Official Title:	**A Double-Blind, Randomized, Placebo-Controlled Study to Evaluate the Efficacy and Safety of CTP-543 In Adult Patients With Moderate to Severe Alopecia Areata**
NCT Number:	NCT04518995
Document Date:	SAP Version 1: 15 April 2022

9. DOCUMENTATION OF STATISTICAL METHODS

CP543.3001, Statistical Analysis Plan, Version 1.0, 15 April 2022

[redacted]

STATISTICAL ANALYSIS PLAN

STUDY TITLE:

A DOUBLE-BLIND, RANDOMIZED, PLACEBO-CONTROLLED STUDY TO EVALUATE THE EFFICACY AND SAFETY OF CTP-543 IN ADULT PATIENTS WITH MODERATE TO SEVERE ALOPECIA AREATA

PROTOCOL NUMBER:

CP543.3001

SPONSOR: Concert Pharmaceuticals, Inc.
IND NUMBER: 131,423
PREPARED BY: [redacted]

ACKNOWLEDGEMENT AND SIGNATURE SHEET

[redacted] Project Statistician Approval	Concert Approval
[redacted], [redacted]	[redacted]
Signature and Date	Signature and Date
DocuSigned by: [redacted] Signer Name: [redacted] Signing Reason: I approve this document Signing Time: 15-Apr-2022 | 3:25:53 PM EDT [redacted]	DocuSigned by: [redacted] Signer Name: [redacted] Signing Reason: I approve this document Signing Time: 15-Apr-2022 | 3:28:17 PM EDT [redacted]

VERSION HISTORY

SAP Version	Version Date	Change(s)	Rationale
1.0	2022-04-15	First version	

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. LIST OF ABBREVIATIONS

Table 1: List of Abbreviations

Abbreviation	Term
ANCOVA	Analysis of Covariance
ATC	Anatomical-Therapeutic-Chemical
BELA	Brigham Eyelash Tool for Alopecia
BETA	Brigham Eyebrow Tool for Alopecia
BID	Twice daily dosing
CGI-I	Clinician Global Impression of Improvement
CGI-S	Clinician Global Impression of Severity
CI	Confidence interval
CSR	Clinical study report
CTCAE	Common terminology criteria for adverse events
DMC	Data Monitoring Committee
ECG	Electrocardiogram
eCRF	Electronic case report form
FCS	Fully conditional specification
FDA	Food and Drug Administration
HADS	Hospital Anxiety and Depression Scale
IQR	Interquartile range
LS	Least squares
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MMRM	Mixed effect model of repeated measures
MNAR	Missing not at random
PCS	Potentially clinically significant
PGI-I	Patient Global Impression of Improvement
PGI-S	Patient Global Impression of Severity
PML	Progressive multifocal leukoencephalopathy
QPRO	Quality of Hair Patient Reported Outcome
SALT	Severity of Alopecia Tool
SAP	Statistical analysis plan

SPRO	Satisfaction of Hair Patient Reported Outcome
TEAE	Treatment-emergent adverse events
WHO	World Health Organization

2. PURPOSE OF THE ANALYSES

This statistical analysis plan (SAP) provides a detailed description of the strategy and statistical methodology to be used for analysis of data from the CP543.3001 protocol, amendment 5.0, dated 28APR2021.

The purpose of the SAP is to describe the pre-specified statistical approaches to the analysis of study data prior to database lock. This analysis plan is meant to supplement the study protocol. If differences occur between analyses described in the SAP and the current protocol, those found in this SAP will assume primacy. Any deviations from this plan will be described in the Clinical Study Report (CSR).

3. PROTOCOL SUMMARY

This is a double-blind, randomized, placebo-controlled multicenter study to evaluate the efficacy and safety of CTP-543 in adult patients with moderate to severe alopecia areata. Patients will be between 18 and 65 years of age and experiencing an episode of hair loss associated with alopecia areata lasting at least 6 months and not exceeding 10 years. Patients not currently being treated for alopecia areata or with other treatments that might affect hair regrowth or immune response must have at least 50% hair loss as measured by Severity of Alopecia Tool (SALT) at Screening and Baseline.

An independent Data Monitoring Committee (DMC) will perform regular safety assessments based on a review of cumulative safety data. The DMC may advise study and/or treatment arm cessation due to intolerability at any time.

The Screening Period may last up to 28 days prior to initiation of study drug. The Treatment Period is a 24-week, double-blind, placebo-controlled period. Following the 24-week Treatment Period, patients will have the opportunity to continue receiving treatment in an Open-Label Extension study. If a patient does not wish to continue into the Open-Label Extension study, they will complete treatment at Week 24 and return in 4 weeks for the Post-Treatment Safety Follow-up to assess safety following treatment completion. Study design diagram can be found in Appendix 15.1.

Patients will provide appropriately-obtained informed consent prior to initiating any screening procedures. Patients meeting initial screening criteria will be eligible to continue to the Day 1 visit for review of eligibility and baseline assessments, including SALT assessment, physical examination, clinical laboratory assessments, and vital signs. Patients meeting all inclusion criteria and none of the exclusion criteria will be randomized to twice-daily treatment with either 12 mg CTP-543, 8 mg CTP-543 or placebo. Patients will be randomized in a 3:5:2 ratio (12 mg BID:8 mg BID:placebo). Randomization will be stratified by scalp hair loss into one of the following two categories: 1) Partial scalp hair loss (SALT $\geq$ 50 and $<$ 95); 2) complete or near-complete scalp hair loss (SALT $\geq$ 95).

Patients will take the first dose of study drug in the clinic on Day 1 and will be instructed to take study drug daily approximately every 12 hours for the duration of the Treatment Period. Patients should dose study drug in the clinic on all Study Visit Days after clinical laboratory blood draws are completed. Other baseline assessments will include patient and clinician global impression of disease severity (PGI-S and CGI-S), Patient Reported Outcome for Satisfaction (SPRO) and Hair Quality (QPRO), the Hospital Anxiety and Depression Scale (HADS), and assessments for eyelash and eyebrow involvement. The schedule of assessments can be found in Appendix 15.2.

The double-blind, placebo-controlled Treatment Period will last 24 weeks. Assessment of treatment response using SALT for efficacy will occur at 4, 8, 12, 16, 20 and 24 weeks.

Patient safety will be monitored throughout the trial by the Investigator and supported by regular review by the Medical Monitor and DMC. Chemistry and hematology laboratory values will be assessed under fasted conditions bi-weekly for the first month of the treatment period, followed by every 4 weeks through the remainder of the study. Lipid levels will be assessed every 4 weeks throughout the Treatment Period. In light of the COVID-19 pandemic, challenges in the

conduct of study visits may arise from quarantines, site closures, and/or travel limitations. To ensure the safety of trial participants and to minimize risk to trial integrity, alternative methods for drug dispensation and assessments, including telemedicine visits and clinical laboratory blood draws at a local laboratory or via a Home Health Care agency, may be offered for those trial participants who may no longer have access to the investigational site. Significant cytopenias or other hematologic abnormalities will be managed by severity through dose interruption or discontinuation, and signs and symptoms of infection will be closely monitored and treated promptly. Patients who experience intolerable symptoms during treatment may discontinue treatment at the judgement of the Investigator, but will be followed for safety and efficacy assessments until Week 24, unless they withdraw consent. Patients may withdraw consent at any time.

Following the 24-week Treatment Period, patients will have the opportunity to continue receiving treatment in an Open-Label Extension study. If a patient does not wish to continue into the Open-label Extension study, they will complete treatment at Week 24 and return in 4 weeks for the Post-Treatment Safety Follow-up to assess safety following treatment completion.

4. GENERAL ANALYSIS AND REPORTING CONVENTIONS

The following is a list of general analysis and reporting conventions to be applied to this study, unless otherwise specified. Departures from these general policies will be described, if applicable, in the appropriate sections of this SAP. When this situation occurs, the rules set forth in the specific section take precedence over the general policies.

All data displays (tables, listings, and figures) will have a header showing the sponsor company name, protocol number, page number, and display status (i.e. "DRAFT" or "FINAL"), as well as a footer indicating file name and run date/time. Summary tables and data listings will be summarized by active treatment versus placebo-treated patients (i.e., by treatment group) for efficacy and exposure, and by treatment group and overall otherwise. All data for analysis will be listed by-patient.

Categorical variables will be summarized using counts (n) and percentages (%) and will be presented in the form "n (xx.x)." If a count is 0, 0% will be shown for the percentage. If a percentage is 100%, 100.0% will be shown. To ensure completeness, summaries for categorical variables will include all categories, even if no patients had a response in a particular category. Unless otherwise specified, the denominator for each percentage will be based on the number of patients in the population being summarized (header n). Percentages will be rounded to one decimal place.

Continuous variables will be summarized using mean, standard deviation, minimum, maximum, median, and number of patients. The minimum and maximum will be the same precision as the original collected data. The mean and median will be reported to an additional level of precision than the original observations, and the standard deviation will be reported to two additional levels of precision than the original observations. The precision for some original collected data and derived variables are longer than required for clinical interpretation (eg, >15 decimals). In these situations, the number of decimal places will be determined by the table of precision included as an attachment.

Results of statistical models (ie, least-squares means, confidence intervals, risk differences) will be rounded to two decimal places. Standard error will be reported to an additional level of precision. Results which cannot be determined will be displayed as "NA", "NE", or similar.

For tests of hypothesis of treatment group differences, the associated p-value will be reported in rounded format. All p-values will be rounded to four decimal places after assignment of statistical significance; p-values that round to 0.0000 or are reported by SAS as zero will be presented as "≤0.0001"; p-values that round to 1 or are reported by SAS as 1 will be presented as ">0.9999". P-values which are significant at a 0.05, 0.025, 0.01, 0.001, or 0.0001 level will be marked with asterisks. P-values which cannot be determined will be displayed as "NA". P-values are descriptive for outcomes and analyses not included in the multiplicity algorithm.

Date of last dose is collected on the *Treatment Discontinuation/Completion* eCRF page. For patients with missing or partial date of last dose, the date of last dose (ie, treatment end date) will be imputed. Date of last dose will be imputed as follows:

- Derive the number of tablets taken at the last visit as the dispensed amount minus the returned amount. For treatment compliance (Section 7.3) and exposure (Section 9.2), if the returned amount is missing, it will be assumed that zero tablets were taken. Otherwise, it will be assumed that 60 tables were taken.

- An estimated date of possible last dose will be calculated by adding the number of tablets taken divided by 2 to the visit date associated with the last visit where drug accountability was collected.

- If there is a partial date of last dose, missing day will be imputed using the day from treatment start date, or the last day of the month if that month has fewer days than the treatment start month.

- Treatment end date will be the minimum of the partial date of last dose, calculated estimated date of possible last dose, and collected study discontinuation/completion date.

Summary tables and data listings:

- Unless otherwise specified, no preliminary rounding will be performed; rounding will only occur after analysis.
- Data from patients excluded from an analysis population will be presented in the data listings but will not be included in the calculation of summary statistics, where applicable.
- Data from each patient will be separated by a blank line. Within a data listing, if a descriptive item appears line after line (e.g., repetition of a patient number, date, visit, etc.), only the first occurrence will be displayed (e.g., in Listing of Vital Signs, patient number, date and visit will only be displayed on first row when presenting all parameters collected at same visit). Repetition of actual results or outcomes (e.g., adverse events, lab results, vital sign values, etc.) will not be collapsed.
- Data listings will be sorted by treatment, patient, and date and/or time of assessment, as applicable.
- When change from baseline or change to baseline is calculated, baseline is the last observation obtained prior to first dosing of the study drug.
- Unless otherwise noted, values reported as greater than or less than some quantifiable limit (e.g., "< 2.0") will be summarized as using the numeric equivalent in summary tables and figures (e.g. "<2.0" will be analyzed as 2.0). The reported value will be used in listings.

Mock tables and data listings will be provided as attachments to this analysis plan. Minor changes to the mocks after formal SAP approval will not necessitate re-approval unless changes to the text of the SAP are required.

All statistical deliverables will be produced, validated, and reviewed for accuracy/consistency in accordance with [redacted]. standard operating procedures and the processes described in the statistical validation plan.

SAS® (SAS Institute, Cary, North Carolina) statistical software, version 9.4 or later, will be used for all analyses. Adverse Events and Medical History will be coded in Medical Dictionary for Regulatory Activities (MedDRA) Version 23.1. Concomitant medications will be coded in World Health Organization (WHO) Drug Version 2020-SEP and Anatomical-Therapeutic-Chemical (ATC) classification and preferred term.

5. ANALYSIS SAMPLES

The Efficacy Population will include all patients who are randomized in the study and dispensed study drug during the Treatment Period.

The Safety Population will include all patients who receive study drug during the Treatment Period.

The Per Protocol Population will include all patients in the Efficacy Population who were dosed according to protocol and have no major protocol deviations. Inclusion in the Per Protocol Population will be determined prior to breaking the study blind. Further details on the identification of major protocol deviations are provided in Section 7.1.

In the event patients incorrectly receive treatment associated with a different treatment group, they will be analyzed “as randomized” for efficacy and “as treated” for safety.

6. STUDY SUBJECTS

6.1. Disposition of Subjects

Disposition will be summarized by randomized treatment group and overall for all patients screened in the study. The following disposition information will be summarized (percentages based on the number randomized, with the exception of the reasons for discontinuation):

- The number of patients screened. Patients who rescreened will be counted only once.
- The number of patients randomized.
- The number and percentage of randomized patients within the Efficacy, Safety, and Per Protocol Populations.
- The number and percentage of patients who completed treatment, defined as patients with either "Complete-Entering Open Label" or "Complete" noted for the question "Subject Status" on the *Treatment Discontinuation/Completion* electronic case report form (eCRF) page.
- The number and percentage of patients who completed treatment and enrolled in the Open Label Extension, defined as patients with "Complete-Entering Open Label" noted for the question "Subject Status" on the *Study Treatment Discontinuation/Completion* eCRF page.
- The number and percentage of patients who completed treatment and completed the follow-up visit (for those patients who did not enroll in the Open Label Extension), defined as patients with "Complete" noted for the question "Subject Status" on the *Treatment Discontinuation/Completion* eCRF page.
- The number and percentage of patients who prematurely discontinued treatment, and the frequency and percentage of each discontinuation reason. The denominator for the percentage of each discontinuation reason will be the number of patients who discontinued treatment.
- The number and percentage of patients who completed the study, defined as patients with either "Complete-Entering Open Label" or "Complete" noted for the question "Subject Status" on the *Study Discontinuation/Completion* eCRF page.
- The number and percentage of patients who completed the study and enrolled in the Open Label Extension, defined as patients with "Complete-Entering Open Label" noted for the question "Subject Status" on the *Study Discontinuation/Completion* eCRF page.
- The number and percentage of patients who completed the study and who did not enroll in the Open Label Extension, defined as patients with "Complete" noted for the question "Subject Status" on the *Study Discontinuation/Completion* eCRF page.
- The number and percentage of patients who prematurely discontinued the study, and the frequency and percentage of each discontinuation reason. The denominator for the percentage of each discontinuation reason will be the number of patients who discontinued the study.

The number and percentage of patients randomized within each treatment group will also be summarized by investigative site.

Disposition and patient visits will also be presented for each patient in patient data listings. Disposition patient data listing will list date of informed consent, date of first/last treatment, date of end of treatment/early treatment termination, date of end of study/early study termination, and reasons for treatment or study discontinuation. A Kaplan Meier curve of time to early treatment termination for any reason will also be presented for the Safety Population. Patients who have completed treatment will be censored at date of last treatment. Screen failures will be provided in a separate listing. Rescreened patients who later enrolled will not be listed as screen failures.

6.2. Demographic and Other Baseline Characteristics

Demographic characteristics (ie, sex, ethnic origin, race, date of birth, and calculated body mass index) will be collected at the Screening Visit, between Day -28 and -1, and detailed on the eCRF. Demographic characteristics will be summarized using descriptive statistics for all randomized patients by treatment group and overall.

Baseline characteristics will be summarized using descriptive statistics for all randomized patients by treatment group and overall and will include the following:

- baseline total SALT score
- Alopecia areata classification (randomization strata via IWRS): partial scalp hair loss (SALT ≥ 50 and < 95); complete or near-complete scalp hair loss (SALT ≥ 95)
- duration of current episode in years (calculated as: [date of randomization – date of current episode onset + 1] / 365.25)
- current nail involvement: yes; no (collected via eCOA system as ‘HAND IMAGE CAPTURE’)
- nasal hair involvement: yes; no (collected on the Physical Examinations eCRF page)
- eyebrow involvement: yes; no (collected as part of the BETA assessment)
- eyelash involvement: yes; no (collected as part of the BELA assessment)
- whether the patient has experienced any past and/or concomitant diseases, conditions, or exposures to serious infections such as HIV, or past surgeries: yes; no (collected via the Medical History Summary eCRF page)

Other co-morbidities may be identified and included as baseline characteristics.

All demographics and baseline characteristics will also be listed within a by-patient data listing for all randomized patients.

6.3. Prior and Concomitant Medications

Concomitant medications will be recorded in the eCRF from Screening through study completion at Week 24 or the Follow-Up Visit at Week 28. All medications will be coded using the WHO Drug Dictionary Version 2020-SEP.

Medications are classified as prior if the medication started and stopped prior to the first dose date of study drug or as concomitant if used on or after the first dose date of study drug.

Prior and concomitant medications will be summarized with counts and percentages separately for the Safety Population by treatment group, WHO Drug ATC classification level 2 and preferred term. For each summary, a subject will be counted only once for each medication.

Partial and completely missing dates will be imputed for the purposes of classifying concomitant medications as follows:

- Partial dates will be imputed following the same algorithm as in Section 9.3 for Treatment-emergent adverse events (TEAEs).
- For an entirely missing stop date (ie day, month, and year are missing), the medication will be treated as ongoing.

Prior and concomitant medications will also be listed by-patient.

6.4. Medical History

Medical history will be coded with the MedDRA terminology Version 23.1 and summarized as counts and percentages for the Safety Population by treatment group, system organ class, and preferred term. A patient will be counted only once for each condition. Conditions will be listed, including the verbatim investigator description of the relevant medical condition, the coded terms (system organ class, preferred term), start date, end date, and whether or not the condition is ongoing.

7. STUDY OPERATIONS

7.1. Protocol Deviations

A protocol deviation is any noncompliance with the clinical trial protocol or International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use Good Clinical Practice requirements. The noncompliance may be either on the part of the participant, the investigator, or the study site staff. Protocol deviations will be recorded on the source documents and with an explanation for the deviation. All protocol deviations will also be recorded as specified in the monitoring plan.

Protocol deviations will be identified by site staff, through medical reviews, and by clinical research associates during site monitoring. A data review will be conducted before database lock by the sponsor to classify protocol deviations as minor or major. Deviations that may alter or confound interpretation of the study results will be classified as major deviations. Protocol deviations will be summarized as counts and percentages by treatment group, deviation severity, and category for all randomized patients. Deviations will also be presented by patient in a data listing for all randomized patients.

7.2. Randomization

Approximately 700 patients are planned to be randomized in the study. Patients will be randomized in a 3:5:2 ratio (12 mg BID:8 mg BID:placebo) to each of three study arms:

- CTP-543 12 mg BID
- CTP-543 8 mg BID
- Placebo BID

Randomization will be stratified by scalp hair loss into one of the following two categories: 1) Partial scalp hair loss (SALT $\geq$ 50 and < 95); 2) Complete or near-complete scalp hair loss (SALT $\geq$ 95). The randomization schedule will be generated prior to study start.

All study patients, Investigators, and site study staff will be blinded to study drug assignment for the duration of the study. Tablets and packaging of CTP-543 and placebo will be identical in appearance.

A by-patient listing of randomized treatment group and randomization number will be presented.

7.3. Measures of Treatment Compliance

Patients will strive for 100% compliance with the daily dosing schedule. Treatment compliance will be summarized as percent of planned dose received for each dosing regimen. Percent of planned dose received will be calculated for the entire treatment period as follows:

100*(Tablets Dispensed-Tablets Returned)/(Tablets Expected)

Tablets Expected is defined as the time on treatment multiplied by the expected number of pills taken daily (x2 for the 12 mg BID group; x2 for the 8 mg BID group; x2 for the placebo group), where time on treatment is defined as treatment end date – treatment start date + 1. Dose interruptions will be ignored in this calculation. Derived patient compliance, compliance per the

eCRF (80% or higher versus less than 80%), and dosing exceptions will be listed in by-patient data listings.

For patients who roll over into one of the open label extension studies, it will be assumed that only 1 dose of study drug was administered for the last dose of study drug. For patients who do not roll over, it will be assumed that 2 doses of study drug were administered for the last dose of study drug.

8. ENDPOINT EVALUATION

8.1. Overview of Efficacy Analysis Methods

8.1.1. Multicenter Studies

Patients will be enrolled at up to 70 sites; site identifiers will be included in by-patient listings.

8.1.2. Assessment Time Windows

The visit schedule for all study assessments is provided in Appendix 15.2. Patients will be considered completed for efficacy analyses after Week 24 (Visit 9). Patients will either complete treatment at Week 24 and have a safety follow-up visit at Week 28 (Visit 10) prior to exiting the study, or if eligible, may enroll into an open-label extension study of CTP-543 at their Week 24 visit.

Generally, for scheduled visits where valid assessments were obtained, there will be no reassignment of the analysis visit based on date, and all data will appear in summary tables based on the nominal timepoint. For safety parameters, an unscheduled visit may take priority over a scheduled visit if the unscheduled visit was performed due to invalid collection.

Unscheduled visits and early termination visits will be reassigned based on the analysis windows in the below table and the schedule of events. Assessments may only be re-assigned to visits where they were to be collected per the schedule of events. If an assessment was already documented at that visit, the scheduled data will be used in summaries. If the scheduled data were missing or invalid, then the re-assigned unscheduled visit/early termination visit will be used in summaries. All visits will be included in by-patient data listings.

Table 2: Analysis Windows for Unscheduled and Early Termination Visits

	Safety Analyses			Efficacy Analyses		
Nominal Visit	**Lower**	**Target**	**Upper**	**Lower**	**Target**	**Upper**
Screening	-28	-1	-1	-28	-1	-1
Visit 2-Randomization	1	1	1	1	1	1
Visit 3-Week 2	2	15	22	12	15	18
Visit 4-Week 4	23	29	43	26	29	32
Visit 5-Week 8	44	57	71	54	57	60
Visit 6-Week 12	72	85	99	82	85	88
Visit 7-Week 16	100	113	127	110	113	116
Visit 8-Week 20	128	141	155	138	141	144
Visit 9-Week 24	156	169	183	166	169	172
Visit 10-Week 28	184	197	211	194	197	200

8.1.3. Timing of Analyses

All final, planned analyses will be performed after the last patient has completed all study assessments, all relevant study data have been processed and integrated into the analysis database, and the database has been locked.

Any post hoc, exploratory analyses completed to support planned study analysis, which were not identified in this SAP, will be documented and reported in appendices to the CSR. Any results from these unplanned analyses (post hoc) will also be clearly identified in the text of the CSR.

8.1.4. Multiple Comparisons/Multiplicity

Hierarchical testing will be used to control for multiple comparisons (Figure 1). For the primary efficacy analysis and the key secondary efficacy analyses, a ≤0.025 (two-sided) alpha-level will be allocated to each dose for comparison with placebo to adjust for multiplicity of testing associated with two doses. Once p>0.025 is observed for a treatment group comparison, inference for additional endpoints is no longer alpha-level protected for that dose group; however, nominal p-values will still be reported.

Subsequently, conditional upon significance (≤ 0.025) for the primary endpoint and first key secondary endpoint for each dose, a linear hierarchical testing approach with ≤ 0.025 (two-sided) alpha-level will be used to test the remaining key secondary endpoint comparisons versus placebo. Once p>0.025 is observed for a treatment group comparison, inference for additional endpoints is no longer alpha-level protected for that dose group; however, nominal p-values will still be reported.

All other statistical tests outside of the hierarchy will be carried out at a two-sided significance level of ≤ 0.05. Nominal p-values will be presented without adjustment for multiple comparisons.

8.2. Primary Endpoint

The primary efficacy endpoint is the percentage of patients achieving an absolute SALT score of ≤20 at Week 24 (i.e., the binary endpoint denoting a SALT score ≤ 20).

8.2.1. Computation of the Primary Endpoint

The SALT is a measure of hair loss that quantifies the amount of scalp surface without hair in a pre-specified quadrant of the scalp (right side, top, left side, back), that is further summed after applying a quadrant-specific multiplier indicative of scalp surface area contribution, to provide an overall score of total hair loss. A total SALT score equal to 0% indicates complete hair growth. A total SALT score equal to 100% indicates complete hair loss. An example of the SALT assessment can be found in Appendix 17.2 of protocol. The overall total SALT score is calculated in the eCOA system and will not re-derived. SALT scores are considered missing if the overall total could not be calculated.

The primary endpoint is derived by categorizing the individual overall SALT scores at Week 24 (≤ 20 versus > 20).

A multiple imputation (MI) approach will be used to estimate missing SALT scores.

One hundred independent imputations will be done with SAS® PROC MI using the Fully Conditional Specification (FCS) regression method. The FCS method combines both the monotone and arbitrary missing data patterns into one procedure, imputing missing values with a sequential regression approach.

For the primary efficacy analysis, missing SALT scores will be imputed under a missing at random (MAR) assumption. This assumes that patients who have missing data would have similar SALT scores as other patients in their respective treatment arm who have complete data. Any missing post-baseline SALT scores will be imputed to better inform the estimates of Week 24 missing SALT scores. Patients who have discontinued treatment but remain on study will not have their SALT scores censored, but any arbitrary missing SALT scores after the patient discontinued treatment will be imputed in the same manner. Utilizing the FCS method, missing SALT scores will be imputed in a sequential manner using regression models with baseline SALT score and the observed or imputed values of previous visits as covariates. Imputations will be implemented separately for each treatment under the assumption that different treatments may have distinct posterior distributions.

In addition, prior to database lock, the classification of the missing data pattern for missing SALT scores will be finalized based on blinded data review and documented as either considered to be missing at random (MAR) or potentially missing not at random (MNAR), consistent with general categories of reasons for missingness. Subjects with missing data due to COVID-19 will have data reviewed to determine the specific limitation imposed by COVID-19 leading to the inability to have captured the protocol-specified assessments. For example, missing at random will be considered plausible for COVID-19 related reasons such as site closures and travel limitations. In general, however, if a MAR assumption is not considered plausible, MNAR will be assumed. Patterns of missing values across visits will be listed and summarized with numbers and percentages by treatment and visit.

Multiple imputation will be done using the Efficacy Population and will not be repeated separately for the Per Protocol Population.

8.2.2. Primary Analysis of the Primary Endpoint

Pairwise treatment group differences from placebo will be assessed with the Mantel-Haenszel test (common risk difference) using baseline scalp hair loss (partial vs complete/near-complete) as the stratification factor, for each active treatment group versus placebo.

For each imputed dataset, treatment differences will be assessed using the Mantel-Haenszel estimate of a common risk difference, with the variance estimator (used to construct confidence limits for the Mantel-Haenszel estimate of the common risk difference in the SAS FREQ Procedure). The resulting 100 estimates of the treatment differences and standard errors will then be combined using SAS PROC MIANALYZE. The following SAS options for PROC FREQ will be used to estimate a common risk difference, confidence intervals, and p-values by imputation for placebo compared to each dose separately:

proc freq data = < > ;

by IMPUTATION;

tables STRATA*TREATMENT*RESPONSE / *commonriskdiff(cl=mh test=mh);*

run;

The observed number and percentage of patients with an absolute SALT score ≤ 20 among non-imputed data will be reported by treatment group along with the appropriate p-value, the combined treatment differences, combined standard errors, and 95% confidence intervals (CI).

This analysis will be completed for the Efficacy Population and the Per Protocol Population. The analysis utilizing the Per Protocol Population is considered supportive. The results may also be displayed graphically.

8.2.3. Sensitivity Analyses of the Primary Analysis

Sensitivity analyses of the primary efficacy endpoint will be conducted to evaluate the robustness of study results under different assumptions and imputation algorithms. For the primary endpoint, categorized as absolute SALT score ≤ 20 at Week 24, the following sensitivity analyses will be performed for the Efficacy Population only.

Sensitivity 1: Multiple imputation; missing not at random (MNAR)

This sensitivity analysis is an implementation of a pattern mixture model that draws from different populations based on the reason for withdrawal from study. Patients who discontinued the study early due to a treatment-emergent adverse event (including COVID-19) or due to lack of efficacy will be considered MNAR. Treatment-emergent adverse event and lack of efficacy are reasons collected under "If 'Withdrew Consent', specify reason" on the *Study Discontinuation/Completion* eCRF page.

For patients who discontinued the study early due to a treatment-emergent adverse event or lack of efficacy, SALT scores missing after discontinuation will be imputed using the control arm. This assumes that patients, regardless of treatment, who discontinue the study early for those reasons would have similar SALT scores to placebo patients with complete data.

For patients who discontinued the study due to reasons other than treatment-emergent adverse event or lack of efficacy, missing SALT scores after subjects discontinue the study early will be multiply imputed from subjects within the same treatment group who have complete data at that time. Terms will include baseline values and the weekly data through the time point being imputed.

The proportion of patients who have a SALT score ≤ 20 will be calculated for each imputed dataset. The primary analysis, as described in Section 8.2.2, will be repeated for the Efficacy Population using the MI with MNAR assumption datasets.

Sensitivity 2: Tipping point

Additionally, a tipping point sensitivity analysis will be conducted for each dose compared to placebo at Week 24 if the result of the MI analysis is statistically significant at the alpha ≤ 0.025 level, in favor of treatment, for the given dose. The tipping point analysis will use the primary MAR assumption datasets, where all imputed values are adjusted.

The tipping point analysis will apply delta adjustments for both treated and placebo. For the tipping point analyses, δ_T values (adjustments for treated patients) will represent a percentage difference between the imputed value and baseline and can vary from 0% to 100% of the difference for subjects with an imputed SALT score less than baseline, or zero otherwise. For the placebo arm, δ_P values will represent a percentage difference between the imputed value and 0 hair loss. Accordingly, the maximum delta value of 100% provides for an adjusted imputed value equal to either baseline (for treated subjects, representing non-response) or zero hair loss for placebo, representing response. Tipping points will therefore be provided for combinations of δ_T and δ_P to provide a range of assumptions about hypothetical improvements on placebo missing

data values and reduction in efficacy for treated imputed values such that there is no longer evidence of efficacy.

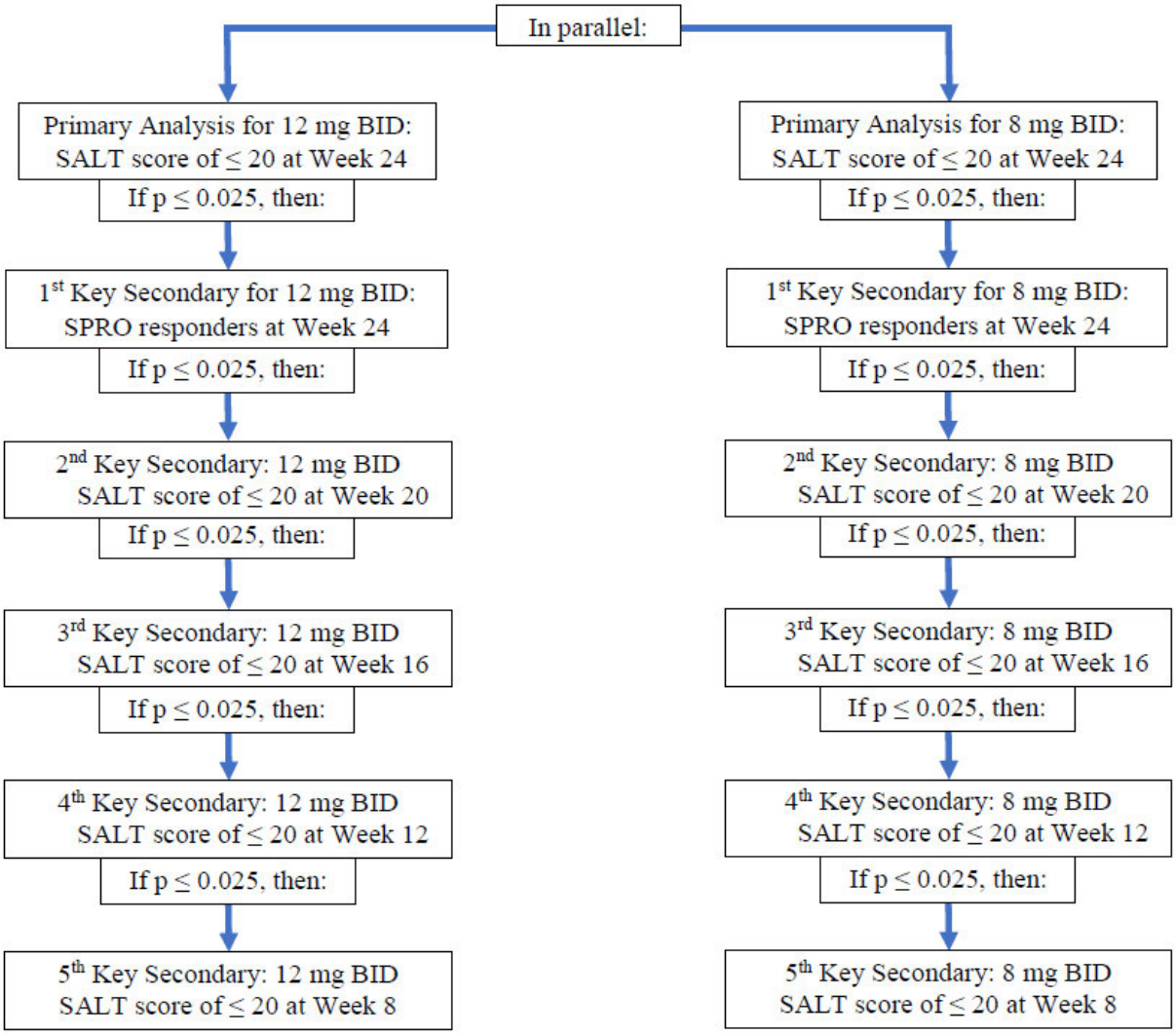
8.2.4. Secondary Analyses of the Primary Endpoint

No secondary analyses specified.

8.3. Key Secondary Endpoints

See Section 8.1.4 and Figure 1 below for hierarchical testing.

Figure 1: Hierarchical Testing



8.3.1. Hair Satisfaction Patient Reported Outcome (SPRO) at Week 24

The first key secondary endpoint is the percentage of responders per the SPRO at Week 24. The SPRO is a single item questionnaire answered by the patient, designed to measure how satisfied alopecia areata patients are with their hair at the time of the assessment. The SPRO responses are on a scale of 1-5, where 1 = very satisfied, 2 = satisfied, 3 = neither satisfied nor dissatisfied, 4 = dissatisfied, and 5 = very dissatisfied. An example of the SPRO assessment can be found in Appendix 17.4.1 of the protocol.

A responder is defined as patients with answers of "very satisfied" or "satisfied" at Week 24 on the SPRO. This endpoint will be assessed similar to the primary analysis for the primary endpoint, according to Section 8.2.2.

Missing data will be imputed similarly to the primary endpoint according to Section 8.2.1. However, since the SPRO scale is ordinal, the FCS logistic regression method will be used in lieu of the FCS regression method. The missing SPRO scores will be imputed in a sequential manner using logistic regression models with baseline SALT score, baseline SPRO score, and the observed or imputed SPRO values of previous visits as covariates.

The observed number and percentage of SPRO responders by treatment group will be reported along with the appropriate p-values. Both CTP-543 doses will be tested in parallel against placebo with a ≤ 0.025 alpha level. This analysis will be completed for the Efficacy Population and the Per Protocol Population. The Per Protocol Population will be considered supportive. If the primary endpoint is unsuccessful, nominal p-values will be reported.

8.3.2. SALT Score ≤ 20 at Weeks 20, 16, 12, and 8

The next key secondary endpoints are the percentage of patients achieving an absolute SALT score of ≤20 (i.e., the binary endpoint denoting a SALT score ≤ 20) at Weeks 20, 16, 12, and 8. Missing data will be imputed according to Section 8.2.1. These endpoints will be assessed according to Section 8.2.2. If the primary endpoint and first key secondary endpoint are unsuccessful, nominal p-values for the SALT key secondary endpoints will be reported. This analysis will be completed for the Efficacy Population and the Per Protocol Population. The Per Protocol Population will be considered supportive.

8.4. Secondary Endpoints

All statistical tests outside of the hierarchy defined in Section 8.1.4 and Figure 1 will be carried out at a two-sided significance level of ≤ 0.05. Nominal p-values will be presented. Analyses will be performed using the Efficacy Population.

Missing values will not be imputed for efficacy endpoints which are not derived from SALT or SPRO scores.

For the secondary endpoints of proportions of patients achieving specified improvement criteria (hereinafter, "secondary responder endpoints"), pairwise treatment group differences from placebo at each scheduled evaluation will be assessed with the Cochran-Mantel-Haenszel test using baseline scalp hair loss as the stratification factor. Observed counts and percentages of responders will be summarized by treatment group and the p-values reported. Secondary responder endpoints derived from SALT or SPRO scores will be analyzed similarly to the primary endpoint according to Section 8.2.2.

For the secondary endpoints of absolute or relative changes (hereinafter, "continuous secondary endpoints"), treatment differences will be assessed with a mixed-effect model of repeated measures (MMRM). Model covariates are described in applicable sections below. Treatment group comparisons at each visit will be based on least squares (LS) mean estimates. An unstructured covariance matrix will be used to model the within-subject errors. Should the model fail to converge, an alternative autoregressive covariance matrix will be used instead. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. Assuming that the data are MAR, the estimates of the treatment differences calculated from the MMRM described above are unbiased. Standard descriptive statistics will be reported for each time point on the values and changes from baseline along with the LS means, standard errors, 95% CIs and differences between treatment groups reported with LS means, standard errors, and 95% CIs.

8.4.1. Secondary SALT Endpoints

The following secondary SALT responder endpoints will be assessed:

- Percentage of patients achieving at least a 75% relative reduction in SALT score from Baseline at Weeks 12 and 24
- Percentage of patients achieving at least a 90% relative reduction in SALT score from Baseline at Weeks 12 and 24
- Percentage of patients achieving an absolute SALT score of ≤20 at Week 4
- Percentage of patients achieving an absolute SALT score of ≤10 at Week 24

In addition, the continuous secondary SALT endpoint of relative change in SALT score from baseline at Weeks 4, 8, 12, 16, 20, and 24 will be assessed with a MMRM. The model will include fixed factors for treatment group, study visit, and treatment by study visit interaction, with baseline SALT score as a covariate. Imputed values will not be used for the MMRM.

Relative reduction is based on relative change from baseline.

Relative change in SALT score is derived as percent change of the follow-up SALT score, where baseline SALT score is the denominator (i.e. absolute change divided by the baseline score, multiplied by 100).

8.4.2. Secondary SPRO Endpoints

The following secondary SPRO responder endpoints will be assessed:

- Percentage of responders at Weeks 12, 16, and 20, where responders is defined as "very satisfied" or "satisfied"
- Percentage of patients achieving a ≥ 2-point change from baseline in the SPRO scale at Weeks 12, 16, 20, and 24

In addition, the continuous secondary SPRO endpoint of change from baseline at Weeks 12, 16, 20, and 24 will be assessed with a MMRM. The model will include fixed factors for treatment group, study visit, and treatment by study visit interaction, with baseline SALT score and baseline SPRO score as covariates. Imputed values will not be used for the MMRM.

8.4.3. Clinical Global Impression of Improvement (CGI-I)

The CGI-I will be assessed by the Investigator at Weeks 12, 16, 20, and 24. Compared to the patient's alopecia areata prior to treatment at baseline, the patient's current state of alopecia areata will be assessed according to the Investigator's perceived change. The Investigator may select one of seven numeric choices representing "Very Much Worse" to "Very Much Improved". To reduce variability, one rater should perform the CGI-I assessment for the patient for the duration of the study. An example of the CGI-I assessment can be found in Appendix 17.3.2 of the protocol.

The secondary CGI-I responder endpoint of percentage of responders (defined as "much improved" or "very much improved") at Weeks 12, 16, 20, and 24 will be assessed.

8.4.4. Clinical Global Impression of Severity (CGI-S)

The CGI-S will be assessed by the Investigator and will consider the severity of the patient's alopecia areata at the time of assessment. The Investigator may select one of seven numeric choices representing "Among the most extreme hair loss" to "Normal, no hair loss". To reduce variability, one rater should perform the CGI-S assessment for the patient for the duration of the study. An example of the CGI-S assessment can be found in Appendix 17.3.1 of the protocol.

The continuous secondary endpoint of change from baseline in the CGI-S at Weeks 12, 16, 20, and 24 will be assessed with a MMRM. The model will include fixed factors for treatment group, study visit, and treatment by study visit interaction, with baseline SALT score and baseline CGI-S score as covariates.

8.4.5. Patient Global Impression of Improvement (PGI-I)

The PGI-I will be assessed by the patient at Weeks 12, 16, 20, and 24. Compared to the patient's alopecia areata prior to treatment at baseline, the patient's current state of alopecia areata will be assessed according to his/her perceived change. The patient may select one of seven numeric choices representing "Very Much Worse" to "Very Much Improved". An example of the PGI-I assessment can be found in Appendix 17.3.4 of the protocol.

The secondary PGI-I responder endpoint of percentage of responders (defined as "much improved" or "very much improved") at Weeks 12, 16, 20, and 24 will be assessed.

8.4.6. Patient Global Impression of Severity (PGI-S)

The PGI-S will be assessed by the patient and will consider the severity of his/her alopecia areata at the time of assessment. The patient may select one of seven numeric choices representing "Among the most extreme hair loss" to "Normal, no hair loss". An example of the PGI-S assessment can be found in Appendix 17.3.3 of the protocol.

The continuous secondary endpoint of change from baseline in the PGI-S at Weeks 12, 16, 20, and 24 will be assessed with a MMRM. The model will include fixed factors for treatment group, study visit, and treatment by study visit interaction, with baseline SALT score and baseline PGI-S score as covariates.

8.4.7. Brigham Eyebrow Tool for Alopecia (BETA)

The BETA is a clinician-rated scale that assesses the total eyebrow hair present. The scale is used for each eyebrow in which the Individual Eyebrow Score is derived by assessment of density (D) x surface area (SA). An expert panel will perform central readings based on photographs of the eyebrows and will provide a score for each patient with eyebrow involvement. The BETA will be performed at Baseline, Week 12, and Week 24.

The continuous secondary endpoint of change from baseline in the BETA at Weeks 12 and 24 will be assessed with a MMRM. The model will include fixed factors for treatment group, study visit, and treatment by study visit interaction, with baseline SALT score and baseline BETA score as covariates.

8.4.8. Brigham Eyelash Tool for Alopecia (BELA)

The BELA is a clinician-rated scale that assesses the total eyelash hair present. The scale is used for each eye in which the Total Eyelash Score is derived by assessment of density and distribution on the right and left eye. An expert panel will perform central readings based on photographs of the eyelashes and will provide a score for each patient with eyelash involvement. The BELA will be performed at Baseline, Week 12, and Week 24.

The continuous secondary endpoint of change from baseline in the BELA at Weeks 12 and 24 will be assessed with a MMRM. The model will include fixed factors for treatment group, study visit, and treatment by study visit interaction, with baseline SALT score and baseline BELA score as covariates.

8.4.9. Hair Quality Patient Reported Outcome (QPRO)

The QPRO questionnaire is a 4-item assessment which provides additional details on key attributes of patient hair and helps provide context to the SPRO response. Like the SPRO, the QPRO responses are on a scale of 1-5, where 1 = very satisfied, 2 = satisfied, 3 = neither satisfied nor dissatisfied, 4 = dissatisfied, and 5 = very dissatisfied. An example of the QPRO assessment can be found in Appendix 17.4.2 of the protocol.

The continuous secondary endpoints of change from baseline each individual item for QPRO at Weeks 12, 16, 20, and 24 will be assessed with a MMRM. The model will include fixed factors for treatment group, study visit, and treatment by study visit interaction, with baseline SALT score and baseline QPRO item score as covariates.

8.4.10. Hospital Anxiety and Depression Scale (HADS)

The HADS is a 14-item measure designed to assess anxiety and depression symptoms in medical patients. Items are rated on a 4-point severity scale. The HADS produces two scales, one for anxiety (HADS–A) and one for depression (HADS–D). An example of the HADS assessment can be found in Appendix 17.5 of the protocol.

The continuous secondary endpoints of change from baseline in HADS-A at Week 24 and change from baseline in HADS-D at Week 24 will be assessed with an ANCOVA. The model will include fixed factors for treatment group with baseline SALT score and baseline HADS score (HADS-A or HADS-D as appropriate) as covariates.

8.5. Other Endpoints

No other efficacy endpoints are specified.

8.6. Examination of Subgroups

No subgroup analyses are planned.

9. SAFETY EVALUATION

9.1. Overview of Safety Analysis Methods

All safety summaries will be descriptive with no statistical hypothesis testing and based on the Safety Population. Patients will be summarized according to the study drug received (ie, as treated), should it differ from the randomized treatment arm. Safety assessments will be summarized by treatment group and overall. All safety endpoints will be listed in by-patient data listings.

9.2. Extent of Exposure

Study drug exposure will be summarized for each treatment group and overall. The number of days on which study drug was dosed (at least partially), the number of days in the study, and total cumulative dose will be summarized for each treatment group.

The number of days on which study drug was dosed (at least partially) will be calculated as (treatment end date – treatment start date + 1) minus the total number of days with a dosing exception per the *Dosing Exceptions* eCRF page of partial dose (counted as 0.5 day missed), missed dose (counted as 1 day missed), or dose interruption (counted as 1 day missed). The number of days in the study will be calculated as (treatment end date – treatment start date + 1). Total cumulative dose will be calculated as the total number of tablets taken multiplied by 8 mg, 12 mg, or 0 mg for each applicable treatment.

The number and percentage of subjects with at least one dose interruption will also be summarized, where dose interruption is defined as a value of "Dose interruption" for the question "Reason for Exception" on the *Dosing Exceptions* eCRF page. The sponsor may review the free text field associated with the "Reason for Exception" prior to database lock in order to identify reasons related to the dose interruption criteria outlined in Protocol Section 9.2.1. Should these reasons be identified, the definition of dose interruption would be refined based on those findings.

Drug compliance will also be summarized, per Section 7.3.

9.3. Adverse Events

An adverse event is any untoward medical occurrence that may appear or worsen in a patient during the course of a study. It may be a new intercurrent illness, a worsening concomitant illness, an injury, or any concomitant impairment of the patient's health, including laboratory test values, regardless of etiology. Any worsening (ie, any clinically significant adverse change in the frequency or intensity of a pre-existing condition) should be considered an adverse event. A diagnosis or syndrome should be recorded on the adverse event page of the eCRF rather than the individual signs or symptoms of the diagnosis or syndrome. A worsening of the condition under study, alopecia areata, will not be reported as an adverse event.

All patients will be monitored for adverse events during the study. Assessments may include monitoring of the following parameters: the patient's clinical symptoms, laboratory, physical examination findings, or findings from other tests and/or procedures.

Adverse events will be coded using MedDRA version 23.1 and summarized by treatment group, system organ class, and preferred term.

An adverse event reported after informed consent, but before the first dose of study drug (i.e., Day 1), will be considered a pre-treatment adverse event. Treatment-emergent adverse events will be defined as any adverse event that occurs after administration of the first dose of study drug. Adverse events will be limited to those collected within this protocol (ie, for rollover subjects, AEs collected under the protocol of the rollover study will not be in this study). The number and percentage of patients who report TEAEs will be summarized by treatment group, system organ class, and preferred term.

Treatment emergent adverse events will also be summarized by intensity as well as relationship to study drug.

Partial dates will be imputed for the purposes of defining TEAEs as follows:

- For a missing start day where the month and year are present, the start day will be set to the first day of the month, unless 1) the first day of the month is before the date of administration of study drug and the month and year are the same as the month and year of the date of administration of study drug, and 2) the end date is on or after the date of administration of study drug or the end date is completely missing, in which case the start day will be set to the first day of administration of study drug.
- For a missing start day and month where the year is present, the start day and month will be set to January 1st, unless 1) January 1st is before the date of administration of study drug, and 2) the end date is on or after the date of administration of study drug or the end date is completely missing, in which case the start day and month will be set to that of the first date of administration of study drug.
- For an entirely missing start date (ie day, month, and year are missing), the start date will be set to the date of administration of study drug unless the stop date is prior to the date of administration of study drug, in which case the start date will be set to the stop date.
- For a missing end day where the month and year are present, the end day will be set to the last day of the month, unless the month and year are the same as the month and year of the last visit date for the patient, in which case the end day will be set to that of the patient's last visit date.
- For a missing end day and month where the year is present, the end day and month will be set to the patient's last visit date, unless the year of the patient's last visit date is greater than the end year, in which case the end day and month will be set to December 31st.
- For an entirely missing stop date (i.e. day, month, and year are missing), the stop date will be imputed as the patient’s last visit date.

Patients who report the same preferred term on multiple occasions will be counted once for the preferred term: under the highest severity when summarized by severity and under the closest relationship to study drug when summarized by relationship. Missing severity or relationship will not be imputed. If a patient reports multiple preferred terms for a system organ class, the patient will be counted only once for that system organ class.

The number and percentage of patients who experience TEAEs will be summarized by treatment group for the following:

- By system organ class and preferred term
- By severity/intensity, system organ class, and preferred term
- By relationship to study drug, system organ class, and preferred term
- Serious adverse events by system organ class and preferred term
- Serious adverse events by relationship to study drug, system organ class, and preferred term
- Adverse events resulting in discontinuation of study drug by system organ class and preferred term
- Adverse events that result in study drug dose interruption by system organ class and preferred term

Additionally, TEAEs will be summarized by time interval: 0 to 12 weeks, and 12 to 24 weeks. For each time interval, an incidence table will summarize only TEAEs with an onset date within the interval and a prevalence table will summarize all TEAEs that have an onset date within the interval or continue into the interval. Differences between the incidence and prevalence tables can provide insight into the duration of TEAEs as well as the recurrence of TEAEs. A preferred term for an individual patient will be reported in multiple time intervals if there are multiple adverse event reports. The denominator for each time interval will be the number of patients who received at least 1 dose of study drug within the interval.

By-patient listings will include all adverse events. TEAEs will be flagged.

9.4. Deaths, Serious Adverse Events, and Other Significant Adverse Events

A serious adverse event is an adverse event that fulfills the following criteria:

- Is fatal (results in death);
- Is life-threatening (Note: the term "life-threatening" refers to an event in which the patient was at risk of death at the time of the event; it does not refer to an event that could hypothetically have caused death had it been more severe);
- Requires inpatient hospitalization or prolongation of existing hospitalization (hospitalization is defined as an inpatient admission, regardless of length of stay);
- Results in persistent or significant disability/incapacity (a substantial disruption of the patient's ability to conduct normal life functions);
- Is a congenital anomaly/birth defect; or
- Constitutes an important medical event that may jeopardize the patient or may require medical or surgical intervention to prevent one of the other outcomes listed above.

Serious adverse events will be summarized according to Section 9.3.

By-patient listings will be provided for any deaths, serious adverse events, and adverse events leading to discontinuation of treatment. Treatment-emergent adverse events that result in dose interruption will also be identified.

9.5. Clinical Laboratory Evaluation

Clinical laboratory samples should be collected at the beginning of each clinic visit and just prior to a dose on all Study Visit Days and will be processed by a central laboratory. In the event a patient is unable to attend an in-clinic visit due to COVID-19, clinical laboratory blood draws may be performed by a Home Health Care agency or a local laboratory.

Clinical laboratory variables will be presented in 3 ways. First, change from baseline to the minimum post-baseline value, maximum post-baseline value, and final on-treatment value will be summarized. Baseline will be defined as the laboratory value obtained before the first dose of study drug on Day 1; if Day 1 values are unavailable, then values obtained at the Screening Visit will be used. Only post-baseline values on treatment will be eligible. The Follow-up Visit will be excluded from possible minimum, maximum, or final post-baseline values.

This analysis will be repeated with outlying values removed from eligibility for minimum and maximum values. An outlier is defined using the interquartile range (IQR). The IQR is equal to the third quartile minus the first quartile and will be calculated by visit and treatment group for each numeric parameter. The IQR will be multiplied by a constant of 1.5. Any value which is above 1.5 times IQR plus the third quartile or below 1.5 times IQR minus the first quartile will be considered an outlier and will be removed from eligibility for minimum, maximum, or final post-baseline values.

Second, shift tables will summarize the number and percentage of patients with categorical changes from baseline defined by the laboratory normal reference range for each laboratory parameter (Below reference range, Within reference range, and Above reference range). The baseline category will be cross-tabulated with post-baseline category (3 × 3 table). The post baseline category will be assigned and summarized in 2 ways for each laboratory parameter: minimum post-baseline value and maximum post-baseline value. Only post-baseline values on treatment will be eligible. Patients will only be reported in the shift tables if they have both a baseline and post-baseline value.

Third, treatment-emergent potentially clinically significant (PCS) laboratory values will be identified. PCS values are defined as those that meet Grade 3 or Grade 4 toxicity criteria from the Common Terminology Criteria for Adverse Events (CTCAE) criteria. Laboratory values will be graded using CTCAE version 5.0, for those values included. CTCAE grades will be based solely on numeric results and without regard to other symptoms based on clinical judgment. For laboratory parameters whose CTCAE grading is independent of baseline criteria, treatment-emergent PCS laboratory values are those in which the baseline value is not PCS and the post-baseline value is PCS. For laboratory parameters whose CTCAE grading is dependent upon baseline criteria, any post-baseline grading of Grade 3 or Grade 4 will be considered treatment-emergent. The number and percentage of patients with treatment-emergent PCS laboratory values will be summarized by treatment group for each clinical laboratory variable.

These analyses will be repeated for serum chemistry results, hematology results, and lipid results (where lipid tests include total cholesterol, low-density lipoprotein, high-density lipoprotein,

triglycerides). Laboratory results will also be listed. Reference ranges for each clinical laboratory parameter will also be summarized in a data listing.

Box plots of hemoglobin, platelets, and neutrophils will also be provided with dose interruption limits included as reference lines (see Protocol Section 9.2.1).

Pregnancy test results will be listed only.

9.6. Vital Signs, Physical Findings, and Other Observations Related to Safety

9.6.1. Vital Signs

Vital signs will be measured after the patient has been in a supine or semi-supine position for at least 5 minutes and will include blood pressure, pulse rate, respiratory rate, and temperature.

Height will only be measured at the Screening Visit. Weight and vital signs will be measured according to the Schedule of Assessments (Appendix 15.2). Weight and height will be used to calculate the patient's body mass index at Screening.

The mean change from baseline to each scheduled assessment will be summarized descriptively by treatment group and overall for each vital sign variable specified in the protocol.

Baseline will be defined as the last vital sign value obtained before the first dose of study drug on Day 1; if Day 1 values are unavailable, then values obtained at the Screening Visit will be used.

The number and percentage of patients with at least 1 PCS vital sign value, defined as for laboratory values in Section 9.5, will be summarized descriptively. A listing of vital signs by patient will also be provided.

9.6.2. Physical Examinations

A complete physical examination will consist of an examination of all major organ systems to include, but not be limited to, chest auscultation, abdominal auscultation and palpation, head, eyes, ears, nose and throat. An assessment for active signs and symptoms of infection including tuberculosis and skin examinations for non-melanoma skin cancers will be performed. Evaluation for progressive multifocal leukoencephalopathy (PML) symptoms such as facial droop, general weakness, clumsiness, trouble speaking, personality changes, memory problems, and vision changes will be assessed.

Additionally, an assessment for the presence or absence of nasal hair will be performed during complete physical exams. Complete physical exams will be performed at screening, baseline, Week 24, and Week 28. The presence or absence of nasal hair will be summarized as counts and percentages by treatment group and visit for the Safety Population. The screening visit will not be summarized, but will be listed only.

Brief physical examinations will include assessment for active signs and symptoms of infection, including tuberculosis, skin examinations for non-melanoma skin, and an evaluation for PML symptoms such as facial droop, general weakness, clumsiness, trouble speaking, personality changes, memory problems, and vision changes. Brief physical examinations will be performed according to the Schedule of Events (Appendix 15.2).

A listing of the date and type of physical examinations performed will be provided. The listing will include presence or absence of nasal hair. Deteriorations from baseline on physical examination will be coded as adverse events and summarized as such.

9.6.3. Other Safety Measures

9.6.3.1. 12-Lead Electrocardiogram

Twelve-lead electrocardiograms (ECG) will be performed at the screening visit, Week 12, and end of study after the patient has rested in a supine or semi-supine position for at least 5 minutes. Individual parameters including heart rate, PR, QT, QTcF (Frederica's correction), QRS, RR intervals will be collected. If a patient is unable to attend an in-person clinic visit due to COVID-19, an ECG may not be performed. Missed assessments will be recorded in the eCRF as due to COVID-19.

The mean change from baseline to each scheduled assessment will be summarized descriptively by treatment group and overall for each ECG variable specified in the protocol.

A listing of ECG results will be provided. A listing of abnormal ECG findings will also be provided.

10. PHARMACOKINETIC EVALUATION

Not applicable for this study.

11. OTHER ANALYSES

In light of the COVID-19 pandemic and recommendations in the FDA guidance document "Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency", the following additional listings and summaries will be provided:

- Listing of patients affected by COVID-19 related study disruption, defined as missed or remote visits or assessments indicated per the eCRF.
- Listing of dosing exceptions related to COVID-19.
- Summary and listing of protocol deviations related to COVID-19, per Section 7.1.
- Summary of TEAEs related to COVID-19 by system organ class and preferred term and TEAEs related to COVID-19 leading to study drug discontinuation by system organ class and preferred term, per Section 9.3. Relationship to COVID-19 will be determined using the Standard MedDRA Query for COVID-19 (Appendix 15.4) and may also include preferred terms related to vaccinations if the collected term indicates relation to COVID-19 (ex: Vaccination complication, Vaccination site pain). Terms may be reviewed by the sponsor prior to database lock to finalize the list of preferred terms.
- Listing of TEAEs related to COVID-19.
- Listing of prior and concomitant medications related to COVID-19, where relationship to COVID-19 is defined by the question "Related to COVID-19?" on the Prior and Concomitant Medications CRF page.
- Summary of number and percentage of randomized subjects with missed visits due to COVID-19.

12. INTERIM ANALYSES AND DATA MONITORING

An independent data monitoring committee (DMC) will perform regular safety assessments based on a review of cumulative safety data. The DMC may advise study and/or treatment arm cessation due to intolerability at any time. There are no plans to perform interim efficacy analyses or to discontinue the study early due to demonstration of efficacy differences. Therefore, no adjustment of Type I error is required for interim analyses.

13. CHANGES TO THE ANALYSES PLANNED IN THE PROTOCOL

The following changes to the protocol were made:

- Protocol Section 12.4.1 Disposition and Baseline Characteristics
    - This section of the protocol does not distinguish treatment discontinuation from study discontinuation. The SAP adds detail and specifies that both treatment and study discontinuation, as well as reasons for discontinuation for both treatment and study will be summarized in line with FDA feedback and the eCRF.
- Protocol Section 12.4.2
    - Primary Efficacy Endpoint
        - The protocol uses the general umbrella term "Cochran-Mantel-Haenszel test" to describe the primary analysis. The SAP specifies "Mantel-Haenszel common risk difference" in reference to analyses which use multiple imputation. This is in line with Protocol Section 12.4.5.1. Analysis of Imputed Datasets and was done to avoid confusion with stratified CMH tests utilized for non-imputed data.
    - Other Efficacy Endpoints
        - The protocol specifies that analyses of HADS will use an MMRM. The SAP updates this analysis to use an ANCOVA as the questionnaire was only collected at one post-baseline visit.
- Protocol Section 12.4.5 – Multiple Imputation
    - The protocol states, "Missing values will be imputed with a sequential regression approach under the assumption of multivariate normality." The SAP removes the phrase "under the assumption of multivariate normality" as it is not necessarily applicable when using the FCS method.
    - The protocol states that the discriminant function will be used for classification variables. The logistic function was specified for SPRO imputation because it is ordinal rather than nominal.
- Protocol Section 12.4.5.2 – Tipping Point Analyses
    - The protocol states that "The patterns of interest for the tipping point analysis will consider two cases and will estimate tipping points for values where only missing values not at random (MNAR) are adjusted and also where all missing values are adjusted (i.e., two separate tipping point analyses)." The SAP specifies only one tipping point analysis due to the following:
        - This statement assumes that the primary imputation method would employ an MNAR assumption. It was determined that MAR would be reasonable upon blinded review of reasons for missing data, as prespecified in Protocol Section 12.4.4.
        - To test the MAR assumption, an additional sensitivity analysis employing an MNAR assumption was added.

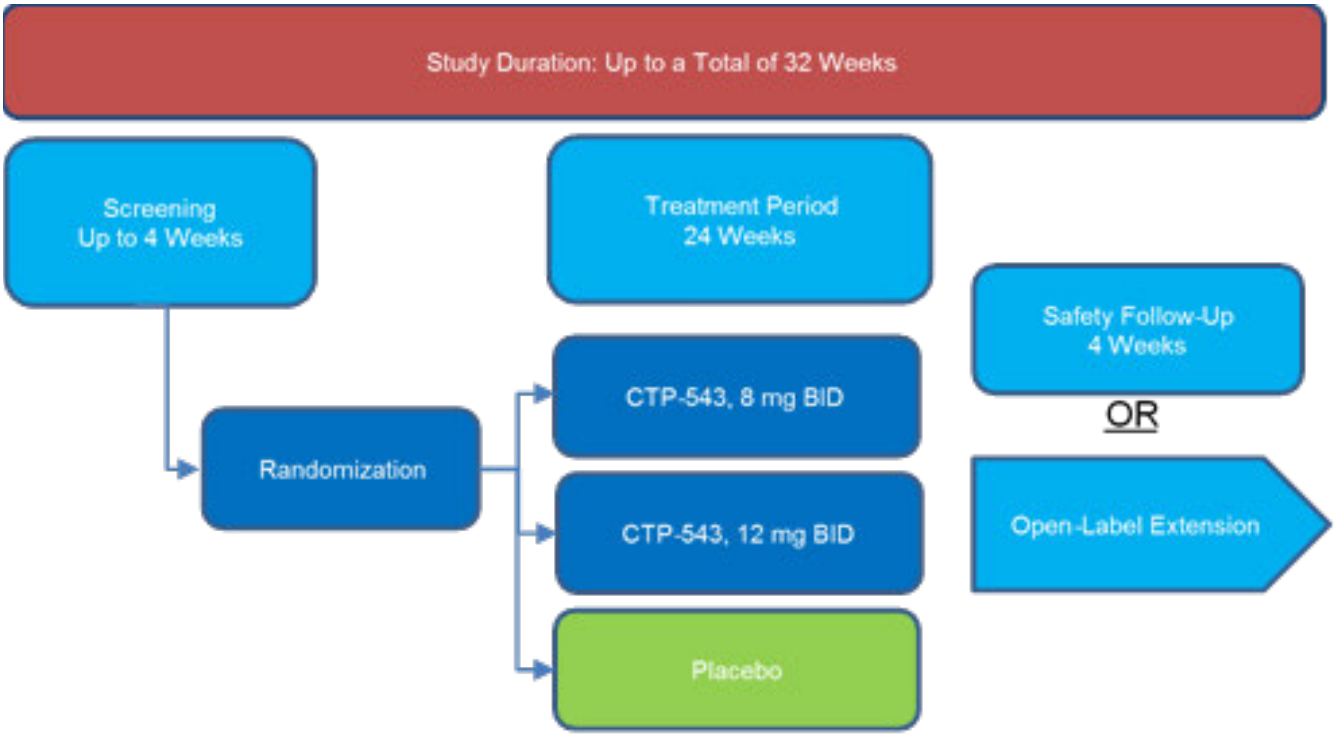
        - The values that will be MNAR for the MNAR sensitivity analysis are study discontinuations due to TEAEs or lack of efficacy. These values will be imputed using the control arm. As such, adjusting these values would shift those patients to be worse than placebo. As placebo in this study is a true placebo, this is not a realistic or informative shift.
    - The protocol states, "Missing values will not be imputed for efficacy endpoints which are not derived from SALT scores." The SAP adds multiple imputation for the SPRO endpoints because SPRO at Week 24 is an alpha-controlled key secondary endpoint and the treatment of missing data should align with the treatment of missing data for the other alpha-controlled endpoints.
- Protocol Section 12.4.7 Safety
    - Adverse Events
        - The protocol states that "Proportions for adverse events that are gender-specific (e.g., dysmenorrhea) will be based on the number of patients from that gender." The SAP specifies that the proportion of adverse events will be based on the number of patients in the safety population, even gender-specific adverse events. This change is in line with previous studies in this program and generally accepted conventions.
    - Clinical Laboratory
        - A 4th presentation of labs was added to the SAP in light of outliers seen in blinded data review throughout the study.
- Other changes:
    - Analyses for COVID-19 were added in Section 11 of the SAP.

14. REFERENCES

FDA. (2020). Guidance for Industry: Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency.

15. APPENDIX

15.1. Study Flow Chart



15.2. Schedule of Events

Event[1]	Screening	Baseline/ Randomization	Treatment Period[4]				Safety Follow-Up
	Day -28 to Day -1[2] (Visit 1)	Day 1[3] (Visit 2)	Week 2 (Visit 3)	Week 4, 8, 12 (Visit 4, 5, 6)	Week 16, 20 (Visit 7, 8)	Week 24 (Visit 9)/ ET Visit	Week 28 (Visit 10)[5]
Informed consent	X						
Eligibility assessment	X	X					
Demographics	X						
Medical history	X	X					
Randomization		X					
Complete physical examination	X	X				X	X
Brief physical examination				X	X		
Height	X						
Weight	X	X		X	X	X	X
Pregnancy test[6]	X	X		X	X	X	X
Tuberculosis test	X						
Clinical laboratory testing[7,8]	X[9]	X	X[10]	X	X	X	X
HBV and HCV test	X						
12-lead electrocardiogram	X			X[12]		X	
Vital signs	X	X		X	X	X	X
Severity of Alopecia Tool assessment[11]	X	X		X	X	X	
Photographs of the scalp	X	X		X	X	X	
Photographs of eyebrows, eyelashes, and hands		X		X[12]		X	
Eyebrow Assessment (BETA)		X		X[12]		X	
Eyelash Assessment (BELA)		X		X[12]		X	

Patient Reported Outcome Scales: SPRO and QPRO		X		X[12]	X	X	
Patient and Clinical Global Impression of Improvement (PGI-I, CGI-I[11])				X[12]	X	X	
Patient and Clinical Global Impression of Severity (PGI-S, CGI-S[11])		X		X[12]	X	X	
Hospital Anxiety and Depression Scale (HADS)		X				X	
Dispense study drug[13]		X		X	X		
Study drug accountability				X	X	X	
Adverse events[14]	X	X	X	X	X	X	X
Concomitant medications[14]	X	X	X	X	X	X	X

HBV = hepatitis B virus; HCV = hepatitis C virus

[1] Due to the COVID-19 pandemic, patient study visits may be performed remotely when the provider and patient are not able to be physically present in the clinic. Study personnel should strive to perform assessments per the schedule of events above with the understanding that all assessments may not be able to be performed (ECGs, vitals, SALT assessments).

[2] Randomization/Day 1 may occur any time after eligibility has been confirmed.

[3] All subsequent visits and week increments should be based on the date of Visit 2 (Day 1).

[4] All visit windows are ±2 days.

[5] The Safety Follow-Up Visit is intended for those patients who do not roll over into the Open-Label Extension.

[6] Serum pregnancy test only for females of childbearing potential

[7] Includes hematology, serum chemistry, and lipids (fasted) performed by the central lab.

[8] Collected pre-dose.

[9] Will include thyroid stimulating hormone and hemoglobin A1c at Screening only.

[10] Will include hematology and serum chemistry **only**.

[11] Should be performed by the same rater for the patient for the duration of the study.

[12] Performed at **Week 12 only**.

[13] Due to COVID-19, Direct to Patient (DTP) shipping services may be warranted in cases where a patient is not able to attend in clinic visits or a site is closed.

[14] Collection is ongoing.

15.3. Random Seeds for Multiple Imputation

The following list of number will be used in order for programming tasks that require random seeds. If additional numbers are required, programming will return to the start of the list and add one to each value for additional seeds.

14164709
31561197
44209199
12317451
37085322
38619061
57110334
35359204
87269781
20314426
73337873
22219275
57685038
66636300
12282889
60913778

15.4. Standardized MedDRA Query Values for COVID-19

Scope	Preferred Term	Preferred Term Code
Narrow	Asymptomatic COVID-19	10084459
Narrow	Coronavirus infection	10051905
Narrow	Coronavirus test positive	10070255
Narrow	COVID-19	10084268
Narrow	COVID-19 immunisation	10084457
Narrow	COVID-19 pneumonia	10084380
Narrow	COVID-19 prophylaxis	10084458
Narrow	COVID-19 treatment	10084460
Narrow	Exposure to SARS-CoV-2	10084456
Narrow	Multisystem inflammatory syndrome in children	10084767
Narrow	Occupational exposure to SARS-CoV-2	10084394
Narrow	SARS-CoV-2 antibody test positive	10084491
Narrow	SARS-CoV-2 carrier	10084461
Narrow	SARS-CoV-2 sepsis	10084639
Narrow	SARS-CoV-2 test false negative	10084480
Narrow	SARS-CoV-2 test positive	10084271
Narrow	SARS-CoV-2 viraemia	10084640
Narrow	Suspected COVID-19	10084451
Broad	Antiviral prophylaxis	10049087
Broad	Antiviral treatment	10068724
Broad	Coronavirus test	10084353
Broad	Coronavirus test negative	10084269
Broad	Exposure to communicable disease	10049711
Broad	Pneumonia viral	10035737
Broad	SARS-CoV-2 antibody test	10084501
Broad	SARS-CoV-2 antibody test negative	10084509
Broad	SARS-CoV-2 test	10084354
Broad	SARS-CoV-2 test false positive	10084602
Broad	SARS-CoV-2 test negative	10084273

16. ATTACHMENTS