Official Title: Topical Ruxolitinib ,Evaluation in A,topic Dermatitis Study 1/2 (TRuE-AD 1/2)
A Phase 3, Double-Blind, Randomized, 8-Week, Vehicle-Controlled Efficacy and Safety Study of Ruxolitinib Cream Followed by a Long-Term Safety Extension Period in Adolescents and Adults With Atopic Dermatitis

NCT Number: NCT03745638/NCT03745651

Document Date: 08NOV2019

STATISTICAL ANALYSIS PLAN

Topical Ruxolitinib Evaluation in Atopic Dermatitis Study 1/2 (TRuE-AD1/2)

A Phase 3, Double-Blind, Randomized, 8-Week, Vehicle-Controlled Efficacy and Safety Study of Ruxolitinib Cream Followed by a Long-Term Safety Extension Period in Adolescents and Adults With Atopic Dermatitis

IND Number	[redacted]
Sponsor:	**Incyte Corporation** **1801 Augustine Cut-Off** **Wilmington, DE 19803**
Protocol Version:	**Protocol Amendment 2 dated 13 FEB 2019**
CRF Approval Date:	**05 MAR 2019**
SAP Version:	**Original**
SAP Author:	[redacted]
Date of Plan:	**08 NOV 2019**

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Term
AD	atopic dermatitis
AE	adverse event
ANOVA	analysis of variance
BID	twice daily
BMI	Body Mass Index
BSA	body surface area
CDLQI	Children Dermatology Life Quality Index
CI	confidence interval
CRF	case report form
CTCAE	Common Terminology Criteria for Adverse Events
DBP	diastolic blood pressure
DLQI	Dermatology Life Quality Index
EASI	Eczema Area and Severity Index
EASI50	≥ 50% improvement in EASI score
EASI75	≥ 75% improvement in EASI score
EASI90	≥ 90% improvement in EASI score
eCRF	electronic case report form
EMA	European Medicines Agency
EQ-5D-5L	EQ-5D is a validated, self-administered, generic, utility questionnaire wherein participants will rate their current health state based on the following criteria: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.
FDA	Food and Drug Administration
flare	an episode of worsening of atopic dermatitis, such that it requires re-initiation or escalation of treatment (IGA of at least 2)
IGA	Investigator's Global Assessment
IGA-TS	Investigator's Global Assessment Treatment Success (IGA score of 0 or 1 with ≥ 2 grade improvement from baseline)
ITT	intent-to-treat
KM	Kaplan-Meier
LTS	long-term safety
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed-model repeated measures
NCI	National Cancer Institute
NRI	nonresponder imputation
NRS	numerical rating scale

Abbreviation	Term
PD	pharmacodynamic
PGIC	Patient Global Impression of Change
PK	pharmacokinetics
POEM	Patient-Oriented Eczema Measure
PP	per Protocol
PROMIS®	Patient-Reported Outcomes Measurement Information System
PT	preferred term
SAP	statistical analysis plan
SBP	systolic blood pressure
SCORAD	SCORing Atopic Dermatitis – tool used to assess extent and severity (intensity) of eczema
SOC	system organ class
TEAE	treatment-emergent adverse event
[redacted]	[redacted]
VAS	Visual Analogue Scale
VC	vehicle-control
WHO	World Health Organization
WPAI:SHP v2.0	Work Productivity and Activity Impairment Questionnaire: Specific Health Problem Version 2.0

1. INTRODUCTION

INCB 18424-303 is a randomized, vehicle-controlled study in adolescent and adult (≥ 12 years old) participants with AD eligible for topical therapy. Approximately 600 participants will be randomized 2:2:1 to ruxolitinib 0.75% cream BID, ruxolitinib 1.5% cream BID, or vehicle cream. In addition, approximately 20% of the overall study population will consist of adolescents. Participants with baseline IGA score of 2 will constitute up to approximately 25% of the overall study population. Participants with AD involvement of 3% to 20% BSA and IGA score of 2 to 3 will receive blinded study treatment for 8 weeks followed by a randomized 44-week LTS extension period. In the LTS period, participants initially randomized to vehicle will receive either ruxolitinib 0.75% or 1.5% cream BID.

The purpose of this SAP is to provide details of the statistical analyses that have been outlined in the study Protocol. The scope of this plan will be executed by the Department of Biostatistics or designee, and the analyses of PK will be executed by the Department of Clinical Pharmacokinetics or designee.

Another Phase 3 study, INCB 18424-304, has the identical study design as INCB 18424-303. This SAP applies to both INCB 18424-303 and INCB 18424-304 studies.

2. STUDY INFORMATION, OBJECTIVES, AND ENDPOINTS

2.1. Protocol and Case Report Form Version

This SAP is based on INCB 18424-303 and -304 Protocol Amendment 2 dated 13 FEB 2019 and CRFs approved 05 MAR 2019. Unless superseded by an amendment, this SAP will be effective for all subsequent Protocol amendments and CRF versions.

2.2. Objectives and Endpoints

Table 1 presents the objectives and endpoints.

Table 1: Objectives and Endpoints

Objectives	Endpoints
Primary	
To establish the efficacy of ruxolitinib cream in participants with AD.	• Proportion of participants achieving IGA-TS at Week 8.[a]
Key Secondary	
To further assess the treatment effects of ruxolitinib cream.	• Proportion of participants who achieve EASI75 at Week 8.[a] • Proportion of participants with a ≥ 4-point improvement in Itch NRS score from baseline to Week 8. • Proportion of participants with a clinically meaningful (≥ 6-point) improvement in the PROMIS Short Form – Sleep Disturbance (8b – 24-hour recall) score at Week 8. • Proportion of participants with a clinically meaningful (≥ 6-point) improvement in the PROMIS Short Form – Sleep-Related Impairment (8a – 24-hour recall) score at Week 8.
Secondary	
To evaluate the safety and tolerability of ruxolitinib cream.	The frequency and severity of AEs; physical examinations; vital signs; and laboratory data for hematology, serum chemistry, and urinalysis.
To further evaluate efficacy of ruxolitinib cream.	• Proportion of participants achieving an IGA-TS at Weeks 2 and 4. • Proportion of participants achieving an IGA of 0 or 1 at each visit. • Proportion of participants with a ≥ 4-point improvement in Itch NRS score from baseline to Weeks 2 and 4. • Proportion of participants who achieve EASI50 at each visit during the VC period. • Proportion of participants who achieve EASI75 at Weeks 2 and 4. • Proportion of participants who achieve EASI90 at each visit during the VC period. • Mean percentage change from baseline in EASI score at each visit during the VC period. • Mean percentage change from baseline in SCORAD score at each visit during the VC period. • Change from baseline in Itch NRS score at each visit during the VC period. • Time to achieve Itch NRS score improvement of at least 2, 3, or 4 points. • Change from baseline in Skin Pain NRS score at each visit during the VC period.

Table 1: Objectives and Endpoints (Continued)

Objectives	Endpoints
To further evaluate efficacy of ruxolitinib cream.	• Proportion of participants with a clinically meaningful (≥ 6-point) improvement in the PROMIS Short Form – Sleep-Related Impairment (8a) and Sleep Disturbance (8b) 24-hour recall score at Weeks 2 and 4. • Change from baseline in PROMIS Short Form – Sleep-Related Impairment (8a) 24-hour recall and Short Form – Sleep Disturbance (8b) 24-hour recall score at Weeks 2, 4, and 8. • PROMIS Short Form – Sleep-Related Impairment (8a) 7-day recall and Short Form – Sleep Disturbance (8b) 7-day recall score at Weeks 8, 12, 24, and 52. • Change from baseline in AD afflicted %BSA at every visit.
To evaluate the participants' Quality of Life and other patient-reported outcomes.	• Change from baseline in POEM score at each visit. • Change from baseline in DLQI score at Weeks 2, 4, 8, 12, 24, and 52 and at unscheduled visits. • Mean PGIC score at Weeks 2, 4, and 8. • Proportion of participants with each score on the PGIC at Weeks 2, 4, and 8. • Proportion of participants with a score of either 1 or 2 on the PGIC at Weeks 2, 4, and 8. • Change from baseline in EQ-5D-5L score during the VC period. • Change from baseline in WPAI:SHP v2.0 at Weeks 2, 4, 8, 12, 24, 36, and 52.
To evaluate the PK of ruxolitinib cream in plasma.	• Trough plasma concentrations of ruxolitinib at all study visits.
[redacted]	

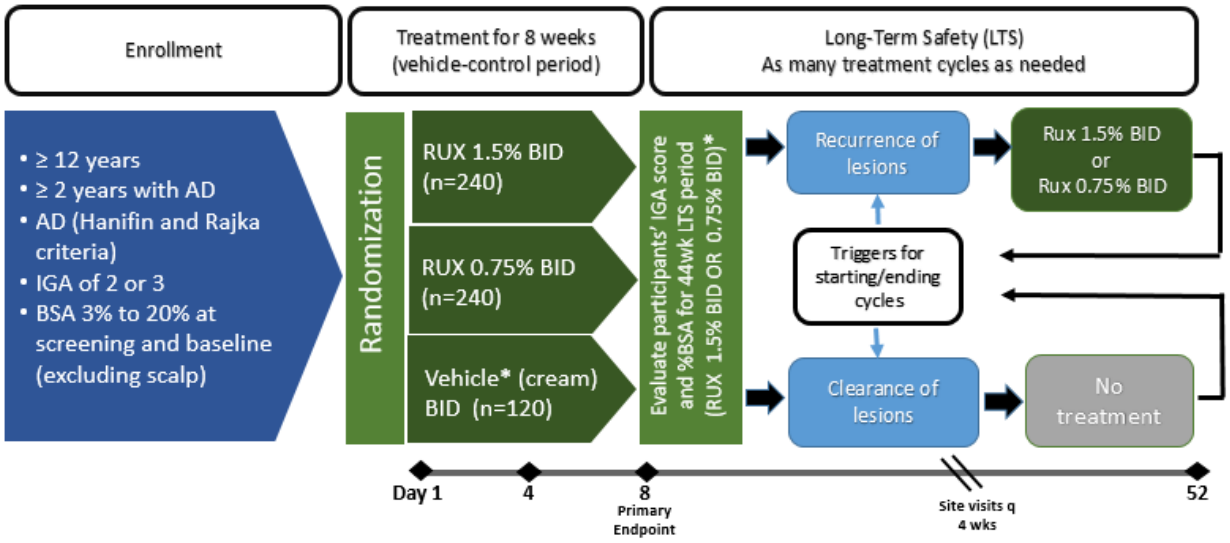
For an EMA submission, if applicable, the primary endpoint is changed to the proportion of participants who achieve EASI75 at Week 8, and the first key secondary endpoint is changed to the proportion of participants achieving IGA-TS at Week 8.

3. STUDY DESIGN

This is a randomized, vehicle-controlled study in adolescent and adult (≥ 12 years old) participants with AD eligible for topical therapy. Approximately 600 participants will be randomized 2:2:1 to ruxolitinib 0.75% cream BID, ruxolitinib 1.5% cream BID, or vehicle cream. In addition, approximately 20% of the overall study population will consist of adolescents. Participants with baseline IGA score of 2 will constitute up to approximately 25% of the overall study population. Participants with AD involvement of 3% to 20% BSA and IGA score of 2 to 3 will receive blinded study treatment for 8 weeks followed by a randomized, double-blinded, 44-week LTS extension period. In the LTS period, participants initially randomized to vehicle will receive either ruxolitinib 0.75% or 1.5% cream BID.

The study schema is shown below in Figure 1. All participants will have follow-up assessments 30 (+ 7) days after the last application of study drug.

Figure 1: Study Design Schema



At week 8 (LTS baseline), participants initially on vehicle will be randomized to either Rux 1.5% or 0.75% cream BID. Additionally, to qualify for the LTS period, the IGA score of 0 to 4 and %BSA of 0% to 20% will be required.

3.1. Randomization

In the treatment period, approximately 600 participants will be randomized 2:2:1 to ruxolitinib 0.75% cream BID, ruxolitinib 1.5% cream BID, or vehicle cream. Participants will be stratified at randomization based on baseline IGA (2, 3) and region (North America or other).

3.2. Control of Type I Error

For the primary and key secondary endpoints, the overall 2-sided Type I error is 0.05.

A graphical procedure with gatekeeping testing strategy for the primary and key secondary analyses, as specified in Protocol Section 10.3, will be implemented to control the overall Type I error. The underlying procedure is derived using the methodology developed in Bretz et al (2009). The method will guarantee a strong control of the family-wise error rate.

In Step 1, two families of 8 elementary hypotheses tests are grouped according to treatment comparison between each ruxolitinib cream group to vehicle group, where

- Family 1 (1.5% BID vs vehicle):
  - H11: proportion of participants who achieve IGA-TS;
  - H12: proportion of participants who achieve EASI-75;
  - H13: proportion of participants with a ≥ 4-grade improvement in Itch NRS over baseline;
  - H14: proportion of participants with a ≥ 6-point improvement in the PROMIS Sleep Disturbance score.
- Family 2 (0.75% BID vs vehicle):
  - H21: proportion of participants who achieve IGA-TS;
  - H22: proportion of participants who achieve EASI-75;
  - H23: proportion of participants with a ≥ 4-grade improvement in Itch NRS over baseline;
  - H24: proportion of participants with a ≥ 6-point improvement in the PROMIS Sleep Disturbance score.

Step 2 has one family of 2 hypotheses tests:

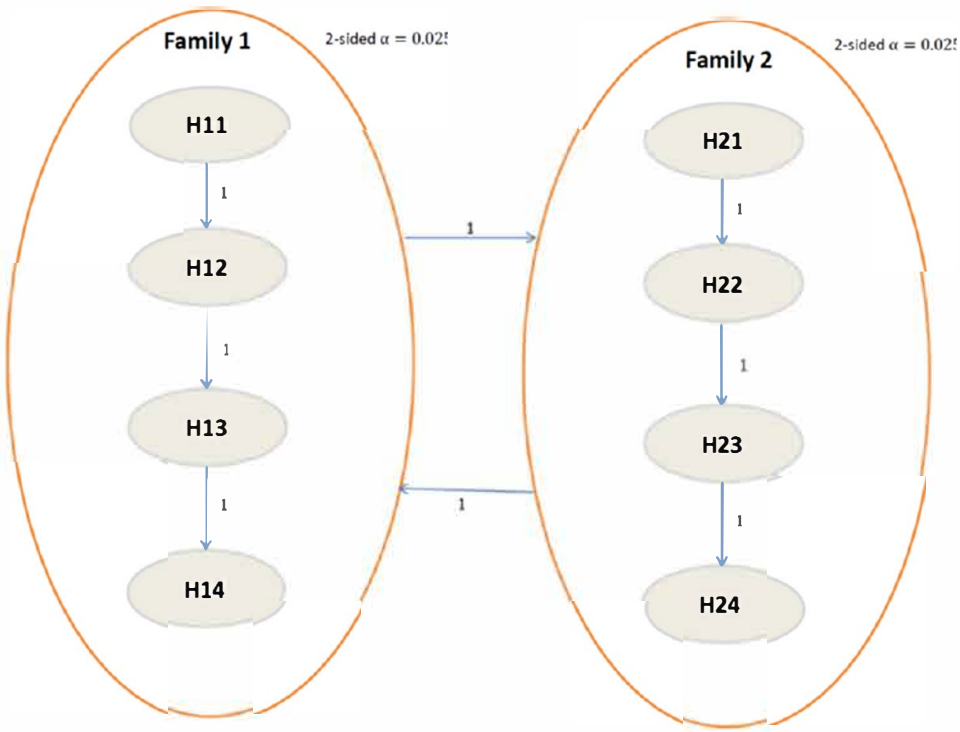
- Family 3 (1.5% BID vs vehicle and 0.75% BID vs vehicle on PROMIS Sleep-Related Impairment score):
  - H31: Proportion of participants with a clinically meaningful improvement (≥ 6-point change over baseline) in Sleep-Related Impairment (8a) score between 1.5% BID vs vehicle;
  - H32: Proportion of participants with a clinically meaningful improvement (≥ 6-point change over baseline) in Sleep-Related Impairment (8a) score between 0.75% BID vs vehicle.

To control the overall Type I error rate, 2-sided $\alpha = 0.05$, the Bonferroni's method will be used. In Step 1, within Family 1 and 2, the endpoints are tested in a fixed sequence at Bonferronized 2-sided $\alpha = 0.025$ level. The endpoint will be tested only if the associated primary (and secondary in previous steps) are rejected. For any dose level, if the 4 related null hypotheses can be rejected, then the fixed sequence for the other dose level can be conducted at 2-sided $\alpha = 0.05$ level. If all null hypotheses in Family 1 and 2 are rejected, in Step 2, H31 and H32 will be tested using Bonferroni- Hochberg's procedure with overall 2-sided $\alpha = 0.05$ level. The approach is visualized in Figure 2.

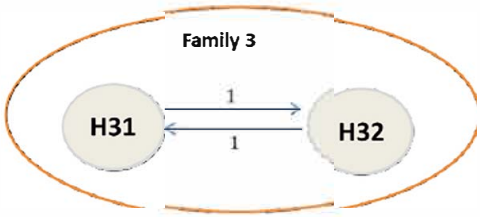
For an EMA submission, if applicable, the primary endpoint is changed to the proportion of participants who achieve EASI75 at Week 8, and the first key secondary endpoint is changed to the proportion of participants achieving IGA-TS at Week 8, and the hypotheses are changed accordingly.

Figure 2: Control of the Overall Type I Error Rate

Step 1:



Step 2:



3.3. Sample Size Considerations

Approximately 600 participants will be randomized 2:2:1 to ruxolitinib 0.75% BID, 1.5% BID, or vehicle and stratified by baseline IGA (2, 3) and region (North America or Europe). The sample size calculation is based on the Fisher exact test for the statistical comparison on the primary efficacy endpoint. The Fisher exact test is used to provide a conservative evaluation of statistical power, and it is accurate when there is a small expected number of responders in the vehicle group. Based on the results from a Phase 2, randomized, dose-ranging study (INCB 18424-206), the IGA-TS at Week 8 is assumed to be 45% and 30% for active arms (1.5% BID and 0.75% BID, respectively) versus 10% for placebo. Using a 2-sided alpha of 0.025, the sample size will have > 95% power to detect a difference between each of the 2 active treatment groups versus vehicle. In addition to provide adequate power for efficacy variables, the sample size is determined to provide an adequate database for safety evaluation.

3.4. Schedule of Assessments

Refer to Protocol Amendment 2 dated 13 FEB 2019 for a full description of all study procedures and assessment schedules (Protocol Tables 3 and 4) for this study.

4. DATA HANDLING DEFINITIONS AND CONVENTIONS

4.1. Scheduled Study Evaluations and Study Periods

4.1.1. Day 1

Day 1 is the date that the first application of ruxolitinib cream or vehicle cream is administered to the participants in the specific period.

For randomized participants not treated with any study drug, Day 1 is defined as the day of randomization.

4.1.2. Study Day

If a visit/reporting date is on or after Day 1, then the study day at the visit/reporting date will be calculated as

Day # = (Visit/Reporting Date – Day 1 date + 1).

If the visit/reporting date is before Day 1, then the study day at the visit/reporting date will be calculated as

Day # = (Visit/Reporting Date – Day 1 date).

A study day of -1 indicates 1 day before Day 1.

4.1.3. Baseline Value

Baseline is the last nonmissing measurement obtained before or on the day of first application of ruxolitinib cream or vehicle cream.

For participants who cross-over treatment in LTS period, baseline is the last nonmissing measurement obtained before or on the day of first application of study treatment in LTS period.

For randomized participants not treated with any study drug, baseline is defined as the last nonmissing assessment before or on the day of randomization for all parameters.

When scheduled assessments and unscheduled assessments occur on the same day and time of the assessment or time of first application is not available, use the following convention to determine baseline:

- If both a scheduled and an unscheduled visit are available on the day of the first application and the time is missing, use the scheduled assessment as baseline.
- If all scheduled assessments are missing on the day of the first application and an unscheduled assessment is available, use the unscheduled assessment as baseline.

4.1.4. Last Available Value

The last available value is the last nonmissing measurement obtained after starting ruxolitinib cream or vehicle cream and within 30 days after the last application of ruxolitinib cream or vehicle cream, or before the first application of ruxolitinib cream in the next period, whichever is earlier.

4.1.5. Handling of Missing and Incomplete Data

For response endpoints, all nonresponders, as well as all participants who are missing postbaseline values, will be defined as nonresponders for the NRI analysis.

For continuous endpoints, any participant who is missing postbaseline values may have missing data handled using MMRM under the missing-at-random assumption. MMRM model implicitly adjusts for missing data through a variance-covariance structure.

For other endpoints, missing observations will be handled as detailed in the specific sections addressing each analysis.

4.2. Variable Definitions

4.2.1. Body Mass Index

Body mass index (BMI) will be calculated as follows:

$$\text{BMI } (kg/m^2) = [\text{weight } (kg)] / [\text{height } (m)]^2$$

4.2.2. Prior and Concomitant Medication

Prior medication is defined as any nonstudy medication started before the first application of study treatment.

Concomitant medication is defined as any nonstudy medication that is started accordingly:

- Before the date of first application of study treatment and is ongoing throughout the study or ends on/after the date of first application of study treatment.
- On/after the date of first application of study treatment and is ongoing or ends during the course of study treatment.

A prior medication could also be classified as "both prior and concomitant medication" if the end date is on or after the first application of study treatment. In the listing, it will be indicated whether a medication is prior-only, concomitant-only, or both prior and concomitant.

For the purposes of analysis, all medications will be considered concomitant medications unless the medications can unequivocally be defined as not concomitant.

5. STATISTICAL METHODOLOGY

5.1. General Methodology

Unless otherwise noted, SAS® software (SAS Institute Inc, Cary, NC; Version 9.1 or later) will be used for the generation of all tables, graphs, and statistical analyses. Descriptive summaries for continuous variables will include, but not be limited to, the number of observations, mean, standard deviation, median, minimum, maximum, 1st quartile, 3rd quartile and 95% CI. Descriptive summaries for categorical variables will include the number and percentage of participants in each category.

5.2. Treatment Groups

This is a randomized, double-blinded vehicle-controlled study followed by a LTS extension period. Data will be summarized based on treatment regimen that was assigned (ITT) or that the participant actually applied (safety).

During the VC period, the treatment groups will be 1.5% BID, 0.75% BID, and Vehicle.

During the LTS period, the treatment groups will be 1.5% BID, 0.75% BID, Vehicle to 1.5% BID, and Vehicle to 0.75% BID.

5.3. Analysis Populations

5.3.1. Intent-to-Treat Population

All participants who are randomized to the study constitute the ITT population. Treatment groups for this population will be defined according to the treatment assignment at the time of randomization regardless of the actual study treatment the participant might apply during his/her participation in the study.

The ITT population will be used for the summary of demographics, baseline characteristics, participant disposition, and analyses of all efficacy data.

5.3.2. Per Protocol Population

Participants in the ITT population who are considered to be sufficiently compliant with the Protocol compose the PP population, which is defined for supportive sensitivity analyses for efficacy endpoints in the treatment period. Participants with important protocol deviations, as defined in Section 6.3, will be excluded from the PP population.

5.3.3. Safety Population

All randomized participants who applied at least 1 application of ruxolitinib cream or vehicle cream will constitute the safety population. Treatment groups for this population will be determined according to the actual treatment the participant applied on Day 1 regardless of assigned study treatment.

All safety analyses will be conducted using the safety population.

5.3.4. Pharmacokinetic/Pharmacodynamic Evaluable Population

The PK/PD evaluable population will include all participants who applied at least 1 application of ruxolitinib cream and provided at least 1 postdose plasma sample (1 PK/PD measurement) that complies with the instruction in the Protocol. Participants in whom it is not possible to obtain a blood sample for PK/PD from an area of the body that was not treated with (exposed to) study drug (eg, due to difficulties of technical/ procedural nature) will not be included in the PK/PD evaluable population because of the material risk of the contamination of such a sample with study drug from the skin through which a needle is passed. The study pharmacokineticist will review data listings of study drug application and sample records to identify participants to be excluded from PK data analyses. The study research investigator will review data listings of PD data and sample records to identify participants to be excluded from analyses of PD data.

6. BASELINE, EXPOSURE, AND DISPOSITION VARIABLES AND ANALYSES

Appendix A provides a list of planned tables, figures, and listings. Sample data displays are included in a separate document.

6.1. Baseline and Demographics, Physical Characteristics, and Disease History

6.1.1. Demographics

The following demographics will be summarized for the ITT population during the VC period: age, age group, sex, race, ethnicity, region, weight, height, and BMI.

6.1.2. Baseline Disease Characteristics

Baseline disease characteristics summarized for the ITT population during the VC period include but are not limited to the following:

- Time since first onset of AD
- Prior history of asthma (No/Yes)
- Prior allergies (food, environmental) (No/Yes)
- History of contact dermatitis (No/Yes)
- Common complications of AD
- Prior treatments for AD
- Time since onset of current AD
- Prior therapy for AD given in the past 30 days (No/Yes)
- Total %BSA involvement in current AD episode
- Facial involvement (No/Yes)
- Number of AD episodes/flare-ups over the last 12 months

- Baseline IGA score (2, 3)
- Baseline EASI score

6.1.3. Medical History

For participants in the ITT population during the VC period, medical history will be summarized by assigned treatment groups. This summary will include the number and percentage of participants with medical history for each body system/organ class as documented in the eCRF.

6.2. Disposition of Participants

The number and percentage of participants who were randomized, treated, and completed the VC period; were treated in the LTS period; completed the LTS period; discontinued the treatment or withdrew from the study during the VC period; and discontinued treatment or withdrew from the study during the LTS period with a primary reason for discontinuation will be summarized for the ITT population during the VC and LTS periods.

6.3. Protocol Deviations

In general, the following are important protocol deviations that may significantly affect the primary and gated secondary analyses:

- Missing data for the primary endpoint;
- Overall application compliance less than 60%;
- Key inclusion/exclusion criteria deviations;
- Use of restricted or prohibited medications.

Participants with one or more such deviations will be excluded from the PP population.

6.4. Exposure

For participants in the safety population during the VC and LTS periods, descriptive statistics will be provided by treatment group for duration of treatment, average daily dose, and total dose. Duration of treatment with ruxolitinib cream or vehicle cream is defined as the number of days from Day 1 to the last record of ruxolitinib cream or vehicle cream application in the specific period.

6.5. Study Drug Compliance

Overall compliance (%) for the application of ruxolitinib cream or vehicle cream during the VC period will be calculated for all participants in the safety population as follows:

Overall application compliance (%) = 100 × [total number of nonmissing applications] / [total number of intended applications]

where the total number of nonmissing applications is the total number of applications that the participant actually applied during the study.

6.6. Prior and Concomitant Medication

For participants in the ITT population during the VC period, prior medications and concomitant medications will be coded using the WHO Drug Dictionary and summarized as number and percentage of participants with prior and concomitant medications by WHO drug class and WHO drug term. For participants in the LTS period, only concomitant medications will be summarized.

Prior medications for AD will be summarized by treatment group as well as listed.

7. EFFICACY

Appendix A provides a list of planned tables, figures, and listings. Sample data displays are included in a separate document.

7.1. General Considerations

For all continuous variables, both the actual value and change and/or percentage from baseline (if available) will be analyzed.

All by-visit analyses will include the follow-up period if the data are available.

7.2. Analysis of the Primary Efficacy Parameters

For FDA submission, the primary endpoint is IGA-TS at Week 8. For EMA submission, if applicable, the primary endpoint is EASI75 at Week 8.

7.2.1. Primary Efficacy Measures

7.2.1.1. Investigator's Global Assessment

The IGA is an overall eczema severity rating on a 0 to 4 scale that will be assessed during site visits. The grades for the IGA are shown in Table 2.

Table 2: Investigator's Global Assessment

Grade	Severity	Status
0	Clear	No erythema or induration/papulation, no oozing/crusting; there may be minor residual discoloration.
1	Almost clear	There may be trace faint pink erythema, with almost no induration/papulation, and no oozing/crusting.
2	Mild	There may be faint pink erythema, with mild induration/papulation and no oozing/crusting.
3	Moderate	There may be pink-red erythema with moderate induration/papulation and there may be some oozing/crusting.
4	Severe	There may be deep or bright red erythema with severe induration/papulation and with oozing/crusting.

Source: University of Nottingham.

The IGA-TS is defined as an IGA score of 0 or 1 with ≥ 2 grade improvement from baseline.

7.2.1.2. Eczema Area and Severity Index Score

Atopic dermatitis will be assessed using EASI scoring system, which is a validated disease measurement for clinical studies that examines 4 areas of the body and weights them for participants of at least 8 years of age. Each of the 4 body regions is assessed separately for erythema (E), induration/papulation/edema (I), excoriations (Ex), and lichenification (l) for an average degree of severity of each sign in each region.

The disease severity strata for the EASI are as follows: 0 = clear; 0.1 to 1.0 = almost clear; 1.1 to 7.0 = mild; 7.1 to 21.0 = moderate; 21.1 to 50.0 = severe; 50.1 to 72.0 = very severe.

The categorical variable EASI75 will be equal to 1 for percentage improvement from baseline in EASI score of 75% or greater and will be equal to 0 for percentage improvement of less than 75%. EASI50 are EASI90 are defined with the same pattern.

7.2.2. Primary Efficacy Analyses

The primary analysis will be based on the ITT population in the VC period. The primary alternative hypothesis (superiority of ruxolitinib 1.5% BID or 0.75% BID compared with vehicle) will be tested using logistic regression. This model will include the treatment group (1.5% BID, 0.75% BID, and vehicle) and stratification factors (baseline IGA and region). The unadjusted p-values between each treatment group versus vehicle will be calculated based on Wald test, which will be compared with the procedure defined in Section 3.2. Exact logistic regression (Mehta and Patel 1995) will be used for all of the comparisons if any of the dose levels have an expected cell count less than 5.

Odds ratio and 95% CI in response rates (ruxolitinib cream vs vehicle) at Week 8 will also be computed. All nonresponders in the VC treatment period, as well as all participants who are missing postbaseline values, will be defined as nonresponders for the nonresponder imputation analysis.

The primary outcome will also be examined for PP population using the same model as the primary analysis.

7.2.3. Subgroup Analyses for Primary Endpoints

Subgroups will be formed based on the following participant characteristics and baseline variables for those participants whose data are available:

- Baseline IGA score (2, 3)
- Baseline EASI score (≤ 7, >7)
- Region (North America, Europe)
- Categorical age
- Sex
- Race

7.2.4. Sensitivity and Supportive Analyses for Primary Endpoints

7.2.4.1. Longitudinal Logistic Regression With Repeated Measures

To adjust for the dependence underlying the hierarchical multilevel data structure (visit, participant, and site), a longitudinal logistic regression with repeated measures will be applied.

The 3 level structures in the model are:

- Level 1: visit;
- Level 2: participant; and
- Level 3: site,

where visits are nested within participants, which are further nested within sites.

The binary response (IGA-TS or EASI75) of each participant at Week 2, Week 4, and Week 8 will be included as the dependent variable. Treatment (1.5% BID, 0.75% BID, and vehicle BID), the randomization stratification factors (baseline IGA and region), visit, and treatment by visit interaction will be included as fixed effects. Site level intercept and participant nested in site level intercept will be included as random effects. The within-participant and within-site errors will be modeled by an unstructured variance-covariance matrix. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom for this model.

7.2.4.2. Multiple Imputation

Multiple imputation with missing at random assumption will be used as an alternative method to handle missing data. The imputation method of choice depends on the type of missing data pattern:

- For data sets with monotone missing patterns, the variables with missing values will be imputed sequentially with covariates constructed from their corresponding sets of preceding variables. With IGA score (0-4) or EASI (0-72), a monotone method (Rubin 1987) with treatment group and stratification factors as predictors will be used.
- For data sets with arbitrary missing patterns, a fully conditional specification method (van Buuren 2007) that assumes the existence of a joint distribution for all variables will be used to impute IGA or EASI score.

After the missing values have been imputed, the binary variable IGA-TS or EASI75 response will be derived based on the definition specified in Section 7.2.1. The imputation will be repeated a number of times to generate corresponding complete data sets, in order to reflect the uncertainty around the true value. The proposed logistic regression described in Section 7.2.2 will be applied to each imputed datasets, and then the results will be combined for the inference.

7.2.4.3. Last Observation Carry Forward

For the participants who are missing postbaseline values, the last observed nonmissing value will be used to fill in missing values at Week 8. Then the proposed logistic regression described in Section 7.2.2 will applied to the imputed dataset.

7.2.4.4. Tipping Point Analysis

A tipping point sensitivity analysis will be conducted to examine the potential effects of missing data. The missing IGA-TS or EASI75 response at Week 8 in each treatment group will be replaced by a range of values to see how far we must change them for the results of the study to tip from significant to not. Between-treatment comparisons will be performed using a chi-square test.

7.3. Analysis of the Key Secondary Efficacy Parameters

7.3.1. Key Secondary Efficacy Measures

7.3.1.1. Itch Numerical Rating Scale Score

The Itch NRS is a once-per-24 hours ("daily") patient-reported measure of itch intensity assessed using an 11-point scale (0 = no itch to 10 = worst imaginable itch). Participants will be asked to record their highest (worst) level of itch in the evening over the 24-hour recall period.

The Itch NRS score for baseline will be determined by averaging the 7 daily NRS scores before Day 1 (Day –7 to Day –1). The by-visit Itch NRS score for postbaseline visits will be determined by averaging the 7 daily NRS scores before the visit day. If 4 or more daily scores are missing (out of the 7), the Itch NRS score at the visit will be set to missing.

The proportion of participants with a clinically relevant change in itch – defined as a ≥ 4-point improvement in Itch NRS score from baseline to Week 8 – will be summarized by treatment groups for participants with baseline Itch NRS ≥ 4.

7.3.1.2. PROMIS Short Form – Sleep Disturbance (8b)

The PROMIS Short Form – Sleep Disturbance (8b) is an 8-item questionnaire. Each item on the measure is rated on a 5-point scale with a range in score from 8 to 40 with higher scores indicating greater severity of sleep disturbance. The recall period will be the past 24 hours for the VC period and the past 7 days for the LTS period. The raw scores on the 8 items should be summed to obtain a total raw score. If 4 or more of the total items on the measure are missing for a given day/assessment, the scores should not be used.

For total raw score, the baseline will be determined by averaging the 7 daily scores on and before Day 1 (Day –6 to Day 1). The by-visit score for post-baseline visits will be determined by averaging the 7 daily scores before the visit day. If 4 or more daily scores are missing (out of the 7), the by-visit score at the visit will be set to missing.

Clinically meaningful difference is defined as a ≥ 6-point change over baseline. The total raw score will be summarized by visit and by treatment. The proportion of participants with a ≥ 6-point improvement in the PROMIS Short Form – Sleep Disturbance (8b – 24-hour recall) score will be summarized by treatment group for participants with baseline ≥ 6.

7.3.1.3. PROMIS Short Form – Sleep-Related Impairment (8a)

The PROMIS Short Form – Sleep-Related Impairment (8a) is an 8-item questionnaire. Each item on the measure is rated on a 5-point scale (1 = Not at all; 2 = A little bit; 3 = Somewhat; 4 = Quite a bit; and 5 = Very much) with a range in score from 8 to 40 with higher scores indicating greater severity of sleep-related impairment. The recall period will be the past 24 hours for the VC period and the past 7 days for the LTS period. The raw scores on the 8 items should be summed to obtain a total raw score. If 4 or more of the total items on the measure are missing, the scores should not be used.

For total raw score, the baseline will be determined by averaging the 7 daily scores before Day 1 (Day –7 to Day –1). The by-visit score for post-baseline VC visits will be determined by averaging the 7 daily scores before the visit day. If 4 or more daily scores are missing (out of the 7), the by-visit score at the visit will be set to missing.

Clinical meaningful difference is defined as a ≥ 6-point change over baseline. The total raw score will be summarized by visit and by treatment. The proportion of participants with a ≥ 6-point (clinically meaningful) improvement in the PROMIS Short Form – Sleep-Related Impairment (8a – 24-hour recall) score will be summarized by treatment group for participants with baseline ≥ 6.

7.3.2. Key Secondary Efficacy Analysis

Key secondary efficacy analyses will be conducted in the ITT population in the VC period. If the primary objective is achieved, the statistical comparisons for key secondary endpoints will be tested with the procedures specified in Figure 2. All of the key secondary endpoints will be analyzed using the similar logistic regression models as specified in the primary efficacy analysis.

7.4. Analysis of Secondary Efficacy Parameters

7.4.1. Secondary Efficacy Measures

7.4.1.1. SCORing Atopic Dermatitis

SCORAD is a tool to assess the extent and severity (ie, intensity) of eczema and will be completed before, during, and after treatment has begun to determine whether the treatment has been effective. This will be performed during all VC study visits, starting at baseline.

- To determine extent, the rule of 9 or handprint method is used to calculate the eczema affected area (A) as a percentage of the whole body.
- To determine intensity, the intensity of each of the following signs of redness, swelling, oozing/crusting, scratch marks, skin thickening (lichenification), dryness (this is assessed in an area where there is no inflammation) is assessed as follows:
  - None (0)
  - Mild (1)
  - Moderate (2)
  - Severe (3)

Intensity scores are added together to give "B" (maximum score of 18).

- Subjective symptoms, that is, itch and sleeplessness, are scored by the participant using a VAS where "0" is no itch (or no sleeplessness) and "10" is the worst imaginable itch (or sleeplessness).

  These scores are added to give "C" (maximum score of 20).

Total score gives approximate weights of 60% to intensity and 20% each to extent and subjective signs (ie, insomnia, etc) for the participant and will be calculated as follows:

Total score = A / 5 + (7 × B) / 2 + C.

7.4.1.2. Time to Itch Response

Time to Itch NRS response is defined as the number of days between the day of randomization and the day of achieving a daily Itch NRS score improvement of at least 2, 3, and 4 points.

7.4.1.3. Skin Pain Numerical Rating Scale

The Skin Pain NRS is a daily patient-reported measure (24-hour recall) of the worst level of pain intensity from 0 (no pain) to 10 (worst imaginable pain). Participants will be asked, "Rate the pain severity from your atopic dermatitis skin changes by selecting a number that best describes your worst level of pain in the past 24 hours."

The Skin Pain NRS score for baseline will be determined by averaging the 7 daily NRS scores before Day 1 (Day –7 to Day –1). The by-visit Skin Pain NRS score for postbaseline visits will be determined by averaging the 7 daily NRS scores before the visit day. If 4 or more daily scores are missing (out of the 7), the Skin Pain NRS score at the visit will be set to missing.

7.4.1.4. Body Surface Area

Total %BSA afflicted by AD will be estimated at each visit in the VC and LTS periods. Body surface area assessment will be approximated to the nearest 0.1% using the Palmar Method as guides, the palm plus 5 digits, with fingers tucked together and thumb tucked to the side (handprint), as 1% BSA and the thumb as 0.1% BSA.

7.4.1.5. Patient-Oriented Eczema Measure

The POEM is a 7-question quality-of-life assessment that asks how many days the participant has been bothered by various aspects of their skin condition during the past 7 days (Charman et al 2004).

Each of the 7 questions carries equal weight and is scored from 0 to 4 as follows:

- No days = 0
- 1 to 2 days = 1
- 3 to 4 days = 2
- 5 to 6 days = 3
- Every day = 4

The POEM is calculated by summing the score of each question resulting in a maximum of 28 and a minimum of 0. The meaning of POEM scores can be categorized as follows:

- 0 to 2 = Clear or almost clear
- 3 to 7 = Mild eczema
- 8 to 16 = Moderate eczema
- 17 to 24 = Severe eczema
- 25 to 28 = Very severe eczema

If 1 question is left unanswered, this is scored 0, and the scores are summed and expressed as usual out of a maximum of 28. If 2 or more questions are left unanswered, the questionnaire is not scored. If 2 or more response options are selected, the response option with the highest score should be recorded.

The total POEM score and POEM scores in each category will be summarized by visit and by treatment.

7.4.1.6. Dermatology Life Quality Index and Children's Dermatology Life Quality Index

The DLQI (Finlay and Khan 1994) is a simple 10-question validated questionnaire for use in participants aged 16 years and over to measure how much the skin problem has affected the participant over the previous 7 days.

The scoring of each question is as follows: Very much = 3; A lot = 2; A little = 1; Not at all = 0; Not relevant = 0; Question 7, "Prevented work or studying" = 3.

The following imputation will be applied for incorrectly completed questionnaires:

- If 1 question is left unanswered, this is scored 0 and the scores are summed and expressed as usual out of a maximum of 30.
- If 2 or more questions are left unanswered, the questionnaire is not scored.
- If question 7 is answered "yes," this is scored 3. If Question 7 is answered "no," but then either "a lot" or "a little" is ticked, this is then scored 2 or 1. If "Not relevant" is ticked, the score for Question 7 is 0. If it is answered "no," but the second half is left incomplete, the score will remain 0.
- If 2 or more response options are ticked, the response option with the highest score should be recorded.
- If there is a response between 2 tick boxes, the lower of the 2 score options should be recorded.
- For DLQI 6 subscales, if the answer to one question in a subscale is missing, that subscale should not be scored.

The DLQI is calculated by summing the score of each question resulting in a maximum of 30 and a minimum of 0. The higher the score, the more quality of life is impaired.

The meaning of total DLQI scores can be categorized as follows:

- 0 to 1 = No effect
- 2 to 5 = Small effect
- 6 to 10 = Moderate effect
- 11 to 20 = Very large effect
- 21 to 30 = Extremely large effect

A change from baseline in DLQI and CDLQI score of at least 4 points is considered clinically important (Basra et al 2015, Waters et al 2010).

The questionnaire is also analyzed under 6 sub-scales as follows:

- Symptoms and feelings (Questions 1 and 2)
- Daily activities (Questions 3 and 4)
- Leisure (Questions 5 and 6)
- Work and school (Question 7)
- Personal relations (Questions 8 and 9)
- Treatment (Question 10)

CDLQI (Lewis-Jones and Finlay 1995) is the youth/children's version of the DLQI and will be completed by adolescents aged ≥ 12 years to < 16 years. The scoring of each question is Very much = 3; Quite a lot = 2; Only a little = 1; Not at all = 0; Question unanswered = 0; Question 7: "Prevented school" = 3. The CDLQI is calculated by summing the score of each question resulting in a maximum of 30 and a minimum of 0. The higher the score, the more quality of life is impaired.

The questionnaire is also analyzed under 6 subscales as follows:

- Symptoms and feelings (Questions 1 and 2)
- Leisure (Questions 4, 5, and 6)
- School or holidays (Question 7)
- Personal relationships (Questions 3 and 8)
- Sleep (Question 9)
- Treatment (Question 10)

The severity banding for CDLQI scores:

- 0 to 1 = No effect on child's life
- 2 to 6 = Small effect
- 7 to 12 = Moderate effect

- 13 to 18 = Very large effect
- 19 to 30 = Extremely large effect

The following imputation will be applied for incorrectly completed questionnaires for CDLQI:

1. If one question is left unanswered, this is scored 0, and the scores are summed and expressed as usual out of a maximum of 30.
2. If 2 or more questions are left unanswered, the questionnaire is not scored.
3. If both parts of Question 7 are completed, the higher of the 2 scores should be counted.

7.4.1.7. Patient Global Impression of Change

The PGIC is a participants' self-reporting measure that reflects their belief about the efficacy of treatment. The PGIC is a 7-point scale depicting a participant's rating of overall improvement and will be captured during site visits during the VC period (1, very much improved; 2, much improved; 3, minimally improved; 4, no change; 5, minimally worse; 6, much worse; 7, very much worse). Missing values will not be imputed.

A participant is said to be a PGIC responder or have achieved a PGIC response if he or she has achieved PGIC of (1) very much improved or (2) much improved.

7.4.1.8. EQ-5D-5L

The EQ-5D-5L (EuroQol Research Foundation 2017) is a validated, self-administered, generic utility questionnaire wherein participants (adolescents and adults) rate their current health state based on the following criteria (dimensions): mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.

During all VC period study visits (starting at screening) and at specific LTS visits (Weeks 12, 24, 36, 52, and follow-up visit), the participant will be asked to indicate his/her health state over the past 7 days.

Missing values will not be imputed. The categorical outcomes for the 5 dimensions (mobility, self-care, usual activity, pain/discomfort, and anxiety/depression) will be summarized. The change from baseline in EQ-5D VAS score will be analyzed.

7.4.1.9. Work Productivity and Activity Impairment Questionnaire: Specific Health Problem Version 2.0

The WPAI:SHP v2.0 (Reilly et al 1993) questionnaire is a validated 6-item instrument, completed during all site visits starting at screening, during the VC period, and at specific LTS visits (Weeks 12, 24, 36, 52 and follow-up visit) that measures the effect of overall health and specific symptoms on productivity at work and regular activities outside of it during the past 7 days.

WPAI outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity, i.e., worse outcomes, as follows:

Questions:

- 1 = currently employed
- 2 = hours missed due to AD
- 3 = hours missed other reasons
- 4 = hours actually worked
- 5 = degree AD affected productivity while working
- 6 = degree AD affected regular activities

Scores:

Multiply scores by 100 to express in percentages.

- Percent work time missed due to AD: Q2 / (Q2 + Q4)
- Percent impairment while working due to AD: Q5 / 10
- Percent overall work impairment due to AD: [Q2 / (Q2 + Q4)] + ({1 – [Q2 / (Q2 + Q4)]} × {Q5 / 10})
- Percent activity impairment due to AD: Q6 / 10

These impairment percentages will be summarized by visit and by treatment. Missing values will not be imputed.

7.4.2. Secondary Efficacy Analysis

All secondary efficacy analyses will be conducted based on the ITT population in the VC and LTS periods.

7.4.2.1. Continuous Efficacy Endpoints

By-visit summary statistics for the following continuous measurements, including change from baseline and percentage change from baseline will be presented:

- EASI score
- By-visit Itch NRS score
- By-visit Skin Pain NRS score
- By-visit PROMIS Short Form – Sleep-Related Impairment (8a) 24-hour recall total raw score
- By-visit PROMIS Short Form – Sleep Disturbance (8b) 24-hour recall total raw score

The summary statistics includes sample size, mean, median, standard deviation, minimum, maximum, 1st quartile, 3rd quartile, and 95% CI will be presented by visits. An MMRM may be fit for the comparisons between ruxolitinib cream groups and vehicle cream group at Week 8. The MMRM will include the fixed effect of treatment, stratification factor, the visit, and

treatment by visit interaction. The variance-covariance matrix of the within-participant errors in MMRM will be modeled as unstructured.

For the following continuous measurements, the change from baseline to Week 8 will be analyzed using an ANCOVA model with treatment and stratification factors and baseline score as covariates if applicable.

- Total AD afflicted %BSA
- Total SCORAD score
- Total POEM score
- Total DLQI/CDLQI score
- DLQI/CDLQI subscores
- EQ VAS score
- WPAI:SHP v2.0 scores

For the above continuous measurements during the LTS period, only summary statistics will be presented.

7.4.2.2. Daily Continuous Efficacy Endpoints

For the following daily continuous measurements, summary statistics including sample size, mean, median, standard deviation, minimum, maximum, 1st quartile, 3rd quartile, and 95% CI will be presented by day. An ANOVA model will be fit for the comparisons between ruxolitinib cream and vehicle cream at Week 8 on change and percentage from baseline. The ANOVA model will include treatment groups.

- Daily Itch NRS score
- Daily Skin Pain NRS score
- Daily PROMIS Short Form – Sleep-Related Impairment (8a) 24-hour recall total raw score
- Daily PROMIS Short Form – Sleep Disturbance (8b) 24-hour recall total raw score

7.4.2.3. Categorical Efficacy Endpoints

For the following categorical parameters, summary statistics including sample size, frequency, and percentages will be presented by visits.

- Proportion of participants achieving an IGA of 0 or 1
- Proportion of participants with a $\geq$ 2-point improvement in Itch NRS score
- Proportion of participants with a $\geq$ 3-point improvement in Itch NRS score
- Proportion of participants with a $\geq$ 4-point improvement in Itch NRS score
- Proportion of participants achieving EASI50
- Proportion of participants achieving EASI75

- Proportion of participants achieving EASI90
- Proportion of participants with a ≥ 6-point improvement in the PROMIS Short Form (8a) – Sleep-Related Impairment 24-hour recall score
- Proportion of participants with a ≥ 6-point improvement in the PROMIS Short Form (8b) – Sleep Disturbance 24-hour recall score
- Proportion of participants with a score of either 1 or 2 of PGIC
- Proportion of participants with a clinically meaningful improvement (≥ 4-point improvement) in DLQI/CDLQI

A logistic regression model with treatment and stratification factors as covariates will be fit at Week 8. The p-values between each ruxolitinib group versus vehicle will be calculated based on Wald test. Exact logistic regression will be used for all of the comparisons if any of the dose levels have an expected cell count less than 5. The NRI will be used to impute postbaseline missing values for binary outcomes based on IGA, EASI, Itch NRS, PROMIS Short Form – Sleep-Related Impairment (8a) scores, PROMIS Short Form – Sleep Disturbance (8b) scores.

For the following categorical endpoints, the number and percentage of participants will be summarized by treatment at each postbaseline visit.

- Proportion of participants in each category of EASI score
- Proportion of participants in each category of IGA score
- Proportion of participants in each category of PGIC score
- Proportion of participants in each category of POEM score
- Proportion of participants in each category of DLQI/CDLQI score
- Proportion of participants in each category of each EQ-5D dimension

7.4.2.4. Time-To-Event Efficacy Endpoints

For the time to achieve Itch NRS score improvement of at least 2, 3, or 4 points, a log-rank test stratified by randomization stratification factors will be used for between treatment group comparisons. The hazard ratio and its 95% CI will be estimated based on the stratified Cox regression model using Efron's method accounting for ties. Kaplan-Meier curves will be presented by treatment groups. The number of participants, number of events and number of censoring will be summarized by treatment groups. The KM estimate of median time will be presented with its 95% CI. The 95% CI will be calculated using the method by Brookmeyer and Crowley (1982).

[redacted]

[redacted]

7.6. Pharmacokinetic Analyses

Trough plasma concentrations of ruxolitinib at all study visits will be summarized using descriptive statistics by treatment group.

[redacted]

8. SAFETY AND TOLERABILITY

Appendix A provides a list of planned tables figures, and listings. Sample data displays are included in a separate document.

8.1. General Considerations

The analyses in this section will be provided for the safety population in the VC and LTS periods. Summary tables may be replaced with listings when appropriate. For instance, an AE frequency table may be replaced with a listing if it only contains a few unique PTs reported on relatively few participants.

Unless otherwise stated, table summaries will be limited to AEs occurring within 30 days of the last application of study drug or the first application in the LTS period (if available) of the study, whichever is earlier.

8.2. Adverse Events

8.2.1. Adverse Event Definitions

A TEAE is any AE either reported for the first time or worsening of a pre-existing event after first administration of study drug. Analysis of AEs (as discussed below) will be limited to TEAEs, but data listings will include all AEs regardless of their timing in relation to study drug administration.

Adverse events will be tabulated by MedDRA PT and SOC. Severity of AEs will be graded using the NCI CTCAE. The CTCAE v4.03 is used for this study. The CTCAE reporting guidelines and grading details are available on the Cancer Therapy Evaluation Program website.

The subset of AEs considered by the investigator to be related to study drug will be considered to be treatment-related AEs. If the investigator does not specify the relationship of the AE to study drug, the AE will be considered to be treatment-related. The incidence of AEs and treatment-related AEs will be tabulated. Serious AEs will also be tabulated.

Any missing onset date, causality, or severity must be queried for resolution. Unresolved missing causality and severity will be handled according to the following rules:

- An unresolved missing causality will be considered treatment-related.
- An unresolved missing severity will be identified as an unknown severity.

For purposes of analysis, all AEs will be considered TEAEs unless the AE can unequivocally be defined as not treatment-emergent.

8.2.2. Adverse Event Summaries

An overall summary of AEs by treatment group will include:

- Number (%) of participants reporting any TEAEs
- Number (%) of participants reporting any serious TEAEs
- Number (%) of participants reporting any Grade 3 or higher TEAEs
- Number (%) of participants reporting any treatment-related TEAEs
- Number (%) of participants who temporarily interrupted study drug because of TEAEs
- Number (%) of participants who permanently discontinued study drug because of TEAEs
- Number (%) of participants who had a fatal TEAE

The following summaries will be produced by MedDRA term (if 2 or fewer participants appear in a table, a listing may be appropriate):

- Summary of TEAEs by SOC and PT
- Summary of TEAEs by PT in decreasing order of frequency
- Summary of Grade 3 or higher AEs by SOC and PT
- Summary of Grade 3 or higher AEs by PT in decreasing order of frequency
- Summary of serious TEAEs by SOC and PT
- Summary of serious TEAEs by PT in decreasing order of frequency
- Summary of treatment-related TEAEs by SOC and PT
- Summary of treatment-related TEAEs by PT in decreasing order of frequency

- Summary of Grade 3 or higher treatment-related TEAEs by SOC and PT
- Summary of treatment-related serious TEAEs by SOC and PT
- Summary of TEAEs with a fatal outcome by SOC and PT
- Summary of TEAEs leading to dose interruption of study drug by SOC and PT
- Summary of TEAEs leading to discontinuation of study drug by SOC and PT
- Summary of TEAEs requiring concomitant medications by SOC and PT

Subgroup analysis for TEAEs by age categories will also be provided.

8.3. Clinical Laboratory Tests

8.3.1. Laboratory Value Definitions

All laboratory assessments will be performed using a central laboratory except for urine pregnancy tests (as applicable). Laboratory values and change from baseline values will be summarized descriptively by visit, and non-numeric test values will be tabulated when necessary.

The baseline value will be determined using the last nonmissing value collected before the first application, prioritizing scheduled assessments for baseline identification over unscheduled visits. The last record before administration in the highest priority will be considered the baseline record. For baseline laboratory candidates with the same date and time in the same priority category, additional rules may be provided after consultation with the medical monitor to delineate which value will be defined as baseline.

8.3.2. Laboratory Value Summaries

Clinical laboratory tests, including hematology and serum chemistry, will be performed at the Protocol-specified visits. If specific safety issues arise, additional unscheduled laboratory tests/analyses may be performed at the discretion of the investigator.

All test results and associated normal ranges from central laboratories will be reported in SI units. All tests with numeric values will have a unique unit per test. Any laboratory test results and associated normal ranges from local laboratories will be converted to SI units.

When there are multiple laboratory nonmissing values for a participant's particular test within a visit window, the laboratory value with the smallest laboratory sequence number will be used in by-visit summaries.

For test results that will be summarized with available normal ranges, the number and percentage of participants with the laboratory values being low (but never high), normal, high (but never low), and both low and high will be calculated for each test. This shift summary will be produced for each test for the safety population in VC period. Shift tables will be presented showing change in CTCAE grade from baseline to worst grade postbaseline as well. The denominator for the percentage calculation will use the number of participants in the baseline category.

Subgroup analysis for laboratory results by age categories will also be provided.

8.4. Vital Signs

Values at each scheduled visit, change, and percentage change from baseline for vital signs, including SBP, DBP, pulse, respiratory rate, and body temperature will be summarized descriptively.

Criteria for clinically notable vital sign abnormalities are defined in Table 3, Table 4, and Table 5. The abnormal values for participants exhibiting clinically notable vital sign abnormalities will be listed along with their assigned treatment group. Alert vital signs are defined as an absolute value outside the defined range and percentage change greater than 25%. The abnormal values for participants exhibiting alert vital sign abnormalities will be listed.

Table 3: Criteria for Clinically Notable Vital Sign Abnormalities for 12 to 15 Years Old

Parameter	High Threshold	Low Threshold
Systolic blood pressure	> 131 mmHg	< 110 mmHg
Diastolic blood pressure	> 83 mmHg	< 64 mmHg
Pulse	> 100 bpm	< 60 bpm
Temperature	> 38.0ºC	< 35.5ºC
Respiratory rate	> 20 breaths/min	< 8 breaths/min

Table 4: Criteria for Clinically Notable Vital Sign Abnormalities for 16 to 17 Years Old

Parameter	High Threshold	Low Threshold
Systolic blood pressure	> 120 mmHg	< 90 mmHg
Diastolic blood pressure	> 85 mmHg	< 50 mmHg
Pulse	> 100 bpm	< 45 bpm
Temperature	> 38.0ºC	< 35.5ºC
Respiratory rate	> 20 breaths/min	< 8 breaths/min

Table 5: Criteria for Clinically Notable Vital Sign Abnormalities for ≥ 18 Years Old

Parameter	High Threshold	Low Threshold
Systolic blood pressure	> 155 mmHg	< 85 mmHg
Diastolic blood pressure	> 100 mmHg	< 40 mmHg
Pulse	> 100 bpm	< 45 bpm
Temperature	> 38ºC	< 35.5ºC
Respiratory rate	> 20 breaths/min	< 8 breaths/min

9. PLANNED ANALYSES

No formal interim analysis is planned. There are 3 formal planned analyses:

- The primary analysis will occur after the primary database lock, when all participants have completed the vehicle-controlled, double-blind treatment period. The sponsor will be unblinded after the primary database lock; however, investigators and participants will remain blinded to the individual study treatment assignment;
- An analysis will occur when approximately 200 participants (≥ 20% adolescents) complete 12 months of treatment across both INCB 18424-303 and INCB 18424-304 studies;
- The final analysis will occur when all participants have completed or withdrew from the study.

10. CHANGES AND MODIFICATIONS TO THE ANALYSIS PLAN

All versions of the SAP are listed in Table 6.

Table 6: Statistical Analysis Plan Versions

SAP Version	Date
Original	08 NOV 2019

10.1. Changes to Protocol-Defined Analyses

- The LTS evaluable population will not be used. Efficacy and safety in LTS will be evaluated in the ITT and safety populations to keep consistency with the measurements in VC.
- Duration of AEs will not be analyzed.
- The sleep-related impairment is considered as an important endpoint. The proportion of participants with a clinically meaningful improvement (≥ 6-point) in the PROMIS Short Form – Sleep-Related Impairment (8a-24-hour recall) score at Weeks 8 has been moved into the list of key secondary endpoints with alpha control.

10.2. Changes to the Statistical Analysis Plan

Not applicable.

11. REFERENCES

Basra MK, Salek MS, Camilleri L, Sturkey R, Finlay AY. Determining the minimal clinically important difference and responsiveness of the Dermatology Life Quality Index (DLQI): further data. Dermatology 2015;230:27-33.

Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. Stat Med 2009;28:586-604.

Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics 1982;38:29-41.

Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure: development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective. Arch Dermatol 2004;140:1513-1519.

EuroQol Research Foundation. EQ-5D-5L. 2017. https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/. Accessed July 17, 2018.

Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)-- a simple practical measure for routine clinical use. Clin Exp Dermatol 1994;19:210-216.

Lewis-Jones MS, Finlay AY. The Children's Dermatology Life Quality Index (CDLQI): initial validation and practical use. Br J Dermatol 1995;132:942-949.

Mehta CR, Patel NR. Exact logistic regression: theory and examples. Stat Med 1995;14:2143-2160.

Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. PharmacoEconomics 1993;4:353-365.

Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York, NY: John Wiley & Sons;1987.

University of Nottingham. Harmonising Outcome Measures for Eczema (HOME). www.homeforeczema.org. Accessed November 6, 2019.

van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res 2007;16:219-242.

Waters A, Sandhu D, Beattie P, Ezughah F, Lewis-Jones S. Severity stratification of Children's Dermatology Life Quality Index (CDLQI) scores. Br J Dermatol 2010;163(suppl 1):121.

APPENDIX A. PLANNED TABLES AND FIGURES

This appendix provides a list of the planned tables, figures, and listing for the clinical study report. Standard tables will follow the conventions in the Standard Safety Tables initial version. Shells are provided for nonstandard tables. In-text tables are identical in structure and content as appendix tables, but follow a Rich Text Format.

The list of tables, figures, listings, and the shells are to be used as guideline. Modifications of the list or shells that do not otherwise affect the nature of the analysis will not warrant an amendment to the SAP.

Tables

Table No.	Title	Population
Baseline Demographic and Characteristic		
1.1.1	Analysis Populations	All
1.1.2	Summary of Participant Disposition	ITT
1.1.3	Summary of Number of Participants Enrolled by Country and Site	ITT
1.1.4	Summary of Protocol Deviations	ITT
1.2	Summary of Demographics and Baseline Characteristics	ITT
1.3	Summary of Baseline Disease Characteristics	ITT
1.4.1	Summary of Prior Medications	ITT
1.4.2	Summary of Prior Medications for Atopic Dermatitis	ITT
1.4.3.1	Summary of Concomitant Medications in VC Period	ITT
1.4.3.2	Summary of Concomitant Medications from Baseline to Week 52	ITT
1.4.3.3	Summary of Concomitant Medications in LTS Period	ITT
1.5	Summary of General Medical History	ITT
Efficacy		
IGA		
2.1.1.1	Summary and Analysis of Participants Achieving IGA-TS From Baseline to Week 52	ITT
2.1.1.2	Summary of Participants Achieving IGA-TS in LTS Period	ITT
2.1.2	Summary and Sensitivity Analysis of Participants Achieving IGA-TS at Week 8 by Tipping Point Analysis	ITT
2.1.3.1	Summary and Analysis of Participants Achieving IGA-TS From Baseline to Week 52	PP
2.1.4	Summary and Analysis of Participants Achieving IGA-TS by Baseline IGA Score From Baseline to Week 52	ITT
2.1.5	Summary and Analysis of Participants Achieving IGA-TS by Baseline EASI Score From Baseline to Week 52	ITT
2.1.6	Summary and Analysis of Participants Achieving IGA-TS by Region From Baseline to Week 52	ITT
2.1.7	Summary and Analysis of Participants Achieving IGA-TS by Age Group From Baseline to Week 52	ITT
2.1.8	Summary and Analysis of Participants Achieving IGA-TS by Sex From Baseline to Week 52	ITT
2.1.9	Summary and Analysis of Participants Achieving IGA-TS by Race From Baseline to Week 52	ITT

Table No.	Title	Population
2.1.10.1	Summary of Participants in Each Category of IGA Score From Baseline to Week 52	ITT
2.1.10.2	Summary of Participants in Each Category of IGA Score in LTS Period	ITT
2.1.11.1	Summary and Analysis of Participants Achieving an IGA of 0 or 1 From Baseline to Week 52	ITT
2.1.11.2	Summary of Participants Achieving an IGA of 0 or 1 in LTS Period	ITT
2.1.12	Shift Summary of IGA at Week 8	ITT
EASI		
2.2.1	Summary and Analysis of Participants Achieving EASI75 in VC Period	ITT
2.2.2	Summary and Analysis of Participants Achieving EASI50 in VC Period	ITT
2.2.3	Summary and Analysis of Participants Achieving EASI90 in VC Period	ITT
2.2.4	Summary and Analysis of EASI Score in VC Period	ITT
2.2.5	Summary of Participants in Each Category of EASI Score in VC Period	ITT
Itch NRS Score		
2.3.1	Summary and Analysis of Participants Achieving ≥ 2-Point Improvement in Itch NRS Score in VC Period	ITT
2.3.2	Summary and Analysis of Participants With a ≥ 3-Point Improvement in Itch NRS Score in VC Period	ITT
2.3.3	Summary and Analysis of Participants With a ≥ 4-Point Improvement in Itch NRS Score in VC Period	ITT
2.3.4	Summary and Analysis of Itch NRS Score in VC Period	ITT
2.3.5	Summary and Analysis of Daily Itch NRS Score in VC Period	ITT
2.3.6	Summary and Analysis of Time to ≥ 2-Point Improvement in Itch NRS Score in VC Period	ITT
2.3.7	Summary and Analysis of Time to ≥ 3-Point Improvement in Itch NRS Score in VC Period	ITT
2.3.8	Summary and Analysis of Time to ≥ 4-Point Improvement in Itch NRS Score in VC Period	ITT
PROMIS Sleep Disturbance Score (8b)		
2.4.1.1	Summary and Analysis of Participants Achieving a ≥ 6-point Improvement in the PROMIS Sleep Disturbance Score (8b) From Baseline to Week 52	ITT
2.4.1.2	Summary of Participants Achieving a ≥ 6-point Improvement in the PROMIS Sleep Disturbance Score (8b) in LTS Period	ITT
2.4.2.1	Summary and Analysis of By-Visit PROMIS Sleep Disturbance Score (8b) From Baseline to Week 52	ITT
2.4.2.2	Summary of By-Visit PROMIS Sleep Disturbance Score (8b) in LTS Period	ITT
2.4.3	Summary and Analysis of Daily PROMIS Sleep Disturbance Score (8b) in VC Period	ITT
PROMIS Sleep-Related Impairment Score (8a)		
2.5.1.1	Summary and Analysis of Participants Achieving a ≥ 6-point Improvement in the PROMIS Sleep-Related Impairment Score (8a)From Baseline to Week 52	ITT
2.5.1.2	Summary of Participants Achieving a ≥ 6-point Improvement in the PROMIS Sleep-Related Impairment Score (8a) in LTS Period	ITT
2.5.2.1	Summary and Analysis of By-Visit PROMIS Sleep-Related Impairment Score (8a) From Baseline to Week 52	ITT

Table No.	Title	Population
2.5.2.2	Summary of By-Visit PROMIS Sleep-Related Impairment Score (8a) in LTS Period	ITT
2.5.3	Summary and Analysis of Daily PROMIS Sleep-Related Impairment Score (8a) in VC Period	ITT
SCORAD		
2.6	Summary and Analysis of SCORAD Score in VC Period	ITT
Skin Pain NRS Score		
2.7.1	Summary and Analysis of By-Visit Skin Pain NRS Score in VC Period	ITT
2.7.2	Summary and Analysis of Daily Skin Pain NRS Score in VC Period	ITT
BSA		
2.8.1.1	Summary and Analysis of Total %BSA Afflicted by AD From Baseline to Week 52	ITT
2.8.1.2	Summary of Total %BSA Afflicted by AD in LTS Period	ITT
POEM		
2.9.1.1	Summary and Analysis of Total POEM Score From Baseline to Week 52	ITT
2.9.1.2	Summary of Total POEM Score in LTS Period	ITT
2.9.2.1	Summary of Participants in Each Category of POEM Score From Baseline to Week 52	ITT
2.9.2.2	Summary of Participants in Each Category of POEM Score in LTS Period	ITT
DLQI/CDLQI		
2.10.1.1	Summary and Analysis of Total DLQI Score From Baseline to Week 52	ITT
2.10.1.2	Summary and Analysis of Total CDLQI Score From Baseline to Week 52	ITT
2.10.1.3	Summary of Total DLQI Score in LTS Period	ITT
2.10.1.4	Summary of Total CDLQI Score in LTS Period	ITT
2.10.2.1	Summary and Analysis of DLQI Subscore From Baseline to Week 52	ITT
2.10.2.2	Summary and Analysis of CDLQI Subscore From Baseline to Week 52	ITT
2.10.2.3	Summary of DLQI Subscore in LTS Period	ITT
2.10.2.4	Summary of CDLQI Subscore in LTS Period	ITT
2.10.3.1	Summary of Participants in Each Category of DLQI Score From Baseline to Week 52	ITT
2.10.3.2	Summary of Participants in Each Category of CDLQI Score From Baseline to Week 52	ITT
2.10.3.3	Summary of Participants in Each Category of DLQI Score in LTS Period	ITT
2.10.3.4	Summary of Participants in Each Category of CDLQI Score in LTS Period	ITT
2.10.4.1	Summary and Analysis of Participants Achieving ≥ 4 Improvement in DLQI Score From Baseline to Week 52	ITT
2.10.4.2	Summary and Analysis of Participants Achieving ≥ 4 Improvement in CDLQI Score From Baseline to Week 52	ITT
2.10.4.3	Summary of Participants Achieving ≥ 4 Improvement in DLQI Score in LTS Period	ITT
2.10.4.4	Summary of Participants Achieving ≥ 4 Improvement in CDLQI Score in LTS Period	ITT
PGIC		
2.11.1	Summary and Analysis of Participants Achieving a PGIC Score of 1 or 2 in VC Period	ITT
2.11.2	Summary of Participants in Each Category of PGIC Score in VC Period	ITT
2.11.3	Summary of PGIC Score in VC Period	ITT

Table No.	Title	Population
EQ-5D-5L		
2.12.1	Summary and Analysis of EQ VAS Score From Baseline to Week 52	ITT
2.12.2	Summary of EQ VAS Score in LTS Period	ITT
WPAI-SHP		
2.13.1.1	Summary and Analysis of WPAI-SHP v2.0 Score From Baseline to Week 52	ITT
2.13.1.2	Summary of WPAI-SHP v2.0 Score in LTS Period	ITT
[redacted]		
Exposure, Adverse Events, Laboratory, and Vital Signs		
Exposure		
3.1.1.1	Summary of Exposure in VC Period	Safety
3.1.1.2	Summary of Exposure From Baseline to Week 52	Safety
3.1.1.3	Summary of Exposure in LTS Period	Safety
3.1.2	Summary of Study Drug Compliance in VC Period	Safety
Adverse Events		
3.2.1.1	Overall Summary of Treatment-Emergent Adverse Events in VC Period	Safety
3.2.1.1.1	Overall Summary of Treatment-Emergent Adverse Events by Age Group in VC Period	Safety
3.2.1.2	Overall Summary of Treatment-Emergent Adverse Events From Baseline to Week 52	Safety
3.2.1.2.1	Overall Summary of Treatment-Emergent Adverse Events by Age Group From Baseline to Week 52	Safety
3.2.1.3	Overall Summary of Treatment-Emergent Adverse Events in LTS Period	Safety
3.2.2.1	Summary of Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in VC Period	Safety
3.2.2.1.1	Summary of Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group in VC Period	Safety
3.2.2.2	Summary of Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term From Baseline to Week 52	Safety
3.2.2.2.1	Summary of Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group From Baseline to Week 52	Safety
3.2.2.3	Summary of Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in LTS Period	Safety
3.2.3.1	Summary of Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency in VC Period	Safety
3.2.3.1.1	Summary of Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency by Age Group in VC Period	Safety
3.2.3.2	Summary of Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency From Baseline to Week 52	Safety
3.2.3.2.1	Summary of Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency by Age Group From Baseline to Week 52	Safety
3.2.3.3	Summary of Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency in LTS Period	Safety

Table No.	Title	Population
3.2.6.1	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in VC Period	Safety
3.2.6.1.1	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group in VC Period	Safety
3.2.6.2	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term From Baseline to Week 52	Safety
3.2.6.2.1	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group From Baseline to Week 52	Safety
3.2.6.3	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in LTS Period	Safety
3.2.7.1	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency in VC Period	Safety
3.2.7.1.1	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency by Age Group in VC Period	Safety
3.2.7.2	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency From Baseline to Week 52	Safety
3.2.7.2.1	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency by Age Group From Baseline to Week 52	Safety
3.2.7.3	Summary of Grade 3 or Higher Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency in LTS Period	Safety
3.2.8.1	Summary of Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in VC Period	Safety
3.2.8.1.1	Summary of Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group in VC Period	Safety
3.2.8.2	Summary of Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term From Baseline to Week 52	Safety
3.2.8.2.1	Summary of Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group From Baseline to Week 52	Safety
3.2.8.3	Summary of Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in LTS Period	Safety
3.2.9.1	Summary of Serious Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency in VC Period	Safety
3.2.9.1.1	Summary of Serious Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency by Age Group in VC Period	Safety
3.2.9.2	Summary of Serious Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency From Baseline to Week 52	Safety
3.2.9.2.1	Summary of Serious Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency by Age Group From Baseline to Week 52	Safety
3.2.9.3	Summary of Serious Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency in LTS Period	Safety
3.2.10.1	Summary of Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in VC Period	Safety
3.2.10.1.1	Summary of Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group in VC Period	Safety

Table No.	Title	Population
3.2.10.2	Summary of Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term From Baseline to Week 52	Safety
3.2.10.2.1	Summary of Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group From Baseline to Week 52	Safety
3.2.10.3	Summary of Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in LTS Period	Safety
3.2.11.1	Summary of Treatment-Related Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency in VC Period	Safety
3.2.11.1.1	Summary of Treatment-Related Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency by Age Group in VC Period	Safety
3.2.11.2	Summary of Treatment-Related Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency From Baseline to Week 52	Safety
3.2.11.2.1	Summary of Treatment-Related Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency by Age Group From Baseline to Week 52	Safety
3.2.11.3	Summary of Treatment-Related Treatment-Emergent Adverse Events by MedDRA Preferred Term in Decreasing Order of Frequency in LTS Period	Safety
3.2.14.1	Summary of Grade 3 or Higher Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in VC Period	Safety
3.2.14.1.1	Summary of Grade 3 or Higher Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group in VC Period	Safety
3.2.14.2	Summary of Grade 3 or Higher Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term From Baseline to Week 52	Safety
3.2.14.2.1	Summary of Grade 3 or Higher Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group From Baseline to Week 52	Safety
3.2.14.3	Summary of Grade 3 or Higher Treatment-Related Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in LTS Period	Safety
3.2.15.1	Summary of Treatment-Related Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in VC Period	Safety
3.2.15.1.1	Summary of Treatment-Related Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group in VC Period	Safety
3.2.15.2	Summary of Treatment-Related Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term From Baseline to Week 52	Safety
3.2.15.2.1	Summary of Treatment-Related Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term by Age Group From Baseline to Week 52	Safety
3.2.15.3	Summary of Treatment-Related Serious Treatment-Emergent Adverse Events by MedDRA System Organ Class and Preferred Term in LTS Period	Safety
3.2.16.1	Summary of Treatment-Emergent Adverse Events With a Fatal Outcome by MedDRA System Organ Class and Preferred Term in VC Period	Safety

Table No.	Title	Population
3.2.16.1.1	Summary of Treatment-Emergent Adverse Events With a Fatal Outcome by MedDRA System Organ Class and Preferred Term by Age Group in VC Period	Safety
3.2.16.2	Summary of Treatment-Emergent Adverse Events With a Fatal Outcome by MedDRA System Organ Class and Preferred Term From Baseline to Week 52	Safety
3.2.16.2.1	Summary of Treatment-Emergent Adverse Events With a Fatal Outcome by MedDRA System Organ Class and Preferred Term by Age Group From Baseline to Week 52	Safety
3.2.16.3	Summary of Treatment-Emergent Adverse Events With a Fatal Outcome by MedDRA System Organ Class and Preferred Term in LTS Period	Safety
3.2.19.1	Summary of Treatment-Emergent Adverse Events Leading to Dose Interruption by MedDRA System Organ Class and Preferred Term in VC Period	Safety
3.2.19.1.1	Summary of Treatment-Emergent Adverse Events Leading to Dose Interruption by MedDRA System Organ Class and Preferred Term by Age Group in VC Period	Safety
3.2.19.2	Summary of Treatment-Emergent Adverse Events Leading to Dose Interruption by MedDRA System Organ Class and Preferred Term From Baseline to Week 52	Safety
3.2.19.2.1	Summary of Treatment-Emergent Adverse Events Leading to Dose Interruption by MedDRA System Organ Class and Preferred Term by Age Group From Baseline to Week 52	Safety
3.2.19.3	Summary of Treatment-Emergent Adverse Events Leading to Dose Interruption by MedDRA System Organ Class and Preferred Term in LTS Period	Safety
3.2.20.1	Summary of Treatment-Emergent Adverse Events Leading to Discontinuation of Study Drug by MedDRA System Organ Class and Preferred Term in VC Period	Safety
3.2.20.1.1	Summary of Treatment-Emergent Adverse Events Leading to Discontinuation of Study Drug by MedDRA System Organ Class and Preferred Term by Age Group in VC Period	Safety
3.2.20.2	Summary of Treatment-Emergent Adverse Events Leading to Discontinuation of Study Drug by MedDRA System Organ Class and Preferred Term From Baseline to Week 52	Safety
3.2.20.2.1	Summary of Treatment-Emergent Adverse Events Leading to Discontinuation of Study Drug by MedDRA System Organ Class and Preferred Term by Age Group From Baseline to Week 52	Safety
3.2.20.3	Summary of Treatment-Emergent Adverse Events Leading to Discontinuation of Study Drug by MedDRA System Organ Class and Preferred Term in LTS Period	Safety
3.2.21.1	Summary of Treatment-Emergent Adverse Events Requiring Concomitant Medications by MedDRA System Organ Class and Preferred Term in VC Period	Safety
3.2.21.1.1	Summary of Treatment-Emergent Adverse Events Requiring Concomitant Medications by MedDRA System Organ Class and Preferred Term by Age Group in VC Period	Safety
3.2.21.2	Summary of Treatment-Emergent Adverse Events Requiring Concomitant Medications by MedDRA System Organ Class and Preferred Term From Baseline to Week 52	Safety
3.2.21.2.1	Summary of Treatment-Emergent Adverse Events Requiring Concomitant Medications by MedDRA System Organ Class and Preferred Term by Age Group From Baseline to Week 52	Safety

Table No.	Title	Population
3.2.21.3	Summary of Treatment-Emergent Adverse Events Requiring Concomitant Medications by MedDRA System Organ Class and Preferred Term in LTS Period	Safety
Laboratory		
3.3.1.1	Summary of Laboratory Values – Hematology From Baseline to Week 52	Safety
3.3.1.1.1	Summary of Laboratory Values – Hematology by Age Group From Baseline to Week 52	Safety
3.3.1.2	Summary of Laboratory Values – Hematology in LTS Period	Safety
3.3.2.1	Summary of Laboratory Values – Chemistry From Baseline to Week 52	Safety
3.3.2.1.1	Summary of Laboratory Values – Chemistry by Age Group From Baseline to Week 52	Safety
3.3.2.2	Summary of Laboratory Values – Chemistry in LTS Period	Safety
3.3.3	Shift Summary of Hematology Laboratory Values in CTC Grade – To the Worst Abnormal Value From Baseline to Week 8	Safety
3.3.4	Shift Summary of Chemistry Laboratory Values in CTC Grade – To the Worst Abnormal Value From Baseline to Week 8	Safety
3.3.5	Shift Summary of Hematology Values – To the Worst Abnormal Value	Safety
3.3.6	Shift Summary of Chemistry Values – To the Worst Abnormal Value	Safety
Vital Signs		
3.4.1.1	Summary of Systolic Blood Pressure From Baseline to Week 52	Safety
3.4.1.2	Summary of Systolic Blood Pressure in LTS Period	Safety
3.4.2.1	Summary of Diastolic Blood Pressure From Baseline to Week 52	Safety
3.4.2.2	Summary of Diastolic Blood Pressure in LTS Period	Safety
3.4.3.1	Summary of Pulse From Baseline to Week 52	Safety
3.4.3.2	Summary of Pulse in LTS Period	Safety
3.4.4.1	Summary of Respiratory Rate From Baseline to Week 52	Safety
3.4.4.2	Summary of Respiratory Rate in LTS Period	Safety
3.4.5.1	Summary of Body Temperature From Baseline to Week 52	Safety
3.4.5.2	Summary of Body Temperature in LTS Period	Safety

Figure

Figure No.	Title	Population
4.1	Proportion of Participants Achieving IGA-TS From Baseline to Week 52	ITT
4.2.1	Proportion of Participants Achieving EASI75 in VC Period	ITT
4.2.2.1	Mean and Standard Error Plot of EASI Score in VC Period	ITT
4.2.2.2	Mean and Standard Error Plot of Change From Baseline in EASI Score in VC Period	ITT
4.2.2.3	Mean and Standard Error Plot of Percent Change From Baseline in EASI Score in VC Period	ITT
4.3.1	Proportion of Participants Achieving ≥ 4-point Improvement in Itch NRS Score in VC Period	ITT
4.3.2.1	Mean and Standard Error Plot of Itch NRS Score in VC Period	ITT
4.3.2.2	Mean and Standard Error Plot of Change From Baseline in Itch NRS Score in VC Period	ITT
4.3.2.3	Mean and Standard Error Plot of Percent Change From Baseline in Itch NRS Score in VC Period	ITT

Figure No.	Title	Population
4.3.3.1	Mean and Standard Error Plot of Daily Itch NRS Score in VC Period	ITT
4.3.3.2	Mean and Standard Error Plot of Change From Baseline in Daily Itch NRS Score in VC Period	ITT
4.3.3.3	Mean and Standard Error Plot of Percent Change From Baseline in Daily Itch NRS Score in VC Period	ITT
4.3.4.1	Kaplan-Meier Curve of the Time to ≥ 2-Point Improvement in Itch NRS Score in VC Period	ITT
4.3.4.2	Kaplan-Meier Curve of the Time to ≥ 3-Point Improvement in Itch NRS Score in VC Period	ITT
4.3.4.3	Kaplan-Meier Curve of the Time to ≥ 4-Point Improvement in Itch NRS Score in VC Period	ITT
4.4.1	Proportion of Participants Achieving a ≥ 6-Point Improvement in the PROMIS Sleep Disturbance Score (8b) From Baseline to Week 52	ITT
4.4.2.1	Mean and Standard Error Plot of PROMIS Sleep Disturbance Score (8b) From Baseline to Week 52	ITT
4.4.2.2	Mean and Standard Error Plot of Change From Baseline in PROMIS Sleep Disturbance Score (8b) From Baseline to Week 52	ITT
4.4.2.3	Mean and Standard Error Plot of Percentage Change From Baseline in PROMIS Sleep Disturbance Score (8b) From Baseline to Week 52	ITT
4.4.3.1	Mean and Standard Error Plot of Daily PROMIS Sleep Disturbance Score (8b) From Baseline to Week 52	ITT
4.4.3.2	Mean and Standard Error Plot of Change From Baseline in Daily PROMIS Sleep Disturbance Score (8b) From Baseline to Week 52	ITT
4.4.3.3	Mean and Standard Error Plot of Percent Change From Baseline in Daily PROMIS Sleep Disturbance Score (8b) From Baseline to Week 52	ITT
4.5.1	Proportion of Participants Achieving a ≥ 6-point Improvement in the PROMIS Sleep-Related Impairment (8a) From Baseline to Week 52	ITT
4.5.2.1	Mean and Standard Error Plot of PROMIS Sleep-Related Impairment (8a) From Baseline to Week 52	ITT
4.5.2.2	Mean and Standard Error Plot of Change From Baseline in PROMIS Sleep-Related Impairment (8a) From Baseline to Week 52	ITT
4.5.2.3	Mean and Standard Error Plot of Percentage Change From Baseline in PROMIS Sleep-Related Impairment (8a) From Baseline to Week 52	ITT
4.5.3.1	Mean and Standard Error Plot of Daily PROMIS Sleep-Related Impairment (8a) From Baseline to Week 52	ITT
4.5.3.2	Mean and Standard Error Plot of Change From Baseline in Daily PROMIS Sleep-Related Impairment (8a) From Baseline to Week 52	ITT
4.5.3.3	Mean and Standard Error Plot of Percent Change From Baseline in Daily PROMIS Sleep-Related Impairment (8a) From Baseline to Week 52	ITT
4.6	Kaplan-Meier Curve of the Time to First Retreatment in LTS Period	ITT

Listings

Listing No.	Title
Baseline Demographic and Characteristic	
2.1.1	Participant Enrollment and Disposition Status
2.1.2	Participant Inclusion and Exclusion Criteria Violations
2.2	Protocol Deviations
2.3	Analysis Populations

Listing No.	Title
2.4.1	Demographic Characteristics
2.4.2	Baseline Disease Characteristics
2.4.3	Prior and Concomitant Medications
2.4.4	Prior Medications for Atopic Dermatitis
2.4.5	Medical History
2.5	Study Drug Exposure and Compliance
Efficacy	
2.6.1	IGA Score
2.6.2	EASI Score
2.6.3.1	Itch NRS Score
2.6.3.2	Time to Itch Response
2.6.4	PROMIS Sleep Disturbance Score (8b)
2.6.5	PROMIS Sleep-Related Impairment Score (8a)
2.6.6	SCORAD Score
2.6.7	Skin Pain NRS Score
2.6.8	AD Afflicted %BSA
2.6.9	POEM Score
2.6.10.1	DLQI Score
2.6.10.2	CDLQI Score
2.6.11	EQ-5D-5L Score
2.6.12	WPAI-SHP v2.0 Score
2.6.13	PGIC Score
2.6.14	Treatment Cycle in LTS Period
Adverse Events	
2.7.1	Adverse Events
2.7.2	Adverse Events Leading to Drug Discontinuation
2.7.3	Serious Adverse Events
2.7.4	Treatment-Related Adverse Events
2.7.5	Adverse Events With a Fatal Outcome
2.7.6	Adverse Events Leading to Interruption of Study Drug
2.7.7	Adverse Events Requiring Concomitant Medications
2.7.8	Adverse Events with Grade 3 or Higher
Laboratory Data	
2.8.1	Clinical Laboratory Values - Hematology
2.8.2	Clinical Laboratory Values - Serum Chemistry
2.8.3	Abnormal Clinical Laboratory Values - Hematology
2.8.4	Abnormal Clinical Laboratory Values - Serum Chemistry
2.8.5	PK Blood Sampling Times
Vital Signs	
2.9.1	Vital Signs
2.9.2	Abnormal Vital Sign Values
2.9.3	Alert Vital Sign Values

Signature Manifest

Document Number: IC-STS-SAP-0168 **Revision:** 0

Title: 18424-303 and -304 SAP

All dates and times are in Eastern Standard Time.

18424-303 and -304 SAP

Approval and Release

Name/Signature	Title	Date	Meaning/Reason
[redacted]	[redacted]	08 Nov 2019, 02:57:20 PM	Approved
[redacted]	[redacted]	08 Nov 2019, 03:16:04 PM	Approved
[redacted]	[redacted]	11 Nov 2019, 10:06:50 AM	Approved
[redacted]	[redacted]	11 Nov 2019, 10:52:11 AM	Approved