Day 1 to Week 4)	Placebo or tildacerfont (50 mg QD)
Treatment Period (post-Week 4 to Week 8)	Placebo or tildacerfont (100 mg QD)
Treatment Period (post-Week 8 to Week 12)	Placebo or tildacerfont (200 mg QD)

The actual study drug treatment and dose level will be the study drug and dose level a subject was on when a data record was measured/assessed.

9.3.17.1 Record-Level Actual Treatment and Dose for an Adverse Event

TEAEs are those AEs with onset date from the date of first dose of randomized study drug until the end of the study (Follow-up). The treatment will be the randomized study treatment (tildacerfont or placebo) and the dose level (50 mg QD, 100 mg QD, 200 mg QD) will be the actual tildacerfont dose received on the onset date of the TEAE records.

Post-treatment AEs (within 30 days after study drug discontinuation) will be assigned the last treatment and dose level that the subject was on before study drug discontinuation. Post treatment AEs after 30 days post study drug discontinuation will be assigned to no treatment.

9.3.17.2 Actual Treatment and Dose Level for a Medication Record

Since a medication could be a prior medication and be taken during Treatment Periods, flag variables will be used to identify a medication record that was taken while a subject received placebo, tildacerfont 50 mg, tildacerfont 100 mg, and/or tildacerfont 200 mg.

A medication record will be assigned to the first study treatment and dose level when the subject was on and was taking the medication during the Treatment Period. For example, if a subject starts a new medication while taking tildacerfont 100 mg then later dose escalated to 200 mg QD at Week 8, the treatment and dose level will be tildacerfont 100 mg for that medication record.

9.4 Analysis Windows

Analysis visits will match the protocol-specified visits (Screening, Day 1, Week 4, Week 8, Week 12, and Follow-up). 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 4.

Table 4 Analysis Visit Windows (Clinic Visits Only)

Protocol Specified Visit Number	eCRF Visit Label	Analysis Visit	Target Study Day	Start (days)	Stop (days)
	Screening	Screening	-15	low	-7
Visit 1	Treatment P1 V1	Day 1	1	-6	14
Visit 2	Treatment P1 V2	Week 4	29	15	43
Visit 3	Treatment P2 V3	Week 8	57	44	71
Visit 4	Treatment P3 V4	Week 12	85	72	100
Visit 15	Follow up	Follow-up	Last dose	Last dose +1	Last dose +30

10 STATISTICAL AND ANALYSIS ISSUES

10.1 Adjustments for Covariates

The categorical variables treatment group (0 = placebo, 1 = tildacerfont) will be used in statistical models for all efficacy analyses. The baseline value of continuous efficacy endpoints will be included in statistical models as a continuous covariate. Age will only be included as a continuous covariate in DHEAS analyses because DHEAS levels are known to decline with age. In general, the categorical, randomization stratification variable (i.e., Screening DHEAS value [0 = DHEAS $\leq$ 1.2 $\times$ ULN; 1 = DHEAS > 1.2 $\times$ ULN]) will be used in efficacy analyses, except for analyses where the dependent variable is derived from DHEAS. For the analysis of change from baseline in DHEAS, the model will include the baseline serum DHEAS, as a continuous covariate in place of the randomization stratification variable.

10.2 Evaluating Normality

A probability plot (Quartile-Quartile [QQ] plot) of the residuals of each analysis model will be evaluated. A linear QQ plot indicates that the residuals follow a normal distribution, while any pattern in the QQ plot indicates deviations from normality. Homoscedasticity will be assessed with a scatter plot of residuals by predicted values. A uniform scatter pattern is indicative of homoscedasticity, while any pattern indicates heteroscedasticity.

10.3 Handling Dropouts or Missing Data

Every effort will be made by the Sponsor to ensure completeness of data collection. Missing data will not be imputed unless otherwise specified.

10.3.1 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 (X), 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

10.3.2 Handling of Safety Data

10.3.2.1 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 start dates will be considered TEAEs, if:

- Day and month are missing and the year is equal to or after the year of the first date of study drug dosing;
- Month is missing and the year is after the year of the first date of study drug dosing;
- Day is missing and the year is equal to the year of the first date of study drug dosing and the month is equal to or after the month of the first date of study drug dosing; or
- Year is missing.

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 ensure no missing severity or relationship of an AE to study drug prior to database lock finalization.

10.3.2.2 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 Jan. 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.

10.4 Primary Analyses

The primary analyses will be conducted when all subjects complete Week 12 of study treatment and the study and the data has been frozen for analysis.

10.5 Multicenter Considerations

This is a multicenter trial in the United States. 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.

10.6 Multiple Comparisons, Multiplicity

This Phase II study is exploratory in nature and no adjustment for multiplicity is planned. The existence of multiplicity is acknowledged and interpretation of the primary and secondary analysis results will be interpreted with caution.

10.7 Active-Control Studies

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

10.8 Examination of Subgroups

Subjects will be categorized into the following common subgroups for the purposes of evaluating the primary and secondary efficacy endpoints. Due to the small sample size, subgroups will be summarized descriptively in place of subgroup analyses.

- Screening DHEAS (≤ 1.2 × ULN; > 1.2 × ULN)
- Inclusion Criteria #3: "DHEAS level > age-matched reference ULN at the screening visit." (Subjects enrolled into the study fully compliant with inclusion/exclusion criteria; subjects who marginally do not meet inclusion criteria #3)

11 STUDY SUBJECTS

11.1 Subject Enrollment and Disposition

11.1.1 Screened Subjects

Screening and disposition will be summarized by investigator as well as by treatment group (if randomized) and overall.

The percentages will be based on total subjects who screened for the study. A subject listing will be provided with the above information, as well as site ID, date of informed consent, and reason for screen failure for subjects who were screened but not randomized.

11.1.2 Randomized Subjects

Enrollment and disposition of randomized subjects will be summarized by investigator as well as by treatment group and overall.

A listing of disposition will be provided for all randomized subjects. A separated listing will present subjects who are excluded from ITT, mITT, or SAF 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 freeze 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 (ICH E3) guidance. Major analysis protocol deviations include, but not limited to:

- Randomized subjects who did not satisfy efficacy inclusion and exclusion criteria,
- Randomization error,
- Subjects who received the wrong tildacerfont dose,
- Subjects who received prohibited concomitant medication(s),
- Subjects who are less than 80% compliant with aggregate study treatment

All protocol deviations will be summarized by deviation category and type. In addition, a listing of all deviations and major protocol deviations will be provided. Prior to locking the database, all protocol deviations will be reviewed and a remedy plan will be added as a note to file to adjust the efficacy analyses, as necessary.

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 (18 - 35 years, 36 - 40 years, and 41+ years), child-bearing potential, race, and ethnicity. Demographic characteristics will be summarized by randomized treatment group and overall. A subject listing of demographics will be provided.

11.3.2 Screening and Baseline Characteristics

The following screening characteristics include screening serum DHEAS and screening DHEAS category (≤ 1.2 × ULN; > 1.2 × ULN). The following baseline characteristics include baseline weight, height, BMI, and hormones (DHEAS, A4, ACTH, 17-OHP, 17-OHP category [< 200ng/dL; ≥ 200 ng/dL], and cortisol). Screening and baseline characteristics will be

summarized by randomized treatment group and overall. A subject listing of screening and baseline demographics will be provided.

11.4 Medical History

Medical history will be summarized by randomized treatment group and overall. Individual subject listings will include subject identification number and randomized treatment for all randomized subjects (ITT).

General medical history, be mapped to preferred terms and system organ classes using the Medical Dictionary for Regulatory Activities (MedDRA®) Dictionary (version 24.1). 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.

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 each dose level of the Treatment Period. 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 9.3.13.1) by treatment group. A subject who received multiple dose levels of tildacerfont will be included in each of the dose levels for the purpose of summarization.

The drug accountability information, including number of tablets dispensed, number of tablets returned, and number of tablets lost or destroyed, along with all summarized variables, will be listed in subject listings.

12.2 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 D Global-

B3 Sep2021). The medications will be mapped to Anatomical/Therapeutic/Chemical (ATC) class and preferred names.

Prior and concomitant medications (see definition in Section 9.3.14 will be summarized using WHO DD ATC class and preferred name. The summary results will be presented by treatment/dose level and overall. Prior medications will be summarized by randomized treatment group and overall. Concomitant medications will be summarized by treatment group and overall.

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 are 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

The efficacy analyses are planned after all subjects have completed the 12 weeks of randomized, double-blinded study treatment and the database is frozen. During the 12 weeks of double-blind, placebo-controlled treatment period, the tildacerfont group will escalate doses at Week 4 and Week 8 to 100 mg QD and 200 mg QD, respectively. The scheduled visits included in this treatment period include Day 1 (baseline), Week 4, Week 8, and Week 12. The primary efficacy analysis will compare individual tildacerfont dose groups to the placebo group with respect to the change from baseline in serum DHEAS.

These efficacy analyses will be performed using the mITT Analysis Set (Section 8.2) as the main analyses. The endpoints identified in Table 1 as the primary efficacy endpoint (1.1), secondary efficacy endpoint (2.1), and exploratory efficacy endpoints (3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9) will be analyzed.

13.1 Primary Efficacy Analysis

The primary objective of the study is to evaluate the effect of tildacerfont on DHEAS levels in female adult subjects with PCOS who have elevated adrenal androgens at baseline over the 12-week, double-blind, placebo-controlled, dose-escalated Treatment Period.

13.1.1 Definition of Primary Estimand

The primary "treatment policy" estimand is the log-transformed change from baseline in serum DHEAS for the active arm (dose escalation of tildacerfont from 50 mg QD to 100 mg QD to 200 mg QD) compared against the matching placebo group after 12 weeks of double-blinded, placebo-controlled, dose-escalating treatment. The placebo treatment group will control for any potential placebo effects in PCOS subjects and their impact on the overall tildacerfont treatment effect. The superiority of tildacerfont is evaluated by comparing the change from baseline in the tildacerfont treatment group against the change from baseline in the placebo treatment group.

Population: The analysis population is the mITT analysis set, all PCOS subjects who were randomized into the study, received at least 1 dose of study drug (tildacerfont or placebo), have at baseline DHEAS assessment, and have at least one post-baseline DHEAS assessment.

Variable: The analysis variable is the log-transformed change in the serum DHEAS value from baseline at Week 12 for each treatment group and dose level. The change is calculated as the log-transformed Week 12 assessment subtracted by the log-transformed baseline value.

Intercurrent events: For the primary analysis, intercurrent events such as treatment interruption, treatment discontinuation, and use of prohibited medications will not be taken into consideration for analysis as we assume there are no clinically relevant anticipated intercurrent events in PCOS.

13.1.2 Statistical Hypotheses

For the primary efficacy endpoint, the null hypothesis for the treatment comparison is that there is no difference in change from baseline at Week 12 in DHEAS between tildacerfont and placebo

after the 12-week, double-blinded, placebo-controlled, dose-escalating Treatment Period. The alternative hypothesis is that there is a difference between a tildacerfont and matching placebo.

$$H_0: \mu_{tildacerfont} - \mu_{placebo} = 0;$$

$$H_A: \mu_{tildacerfont} - \mu_{placebo} \neq 0$$

Where $\mu_{tildacerfont}$ and $\mu_{placebo}$ are the tildacerfont and placebo group means of log-transformed change in the serum DHEAS.

13.1.3 Statistical Modeling and Testing

13.1.3.1 Analysis of Covariance

The normality of residuals will be evaluated (Section 10.2). If the distribution of residuals is judged non-normal, then the analysis will be repeated using rank values in place of the original values. Statistical testing for the exploratory efficacy endpoint will be performed using a (rank-based, if non-normal residuals) ANCOVA model. The model will comprise the (rank of, if non-normal residuals) log-transformed change from baseline value as the dependent variable; randomized treatment group (0 = placebo, 1 = tildacerfont) as a fixed effect; and the log-transformed baseline serum DHEAS and age as a continuous covariates. The estimation method for the ANCOVA model will be maximum likelihood.

The ranked value is obtained by ranking all non-missing ratio of post-baseline/baseline values, if a rank-based model is used. The formal testing for the group difference ($\neq 0$) between tildacerfont and placebo treatment groups in change from baseline will be conducted at the scheduled time point of Week 12 visit using the aforementioned ANCOVA model. Statistically significant difference will be declared if two-sided p-value < 0.05.

The statistical model used to characterize the primary endpoint at Week 12 will also be used to characterize the tildacerfont response at earlier weeks, replacing the Week 12 outcome with the Week 8 and the Week 4 outcomes.

13.1.3.2 Mixed Model for Repeated Measures

Additionally, a mixed model repeated measures (MMRM) analysis will be conducted to further assess the primary endpoint. The MMRM will include the log-transformed change from baseline

value as the dependent variable with subject as random effect; time, randomized treatment group (0 = placebo, 1 = tildacerfont) and its interaction with time as fixed effects; as well as the log-transformed baseline serum DHEAS and age as a 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. The p-value for testing the treatment effect from the MMRM model will be reported.

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.

1. Toeplitz covariance structure (assuming measurements from samples taken closer together in time are more highly correlated than those from samples taken farther apart)
2. 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)
3. Compound symmetry covariance structure (assuming equal correlation for measurements from a subject, regardless of how far apart in time when they were taken)

Subjects with missing data DHEAS data will not be imputed for the primary analyses.

13.1.4 Reporting Results

13.1.4.1 Summary of Descriptive Statistics

Descriptive statistics (8-point) for the serum DHEAS assessments, change from baseline in serum DHEAS, and % change from baseline in serum DHEAS will be presented by scheduled time point (i.e., Baseline [serum DHEAS only], Week 4, Week 8, and Week 12). For the baseline assessment only, the geometric mean and CV% will also be summarized.

If the normality assumption does not hold, the descriptive statistics of the change from baseline will be reported by treatment group and will include the median, the 95% distribution-free CI of the median, Q1, Q3, min and max.

13.1.4.2 Summary of Analysis Results

The p-value for testing the treatment effect from the (rank-based, if non-normal residuals) ANCOVA model will be reported. In addition, if the normality assumption holds, the back-transformed least squares (LS) GMR of change from baseline (post-baseline/baseline), two-sided 95% CI of the LS GMR, the back-transformed LS mean ratio difference between treatment group (i.e., tildacerfont/placebo), and 95% CI of the LS mean ratio treatment difference will be reported from the ANCOVA model. If the normality assumption does not hold, the difference of medians and its 95% CI will be reported. These values will be obtained from bootstrap method as the 50th percentile, the 2.5th percentile, and the 97.5th percentile, respectively, of the difference in group medians from 5000 bootstrap random samples drawn with replacement.

The p-value for testing the treatment effect from the (rank-based, if non-normal residuals) MMRM model will be reported. In addition, if the normality assumption holds, the back-transformed least squares (LS) GMR of change from baseline (post-baseline/baseline), two-sided 95% CI of the LS GMR, the back-transformed LS mean ratio difference between treatment group (i.e., tildacerfont/placebo), and 95% CI of the LS mean ratio treatment difference will be reported from the MMRM model.

13.1.4.3 Graph Presentation

A serial change from baseline in serum DHEAS plot will be presented using the mean change from baseline in a serum DHEAS. The y-axis will be the change from baseline level and the x-axis will represent the treatment visits: Baseline, Week 4, Week 8, and Week 12. All treatment groups will be presented on the graph. In addition, the LS GMR and the 95% CI from the MMRM model will be plotted.

For each treatment group, a spaghetti plot of all subjects' DHEAS levels over time (Baseline, Week 4, Week 8, and Week 12) will be plotted with the means at each time point overlaid. These spaghetti plots will be replicated for DHEAS change from baseline.

13.1.5 Subgroups

The above ANCOVA and MMRM analyses will be repeated for two separate subgroups: the subgroup of subjects who fully meet the inclusion criteria #3, and the subgroup of subjects that do not meet this criterion. Furthermore, the primary endpoint analyses will be descriptively summarized for the other subgroup(s) specified in Section 10.8.

13.2 Secondary Efficacy Analysis

The secondary "treatment policy" estimand is the proportion of PCOS subjects with elevated androgens at baseline who were randomized into the study and achieved adequate control of serum DHEAS levels through the 12 weeks of double-blinded, placebo-controlled, dose-escalating Treatment Period. The active arm (dose escalation of tildacerfont from 50 mg QD to 100 mg QD to 200 mg QD) will be compared against the matching placebo group. The placebo treatment group will control for any potential placebo effects in PCOS subjects and their impact on the overall tildacerfont treatment effect. The superiority of tildacerfont is evaluated by comparing the proportion of tildacerfont subjects meeting this criterion is greater than the proportion of placebo subjects meeting this criterion.

Adequate control of serum DHEAS will be evaluated with two measurements: the proportion of subjects with a greater than 30% reduction from baseline and the proportion of subjects with DHEAS ≤ ULN.

Population: The analysis population for the secondary efficacy analyses is the mITT analysis set, including all randomized PCOS subjects who received at least 1 dose of study drug (tildacerfont or placebo), have at baseline DHEAS assessment, and have at least one post-baseline DHEAS assessment.

Variable: The analysis variable is the proportion of subjects in the mITT analysis set who are responders (defined as [a] ≥ 30% reduction in serum DHEAS or [b] DHEAS ≤ ULN). This proportion is calculated for each treatment group (e.g., placebo, tildacerfont) as the number of subjects at each time point (Week 12, Week 8, Week 4) who have adequate control of serum hormones (greater than or equal to 30% reduction from baseline [2.1] or DHEAS ≤ ULN [2.2]),

divided by the total number of subjects in the mITT analysis who were randomized to that specific treatment.

Subjects with missing data are assumed to be non-responders. Subjects who do not have a 30% reduction from baseline (Objective 2.1) or subjects who have DHEAS > ULN (Objective 2.2) are considered non-responders.

The same analysis will be repeated for subjects with 25% and 20% reduction in DHEAS from baseline.

Intercurrent events: For the secondary analyses, intercurrent events such as treatment interruption or treatment discontinuation will be treated as non-responders, and use of prohibited medications will not be taken into consideration.

13.2.1 Statistical Hypotheses

For these secondary efficacy endpoints, the null hypothesis for the treatment comparison is that there is no difference in the proportion of responders in the tildacerfont treatment groups compared to the proportion of responders in the placebo treatment group after the 12-week, double-blind, placebo-controlled treatment period. The alternative hypothesis will be that there is a difference between the proportions of responders in each treatment group.

$$H_0: \pi_{tildacerfont} - \pi_{placebo} = 0;$$

$$H_A: \pi_{ildacerfont} - \pi_{placebo} \neq 0$$

13.2.1.1 Statistical Modeling and Testing

A logistic regression model will be used to evaluate whether the number of subjects who achieve control of DHEAS is different between the tildacerfont treatment groups and the placebo treatment group after 12 weeks of double-blinded, placebo-controlled, dose-escalating treatment. The model will include the responder indicator (Section 9.3.11) as the dependent variable for the Week 12 assessment; treatment group (0 = placebo, 1 = tildacerfont); age as well as baseline DHEAS (for Objective 2.1) or the multiple of baseline DHEAS x ULN (for Objective 2.2) as continuous covariates.

An exact score test derived from the logistic regression model, equivalent to an exact CMH test on small sample sizes, will be used to perform a two-sided test at a significance level of 0.05 to determine whether proportion of responders is different between the tildacerfont treatment group and the placebo treatment group at Week 12. If the logistic model does not converge, three Fisher's exact tests corresponding to tildacerfont compared to placebo will be utilized to assess if differences exist in the proportion of subjects meeting each of the responder criteria.

The statistical model used to characterize the secondary endpoints at Week 12 will also be used to characterize the tildacerfont response at earlier weeks, replacing the Week 12 outcome with the Week 8 and the Week 4 outcome.

Subjects with missing data are assumed to be non-responders in the secondary responder endpoints.

13.2.1.2 Reporting Results

13.2.1.2.1 Summary of Descriptive Statistics

For each responder analysis and treatment group, the number and proportions (expressed as percentages) of responders at each scheduled time point will be calculated. These proportions, along with 95% CIs, will be summarized by scheduled time point. In addition, the differences in proportions between a tildacerfont and placebo subjects and their 95% CIs will be presented. If less than 5 responders are in a treatment group, exact (Clopper-Pearson) CIs will be used.

13.2.1.2.2 Summary of Analysis Results

The parameter estimates (i.e., odds ratios, the 95% CI of odds ratios, nominal p-values) from the logistic model will be reported by scheduled time point (i.e., Week 4, Week 8, and Week 12) in reverse time order.

13.2.1.2.3 Graph Presentation

The proportion of responders from the model-based analyses will be displayed graphically by treatment group over time with the y-axis representing the proportion meeting the responder definition and the x-axis the visits Week 4, Week 8, and Week 12.

A cumulative distribution plot of responders will be plotted against baseline DHEAS. This figure will be produced by treatment group and overall.

13.2.1.3 Supportive Analyses

The analysis of the secondary endpoint will be repeated using the additional thresholds of 25% and 20% reduction from baseline in DHEAS.

13.2.1.4 Subgroups

The primary endpoint will be descriptively summarized for each subgroup specified in Section 10.8 for the mITT Analysis Set.

13.3 Exploratory Efficacy Analyses

The analysis population for the exploratory efficacy analyses is the mITT analysis set (Section 8.2).

The exploratory endpoints explore clinical outcomes: ACTH (Endpoint 3.1), DHEA (Endpoint 3.2), T (Endpoint 3.3), A4 (Endpoint 3.4), 17-OHP (Endpoint 3.5), 11OHA4 (Endpoint 3.6), 11OHT (Endpoint 3.7), 11KA4 (Endpoint 3.8), 11KT (Endpoint 3.9). These biomarkers will be descriptively summarized; no statistical analyses will be performed.

13.3.1 Reporting Results

Descriptive statistics (11-point) for the exploratory biomarker assessments, change from baseline, and % change from baseline will be presented by treatment group and scheduled time point (i.e., Baseline, Week 4, Week 8, and Week 12).

For each treatment group, a spaghetti plot of select hormone levels and change from baseline over time (Baseline, Week 4, Week 8, and Week 12) will be plotted with the means at each time point overlaid. The select hormones are ACTH, 17-OHP, and A4.

14 SAFETY ANALYSES

Safety will be assessed by AEs, SAEs, AEs leading to discontinuation/withdrawal, DLTs, AESIs, physical examination, vital signs, ECGs, clinical laboratory tests, and psychiatric evaluations for suicide risk by C-SSRS. All analyses of the safety data will be performed using

the safety analysis set, based on the actual treatment a subject received. All descriptive statistics (described in Section 9.1) will be presented by treatment group (including dose level).

14.1 Adverse Events

All AE verbatim terms reported on the eCRFs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA® version 24.1).

All reported AEs (including non-TEAEs) will be listed. Separate listings will be provided for SAEs, TEAEs leading to study drug discontinuation, DLTs, AESIs, SUSARS, AEs related to adrenal insufficiency, and AEs related to COVID-19.

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.

An overall summary of TEAEs will be presented. The overall summary will include the number and percentage of subjects experiencing any AEs, TEAEs, study drug related TEAEs, TEAEs by maximum severity (highest toxicity grade), SAEs, study drug related SAEs, DLTs, SUSARs, TEAEs of adrenal insufficiency, TEAEs leading to study drug discontinuation, and deaths.

The number and percentage of subjects experiencing each TEAE, study drug related TEAE, and SAEs will be summarized by treatment group according to system organ class and preferred term. An additional table including only preferred term summarized by treatment for the number and percentage of subjects experiencing each TEAE will be presented as well.

14.2 Clinical Laboratory Evaluation

Laboratory parameters hematology, chemistry, coagulation, lipid panel, thyroid panel, and other tests will be summarized by treatment 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. Subject-level listings will be provided for urinalysis and all the

aforementioned laboratory parameters. In addition, subjects meeting the following Hy's law criteria will be identified and listed:

- Alanine aminotransferase (ALT) > 3 × ULN
- Aspartate aminotransferase (AST) > 3 × ULN
- Total bilirubin (TBL) > 2 × ULN

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. All vital signs parameters will be listed.

14.4 12-Lead Electrocardiogram

Electrocardiogram (ECG) data, such as clinical interpretation of ECGs, heart rate (HR) values, and interval assessments of RR, PR, 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 thresholds are provided by Spruce (See Table 5 below).

Table 5 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 Borderline Prolonged	≤ 450 > 450 to 470 > 470 > 450, > 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

All responses from the Baseline/Screening Version of the Columbia–Suicide Severity Rating Scale (C-SSRS) assess 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".

15 CHANGES RELATIVE TO THE PROTOCOL-SPECIFIED ANALYSIS

The study intended to enroll up to 40 subjects but was discontinued early on the basis of a business decision made by the sponsor. Due to the resulting small sample size, the exploratory efficacy analyses described in Protocol Section 8.3.4 are replaced with summaries of descriptive statistics.

The MMRM analysis introduced in Section 13.1 of this SAP is not mentioned in the protocol. However, to avoid protocol amendment, inclusion of this analysis in the SAP – well in advance of the database lock – is meant to serve the same intent.

The secondary endpoint of 30% reduction from baseline in DHEAS is mentioned in the protocol and sections 6, 6.1.2, 9.3.12 and 13.2 of this SAP. However, since some subjects have entered the study up to 25% lower than anticipated DHEAS into the study, it is necessary to assess respond at lower threshold of reduction at 25% and 20% in addition to originally planned 30% reduction.

16 REFERENCE

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.

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

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

17 Appendix A: Schedule of Assessments

Table 6 Study SPR001-210 Optional DHEAS Screening Visit (Protocol Version 6.1 December 6, 2022)

CLINIC VISIT NUMBER	**Optional Visit**
STUDY DAY	**≤ 60 days before Day 1**
Informed consent for DHEAS screening[1]	X
DHEAS[2]	X

Table 7 Study SPR001-210 Schedule of Activities (Protocol Version 6.1 December 6, 2022)

	Screening	**Treatment Period 1**			**Treatment Period 2**		**Treatment Period 3**		**ETV/ Follow up**
CLINIC VISIT NUMBER		**1**		**2**		**3**		**4**	**5**
STUDY WEEK		**1**	**2**	**4**	**6**	**8**	**10**	**12**	**16**
STUDY DAY	**≤30 days before Day 1**	**1**	**14**	**28**	**42**	**56**	**70**	**84**	**Last dose +30 days**
Clinic Visit (C)[1], Telephone Contact (T)[1]	**C**	**C**	**T**	**C**	**T**	**C**	**T**	**C**	**C[2]**
Informed consent	X								

[1] Remote consenting is allowed
[2] If documentation of prior DHEAS elevation is not present, the blood draw for the DHEAS sample will be obtained 8 A.M. ± 1 hour
[1] All clinic visits and telephone contacts should be performed on the indicated study days. In cases where adherence to the foregoing schedule is not possible, the following visit windows apply: ± 1 days for clinic visits and telephone contacts.
[2] All subjects will return to the clinic (or, at the discretion of the Investigator, be contacted via telephone) for the ETV/Follow-up Visit.

	Screening	Treatment Period 1			Treatment Period 2		Treatment Period 3		ETV/ Follow up
CLINIC VISIT NUMBER		**1**		**2**		**3**		**4**	**5**
STUDY WEEK		**1**	**2**	**4**	**6**	**8**	**10**	**12**	**16**
STUDY DAY	**≤30 days before Day 1**	**1**	**14**	**28**	**42**	**56**	**70**	**84**	**Last dose +30 days**
Clinic Visit (C)[1], Telephone Contact (T)[1]	**C**	**C**	**T**	**C**	**T**	**C**	**T**	**C**	**C[2]**
Inclusion/exclusion criteria	X	X							
Demography	X								
Medical history	X								
Prior medications from past year	X								
Concomitant medications	X	X	X	X	X	X	X	X	X
Hepatitis B & C and HIV	X								
Urine drug screen	X								
Urinalysis	X	X[3]						X	
Pregnancy test for FCP[4]	X	X[3]		X		X		X	
PK[5]		X		X		X		X	
Vital signs, body weight, height[6]	X	X		X		X		X	
Full physical examination[7]		X[3]						X	

[3] Will be conducted pre-dose at Day 1
[4] A serum pregnancy test will be performed at screening. All other pregnancy tests will be urine pregnancy tests.
[5] Blood samples will be drawn for PK measurement at 8 A.M. (±1 hour) on clinic visit days.
[6] Vital signs consist of systolic and diastolic blood pressure, pulse rate, respiration rate, and body temperature. Vital signs will be measured after a 5-minute rest period. Height is measured only at Screening Visit.
[7] A full physical examination may exclude rectal, genitourinary, and breast exams.

	Screening	Treatment Period 1			Treatment Period 2		Treatment Period 3		ETV/ Follow up
CLINIC VISIT NUMBER		**1**		**2**		**3**		**4**	**5**
STUDY WEEK		**1**	**2**	**4**	**6**	**8**	**10**	**12**	**16**
STUDY DAY	**≤30 days before Day 1**	**1**	**14**	**28**	**42**	**56**	**70**	**84**	**Last dose +30 days**
Clinic Visit (C)[1], Telephone Contact (T)[1]	**C**	**C**	**T**	**C**	**T**	**C**	**T**	**C**	**C[2]**
C-SSRS[8]	X	X		X		X		X	
Abbreviated physical examination[9]				X		X			
12-lead ECG	X	X[3]		X		X		X	
Hematology, clinical chemistry, coagulation, lipid panel, eGFR, thyroid panel, cortisol[10]	X	X		X		X		X	
DHEAS, progesterone, T, LH, FSH, SHBG, 17-OHP, prolactin[10]	X								
ACTH, DHEA, DHEAS, T, 17-OHP, A4, estradiol, LH, FSH, SHBGH, 11OHA4, 11OHT, 11KA4, 11KT[10]		X		X		X		X	
Randomization to treatment		X							
Dispense study drug[11]		X		X		X			
Dose escalation				X		X			
Study drug accountability				X		X		X	

[8] See Protocol Section 7.1.6 for details on the administration of the C-SSRS.
[9] Abbreviated physical examinations at Visits 2 and 3 will be conducted directed by AEs.
[10] Blood draws for these laboratory assessment samples will be obtained at 8 A.M. (± 1 hour). On Day 1, blood draws should be performed prior to the first dose of study drug.
[11] Study drug will be taken daily with an evening meal (with approximately <50% fat content), which should be eaten between 6 P.M. and midnight. Study drug may be consumed up to 30 minutes after completing the evening meal, if necessary.

	Screening	Treatment Period 1			Treatment Period 2		Treatment Period 3		ETV/ Follow up
CLINIC VISIT NUMBER		**1**		**2**		**3**		**4**	**5**
STUDY WEEK		**1**	**2**	**4**	**6**	**8**	**10**	**12**	**16**
STUDY DAY	**≤30 days before Day 1**	**1**	**14**	**28**	**42**	**56**	**70**	**84**	**Last dose +30 days**
Clinic Visit (C)[1], Telephone Contact (T)[1]	**C**	**C**	**T**	**C**	**T**	**C**	**T**	**C**	**C[2]**
Telephone contact to assess compliance and AEs			X		X		X		
Review AEs		X	X	X	X	X	X	X	X

11OHA4 = 11β-hydroxyandrostenedione; 11OHT = 11β-hydroxytestosterone; 11KA4 = 11-ketoandrostenedione; 11KT = 11-ketotestosterone; 17-OHP = 17-hydroxyprogesterone; A4 = androstenedione; ACTH = adrenocorticotropic hormone; AE = adverse event; An = androsterone; DHEA = dehydroepiandrosterone; DHEAS = dehydroepiandrosterone sulfate; ECG = electrocardiogram; eGFR = estimated glomerular filtration rate; ETV = Early Termination Visit; FSH = follicle-stimulating hormone; HIV = human immunodeficiency virus; HIV = human immunodeficiency virus; LH = luteinizing hormone; PK = pharmacokinetics; T = testosterone

18 Appendix B: Clinical Laboratory Tests

Table 8 Study SPR001-210 Clinical Laboratory Tests

Laboratory Assessments	Parameters	
Hematology	Platelet count	
	RBC count RBC indices: MCV, MCH, % reticulocytes	
	Hemoglobin	
	Hemoglobin A1c	
	Hematocrit	
	WBC count Differential: neutrophils, lymphocytes, monocytes, eosinophils, basophils	
Clinical Chemistry	Potassium Calcium Sodium BUN Creatinine Total protein Fasting glucose	ALP ALT/SGPT AST/SGOT GGT Total and direct bilirubin Total bile acids Fasting insulin
Coagulation	PT/INR, PTT	
Lipid Panel	Total cholesterol, LDL, HDL, triglycerides	
Thyroid Panel	T3 (free and total), T4 (free and total), TSH	
Routine Urinalysis	Specific gravity pH, glucose, protein, blood, ketones, bilirubin, urobilinogen, nitrite, by dipstick Microscopic examination (if blood or protein is abnormal) Urine pregnancy testing	
Hormone Panel	DHEAS, ACTH, DHEA, T, A4, 17-OHP, 11OHA4, 11OHT, 11KA4, 11KT, FSH, LH, SHBG, prolactin, cortisol, estradiol, progesterone (blood [screening only]), HCG (pregnancy test at screening only)	

Abbreviations: 11KA = 11-ketoandrostenedione; 11KT = 11-ketotestosterone; 11OHA4 = 11β-hydroxyandrstenedione; 11OHT = 11β-hydroxytestosterone; 17-OHP = 17-hydroxyprogesterone; A4 = andrstenedione; ACTH = adrenocorticotropic hormone, corticotropin; ALP = alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase; BUN = blood urea nitrogen; DHEA = dehydroepiandrosterone; DHEAS = dehydroepiandrosterone sulfate; FSH = follicle-stimulating hormone; GGT = gamma-glutamyl transferase; HCG = human chorionic gonadotropic; HDL = high-density lipoprotein; INR = international normalized ratio; LDL = low-density lipoprotein; LH = luteinizing hormone; MCH = mean corpuscular hemoglobin; MCV = mean corpuscular volume; PT = prothrombin time; PTT = partial thromboplastin time; RBC = red blood cell; SGOT = serum glutamic-oxaloacetic transaminase; SGPT = serum glutamic-pyruvic transaminase; SHBG = sex hormone-binding globulin; T = testosterone; TSH = thyroid stimulating hormone; WBC=white blood cell.

19 Appendix C: Prohibited and Cautionary Concomitant Medications

The following is a non-exhaustive list of medications that are prohibited because of their potential for metabolic interactions with tildacerfont. This list is intended to show more commonly encountered drugs that subjects may be taking at Screening. Any drugs of concern should be discussed with the Medical Monitor.

Table 9 Study SPR001-210 Prohibited and Cautionary Concomitant Medications

alfentanil	apixaban	atorvastatin	avanafil	avasimibe
buspirone	carbamazepine	cenobamate	ciprofloxacin	darifenacin
diltiazem	elagolix	eletriptan	ethinyl estradiol (>35 µg)	felodipine
fluconazole	glyburide	isavuconazole	itraconazole	ketoconazole
lesinurad	loperamide	lovastatin	mibefradil	midazolam
nefazodone	oseltamivir	phenobarbital	phenytoin	posaconazole
quetiapine	repaglinide	rifampin	rifapentine	rivaroxaban
simvastatin	St John's wort	teneligliptin	ticagrelor	triazolam
voriconazole				

Many oncology drugs and medications used to treat the hepatitis C virus and HIV are strong inhibitors of CYP3A4 but are not listed above simply because individuals with active cancer, hepatitis C, and/or HIV are excluded from this study.