Protocol B7451094

A RANDOMIZED, OPEN-LABEL, PARALLEL-GROUP STUDY TO EVALUATE THE SAFETY AND EFFICACY OF ABROCITINIB 100 MG AND 200 MG TABLETS IN PARTICIPANTS AGED 12 YEARS AND OLDER WITH MODERATE TO SEVERE ATOPIC DERMATITIS IN INDIA

Statistical Analysis Plan
(SAP)

Version: 3

Date: 13 Jul 2023

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
1 18 Feb 2022	Original 21 Dec 2021	N/A	N/A
2 12 May 2022	Amendment 1 23 Mar 2022	Per India BoH's recommendation	Added SCORAD efficacy assessment for atopic dermatitis as a secondary endpoint. (Sections 2.1, 2.1.2, 3.2.3, 5.2.1, 6.2.3, Appendix 1, Appendix 2.4) Revised "abnormal developmental bone findings", "abnormal bone findings", "bone abnormalities" to "bone safety findings" in Sections 2.1, 2.1.3, 2.2, 3.3.1, 5.2.1, 6.3.1, Appendix 1. Removed text: "Photographic images of the lesion will be taken as part of source documentation, wherever possible." in Appendix 2.2 and Appendix 2.3 to align with Protocol Amendment 1. Added subset analysis based on monotherapy or combination therapy in Section 6.4.
3 13 Jul 2023	Amendment 1 23 Mar 2022	Add safety population	Added safety population and details for FAS in the analysis sets in Section 4. Changed the analysis population from FAS to Safety in Sections 3.1, 6.1.1, 6.6 and Appendix 1. Removed 'randomized' in Sections 6.5.1 and 6.5.2 as tables will be summarized according to actual treatment.

2. INTRODUCTION

This statistical analysis plan (SAP) provides the detailed methodology for summary and statistical analyses of the data collected in Study B7451094. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives, Endpoints, and Estimands

Objectives	Endpoints	Estimands
Primary:	**Primary:**	**Primary:**
• To evaluate the safety of abrocitinib in participants aged 12 years and older with moderate to severe Atopic Dermatitis (AD)	• Incidence of adverse events (AEs) and serious adverse events (SAEs)	• The primary estimand for safety is the incidence of AEs and SAEs in participants aged 12 years and older with moderate to severe AD, from the time of first dose to Week 16, regardless of dosing compliance or treatment discontinuation.
Secondary:	**Secondary:**	**Secondary:**
• To assess the efficacy of abrocitinib in participants aged 12 years and older with moderate to severe AD	• Response based on Investigator's Global Assessment (IGA) score of clear (0) or almost clear (1) and greater than or equal to 2 points improvement from baseline at Week 12 • Response based on achieving ≥75% improvement from baseline in the Eczema Area and Severity Index total score (EASI-75) at Week 12 • Response based on ≥75% improvement in Scoring Atopic Dermatitis total score (SCORAD75) at week 12	• The secondary estimand is the treatment effect of abrocitinib on participants aged 12 years and older with moderate to severe AD, based on IGA response, EASI-75 response and SCORAD75 response from baseline to Week 12, considering patients who discontinued treatment as non-responders and based on outcome measures of POEM and ADCT from baseline to Week 12 and at all scheduled time points. Data collected after treatment discontinuation is excluded.

	• Change from baseline in Patient-Oriented Eczema Measure (POEM) at Week 12 and at all scheduled time points • Change from baseline in Atopic Dermatitis Control Tool (ADCT) at Week 12 and at all scheduled time points	
Primary substudy objective(s):	**Primary substudy endpoints (as secondary endpoints for B7451094):**	**Primary substudy estimand(s):**
• To evaluate the potential effects of abrocitinib on bone development in adolescent participants 12 to <18 years of age, as assessed by knee magnetic resonance imaging (MRI).	• The incidence of bone safety findings in knee MRI 1 year after randomization in adolescent participants exposed to abrocitinib 100 mg and 200 mg once daily (QD).	• The primary substudy estimand is the incidence of bone safety findings in knee MRI after 1 year of being exposed to abrocitinib in adolescent participants 12 to <18 years of age, regardless of dosing compliance or treatment discontinuation.

2.1.1. Primary Estimand(s)

The primary estimand is the incidence of AE/SAE regardless of whether an intercurrent event occurs. The estimand is defined according to the primary objective and is in alignment with the primary endpoint. It includes the following 4 attributes:

- Population: Patients aged 12 years and older with moderate to severe AD, as defined by the inclusion and exclusion criteria, and who are randomized;
- Variable: Binary indicator of whether any AE/SAE occurred at least once;
- Intercurrent event(s): Dosing compliance or treatment discontinuation is not considered for analysis. All data collected (including data after discontinuation of treatment) are included.
- Population-level summary: Incidence rate of AEs/SAEs, defined as the number of participants with AEs/SAEs from the time of first dose to Week 16 divided by the number of participants, in each treatment arm.

2.1.2. Secondary Estimand(s)

The secondary estimand 1 is the treatment effect of abrocitinib based on IGA response. This estimand is defined according to the secondary objective and in alignment with the secondary endpoint response based on IGA score. It includes the following 4 attributes:

- Population: Patients aged 12 years and older with moderate to severe AD, as defined by the inclusion and exclusion criteria, and who are randomized;
- Variable: Response based on IGA score of clear (0) or almost clear (1) and greater than or equal to 2 points improvement from baseline at Week 12; participants who discontinued intervention for any reason or required a rescue medication are considered as non-responders;
- Intercurrent event: The intercurrent event "intervention discontinuation for any reason" is addressed in the variable definition;
- Population-level summary: IGA response rate with 95% confidence interval (CI), in each treatment arm.

The secondary estimand 2 is the treatment effect of abrocitinib based on EASI-75 response. This estimand is defined according to the secondary objective and in alignment with the secondary endpoint response based on EASI-75. It includes the following 4 attributes:

- Population: Patients aged 12 years and older with moderate to severe AD, as defined by the inclusion and exclusion criteria, and who are randomized;
- Variable: Response based on improvement of $\geq$75% from baseline EASI score at Week 12; participants who discontinued intervention for any reason or required a rescue medication are considered as non-responders;
- Intercurrent event: The intercurrent event "intervention discontinuation for any reason" is addressed in the variable definition;
- Population-level summary: EASI-75 response rate with 95% CI, in each treatment arm.

The secondary estimand 3 is the treatment effect of abrocitinib based on SCORAD75 response. This estimand is defined according to the secondary objective and in alignment with the secondary endpoint response based on SCORAD75. It includes the following 4 attributes:

- Population: Patients aged 12 years and older with moderate to severe AD, as defined by the inclusion and exclusion criteria, and who are randomized;

- Variable: Response based on improvement of ≥75% from baseline SCORAD total score at Week 12; participants who discontinued intervention for any reason or required a rescue medication are considered as non-responders;

- Intercurrent event: The intercurrent event "intervention discontinuation for any reason" is addressed in the variable definition;

- Population-level summary: SCORAD75 response rate with 95% CI, in each treatment arm.

The secondary estimand 4 is defined according to the secondary objective and in alignment with the secondary endpoint change from baseline in POEM scores. It includes the following 4 attributes:

- Population: Patients aged 12 years and older with moderate to severe AD, as defined by the inclusion and exclusion criteria, and who are randomized;

- Variable: Change from baseline in POEM scores at Week 12 and at all scheduled time points; data after treatment discontinuation or use of rescue medication at any time during the treatment period will be censored;

- Intercurrent event: The intercurrent event "intervention discontinuation for any reason" is addressed in the variable definition;

- Population-level summary: Mean and standard deviation (SD), in each treatment arm.

The secondary estimand 5 is defined according to the secondary objective and in alignment with the secondary endpoint change from baseline in ADCT scores. It includes the following 4 attributes:

- Population: Patients aged 12 years and older with moderate to severe AD, as defined by the inclusion and exclusion criteria, and who are randomized;

- Variable: Change from baseline in ADCT scores at Week 12 and at all scheduled time points; data after treatment discontinuation or use of rescue medication at any time during the treatment period will be censored;

- Intercurrent event: The intercurrent event "intervention discontinuation for any reason" is addressed in the variable definition;

- Population-level summary: Mean and SD, in each treatment arm.

2.1.3. Additional Estimand(s)

The primary substudy estimand is the incidence of bone safety findings in knee MRI after 1 year of being exposed to abrocitinib in adolescent participants 12 to <18 years of age, regardless of dosing compliance or treatment discontinuation. This estimand is defined according to the primary substudy objective and is in alignment with the primary substudy endpoint. It includes the following 4 attributes:

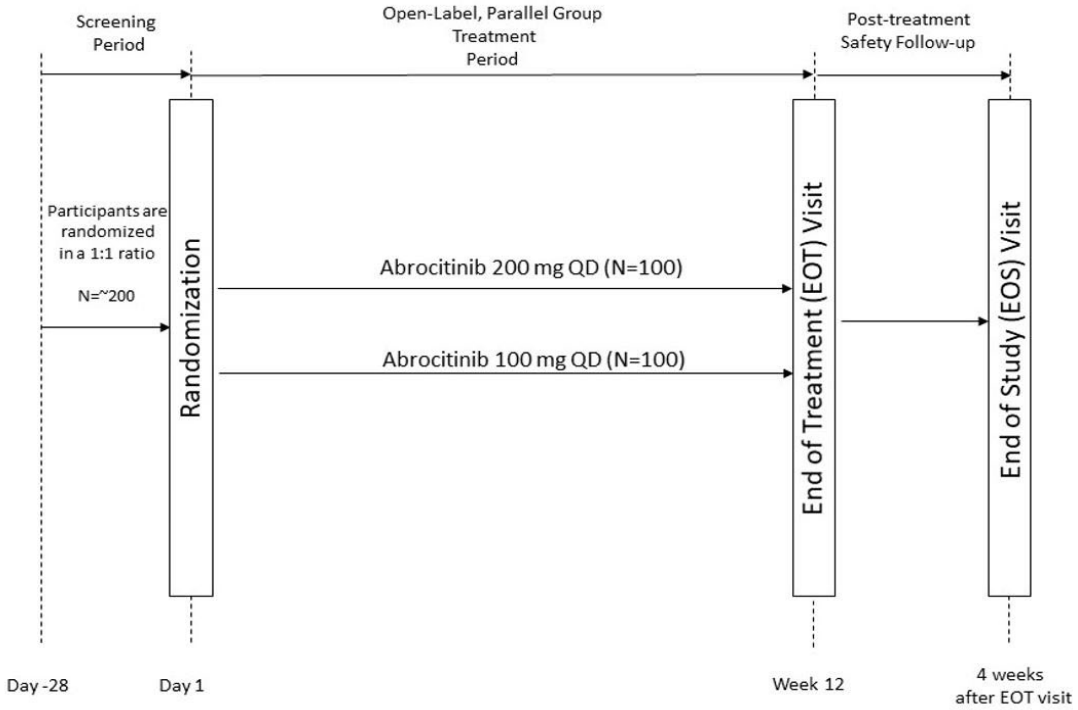
- Population: Patients aged 12 to <18 years with moderate to severe AD, as defined by the substudy inclusion and exclusion criteria;
- Variable: Binary indicator of whether the participant had bone safety findings in knee MRI 1 year after randomization;
- Intercurrent event(s): Dosing compliance or treatment discontinuation is not considered for analysis. All data collected (including data after discontinuation of treatment) are included.
- Population-level summary: Incidence rate of bone safety findings, defined as the number of participants with bone safety findings in knee MRI 1 year after randomization divided by the number of participants that received treatment during the substudy, in each treatment arm.

2.2. Study Design

This is a randomized, open-label, parallel-group study to evaluate the safety and efficacy of abrocitinib 100 mg QD and 200 mg QD tablets in participants aged 12 years and older with moderate to severe AD in India. There is a planned treatment duration of 12 weeks, with 4 weeks of off-treatment safety follow up thereafter. Study participants will be screened within 28 days prior to the first dose of investigational product to confirm that they meet the eligibility criteria for the study.

This study protocol also includes a substudy evaluating whether abrocitinib has any potential effects on adolescent bone with regard to bone safety findings in knee MRI. Adolescent participants (12 to < 18 years of age) will continue to receive investigational product until 1 year after randomization into the main study.

Figure 1. Study Design (Main Study)



3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint(s)

The primary endpoint is the incidence of AEs/SAEs, defined as the number of participants with treatment emergent adverse events, divided by the number of participants in the Safety Population (Section 4).

An adverse event will be considered treatment emergent relative to a given treatment if:

- The event starts on or after the first dose of study drug, but before the last dose plus the lag time of 28 days. The algorithm will not consider any events that started prior to the first dose date. If an AE starts on the same day as the first dose date, it will be considered treatment emergent.

  or

- The event was seen prior to the start of study drug but increased in severity during the effective duration of treatment.

3.2. Secondary Endpoint(s)

3.2.1. Response based on IGA Score

The Investigator's Global Assessment of AD is scored on a 5-point scale (0-4), reflecting a global consideration of the erythema, induration and scaling (Appendix 2.2). Response is defined as a Week 12 score of clear (0) or almost clear (1), which also translates to a reduction of at least 2 points from baseline.

3.2.2. Response based on EASI-75

The EASI quantifies the severity of a subject's atopic dermatitis based on both severity of lesion clinical signs and the percent of body surface area (BSA) affected (Appendix 2.3). Response is defined as improvement of at least 75% from baseline in the EASI total score at Week 12.

3.2.3. Response based on SCORAD75

The SCORAD is a validated scoring index for AD, which combines extent (0-100), severity (0-18), and subjective symptoms (0-20) based on pruritus and sleep loss, each scored using a visual analog scale (0-10). The SCORAD total score for an individual ranges from 0 to 103 (Appendix 2.4). The extent score is set to missing if any of the body areas are missing. The severity score is set to missing if severity is missing for any of the subjective symptoms. The subjective symptom score is set to missing if either the itch score or the sleeplessness score is missing. The total score is set to missing if any of the 3 components is missing. Response is defined as improvement of at least 75% from baseline in the SCORAD total score at Week 12.

3.2.4. Change from baseline in POEM Score

The POEM is a validated 7-item patient-reported outcome (PRO) measure used to assess the impact of AD recalled over the past week.[1-3] Each item is scored from 0 (No days) to 4 (Every day). The POEM score is derived as the sum of scores from all 7 POEM items. The range of POEM score is 0 to 28, with higher scores indicating greater symptom burden. If one item is unanswered, this is scored as 0. If two or more items are unanswered, the POEM score is set to missing. Change from baseline in POEM score is derived for all scheduled time points. A positive mean change score indicates worsening of symptom burden over time.

3.2.5. Change from baseline in ADCT Score

The ADCT is a validated 6-item PRO measure used to assess patient-perceived AD control.[4, 5] Each item is scored 0 to 4. The ADCT score is derived as the sum of scores from all 6 ADCT items. The range of ADCT score is 0 to 24, with higher scores indicating lower AD control. If one item is unanswered, the ADCT score is set to missing. Change from baseline in ADCT score is derived for all scheduled time points. A positive mean change score indicates worsening of AD control over time.

3.3. Other Endpoint(s)

The following additional endpoints will be derived for participants included in the substudy.

3.3.1. Incidence of Bone Safety Findings

The primary substudy endpoint is the incidence of bone safety findings in knee MRI after 1 year of being exposed to abrocitinib in adolescent participants 12 to <18 years of age. Bone safety findings will be adjudicated by the Safety Adjudication Committee.

3.4. Baseline Variables

For all safety and efficacy analyses, baseline value is defined as the result of the last assessment performed prior to first dose of study treatment.

3.5. Safety Endpoints

3.5.1. Adverse Events

See Section 3.1.

3.5.2. Laboratory Data

Below is a list safety laboratory data that will be collected in the study. Change from baseline will be derived for laboratory parameters with scheduled post-baseline assessments.

Table 2. Protocol-Required Safety Laboratory Assessments

Hematology	Chemistry	Lipid Profile[a]	Other
Hemoglobin WBC count Neutrophils (%, Abs) Lymphocytes (%, Abs) Platelets Coagulation Panel (APTT, PT/INR)	Urea sCr sCysC (at Screening) Estimated Creatinine Clearance AST ALT TBili Alkaline phosphatase	Total cholesterol LDL VLDL HDL Triglycerides	At screening only: • FSH[b] • Serum pregnancy test (β-hCG, for WOCBP only)[c] • HBsAg • HIV • HCV • TB test (as per site standards)

a. Fasting lipid profile panel require at least an 8 hour fast. If possible, participants should comply with fasting requirements at those visits detailed above.
b. For confirmation of postmenopausal status only.
c. Urine pregnancy test (for WOCBP only) will be performed at all visits other than screening visits.

3.5.3. Vital Signs

Vital signs data include blood pressure and pulse rate.

3.5.4. Electrocardiograms

Electrocardiograms (ECG) parameters to be summarized include heart rate, PR, QT, QTcF and QRS.

4. ANALYSIS SETS (POPULATIONS FOR ANALYSIS)

Population	Description
Enrolled	All participants who sign the informed consent document.
Full analysis set (FAS)	All participants randomly assigned to study intervention and who take at least 1 dose of study intervention. Participants will be analyzed according to the treatment arm they are randomized to.
Safety	All participants randomly assigned to study intervention and who take at least 1 dose of study intervention. Participants will be analyzed according to the treatment they actually received.
Substudy analysis set (SAS)	All participants who entered the substudy and who take at least 1 dose of study intervention during the substudy.

For safety analyses, participants will be analyzed according to the treatment they actually received. On the other hand, for efficacy analyses, participants will be analyzed according to the treatment arm they are randomized to.

5. GENERAL METHODOLOGY AND CONVENTIONS

The study will be analyzed and reported once all randomized participants have completed all scheduled procedures (or discontinued) in the main study, i.e. at study participant data set release after the completion of the main study. The substudy will be analyzed and reported after the substudy data set release following last participant last visit in the substudy. The substudy report will include analyses that are discussed in Sections 6.1 (using the SAS), 6.3, 6.5, and Appendix 2.5.

5.1. Hypotheses and Decision Rules

This is a descriptive study, therefore there are no decision rules and statistical hypothesis testing to be performed. The objective of the study is to estimate safety and efficacy parameters, separately for each treatment arm.

5.2. General Methods

5.2.1. Analyses for Binary Endpoints

Binary endpoints (response based on IGA, EASI-75 and SCORAD75 and incidence of bone safety findings) will be presented using proportions together with the 95% confidence

interval calculated using the normal approximation (or the Clopper-Pearson exact method when the estimated proportion is 0 or 1). Incidence of adverse events will be presented using frequency counts and percentages.

5.2.2. Analyses for Continuous Endpoints

Continuous endpoints will be summarized using descriptive statistics, including the number of non-missing observations, mean, SD, median, first and third quartiles, minimum and maximum values. Descriptive summaries of the observed values and change from baseline for each time point will be provided, for each treatment arm.

5.2.3. Analyses for Categorical Endpoints

Categorical endpoints will be summarized by frequency counts and percentages. Unless otherwise specified, the calculation of percentages will include the missing category. Therefore, counts of missing observations will be included in the denominator and presented as a separate category.

5.3. Methods to Manage Missing Data

For the analysis of efficacy endpoints using descriptive statistics, missing values will not be imputed. For the analysis of safety endpoints, the sponsor data standard rules for imputation according to Clinical Data Interchange Standards Consortium (CDISC) and Pfizer Standards (CaPS) will be applied.

6. ANALYSES AND SUMMARIES

6.1. Primary Endpoint(s)

6.1.1. Incidence of AEs/SAEs

- Estimand strategy: Treatment policy (Section 2.1.1).
- Analysis set: Safety (Section 4).
- Analysis methodology: Descriptive statistics.
- Intercurrent events and missing data: All adverse event data collected will be analyzed regardless of dosing compliance or treatment discontinuation; missing values will not be imputed.
- The number and proportion of participants who had treatment emergent adverse events will be presented for each treatment arm. Treatment emergent adverse events will be summarized according to type, frequency, severity, seriousness and relationship to abrocitinib in each treatment arm.

6.2. Secondary Endpoint(s)

6.2.1. Response based on IGA Score

- Estimand strategy: Composite (Section 2.1.2).
- Analysis set: FAS (Section 4).
- Analysis methodology: Descriptive statistics and 95% CIs.
- Intercurrent events and missing data: Missing data due to participant discontinuing treatment or from the study for any reason are considered non-response. In addition, participants who required a rescue medication are considered non-responders.
- The response rate along with the 95% CI will be estimated for each treatment arm. In addition, summary of IGA scores for all scheduled time points will be presented for each treatment arm.

6.2.2. Response based on EASI-75

- Estimand strategy: Composite (Section 2.1.2).
- Analysis set: FAS (Section 4).
- Analysis methodology: Descriptive statistics and 95% CIs.
- Intercurrent events and missing data: Missing data due to participant discontinuing treatment or from the study for any reason are considered non-response. In addition, participants who required a rescue medication are considered non-responders.
- The response rate along with the 95% CI will be estimated for each treatment arm. In addition, summary of EASI scores for all scheduled time points will be presented for each treatment arm.

6.2.3. Response based on SCORAD75

- Estimand strategy: Composite (Section 2.1.2).
- Analysis set: FAS (Section 4).
- Analysis methodology: Descriptive statistics and 95% CIs.
- Intercurrent events and missing data: Missing data due to participant discontinuing treatment or from the study for any reason are considered non-response. In addition, participants who required a rescue medication are considered non-responders.

- The response rate along with the 95% CI will be estimated for each treatment arm. In addition, summary of SCORAD total scores for all scheduled time points will be presented for each treatment arm.

6.2.4. Change from baseline in POEM Score

- Estimand strategy: While on treatment (Section 2.1.2).
- Analysis set: FAS (Section 4).
- Analysis methodology: Descriptive statistics.
- Intercurrent events and missing data: Data collected after treatment discontinuation or use of rescue medication at any time during the treatment period will be censored; missing values will not be imputed.
- Summary statistics for observed value and change from baseline in POEM Score will be presented for all scheduled time points, for each treatment arm.

6.2.5. Change from baseline in ADCT Score

- Estimand strategy: While on treatment (Section 2.1.2).
- Analysis set: FAS (Section 4).
- Analysis methodology: Descriptive statistics.
- Intercurrent events and missing data: Data collected after treatment discontinuation or use of rescue medication at any time during the treatment period will be censored; missing values will not be imputed.
- Summary statistics for observed value and change from baseline in ADCT Score will be presented for all scheduled time points, for each treatment arm.

6.3. Other Endpoint(s)

6.3.1. Incidence of bone safety findings in knee MRI

- Estimand strategy: Treatment policy (Section 2.1.3).
- Analysis set: SAS (Section 4).
- Analysis methodology: Descriptive statistics and 95% CIs.
- Intercurrent events and missing data: All data collected will be analyzed regardless of dosing compliance or treatment discontinuation; missing values will not be imputed.

- The incidence rate of bone safety findings in knee MRI, along with the 95% CI will be estimated for baseline and End of Treatment (EOT) in the substudy, for each treatment arm. In addition, the types of bone safety findings will be summarized.

6.4. Subset Analyses

Summary statistics for the efficacy endpoints (Section 6.2) will be presented by subgroups according to whether the participants received monotherapy or combination therapy based on the permitted concomitant AD therapies in the protocol.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

Demographic, baseline disease characteristics, and medical history will be summarized by treatment arm according to CaPS.

6.5.2. Study Conduct and Participant Disposition

Participants' evaluation, disposition, and discontinuation will be summarized by treatment arm according to CaPS.

6.5.3. Study Treatment Exposure

A summary of dosing compliance by treatment arm will be provided. Dosing compliance will be calculated as the percent of the number of doses of study drug the participant took out of the expected total number of doses. Number and percent of participants with compliance of <80 % and >120% will be provided for each treatment arm.

The exposure to study drug (computed as last dosing date – first dosing date + 1) will be summarized by treatment arm and according to duration categories (<1 week, ≥1 week to <4 weeks, ≥4 weeks to < 8 weeks, ≥8 weeks to < 10 weeks, and ≥10 weeks).

6.5.4. Concomitant Medications and Nondrug Treatments

Prior drug and non-drug treatments, concomitant drug and non-drug treatments will be summarized for each treatment arm according to CaPS.

6.6. Safety Summaries and Analyses

All safety analyses will be performed on the Safety population. These will be presented in a tabular format and summarized descriptively for each treatment arm following Pfizer standards as appropriate.

6.6.1. Adverse Events

Medical Dictionary for Regulatory Activities (MedDRA) will be used to classify all AEs with respect to system organ class and preferred term. Summaries will include treatment

emergent AEs/SAEs, withdrawal from active treatment due to AEs, and serious infections, defined as any infection (viral, bacterial, and fungal) requiring hospitalization or parenteral antimicrobials or meets other criteria that require it to be classified as a serious adverse event.

Any AEs that occur after signing Informed Consent and before randomization and the first dose of study drug will be listed but not summarized.

6.6.2. Laboratory Data

Laboratory data will be listed and summarized according to CaPS. Laboratory parameters will be summarized descriptively at baseline, Week 4 and EOT visits, and will include the observed and change from baseline values.

6.6.3. Vital Signs

Vital signs will be summarized descriptively over time. Summaries will include the observed and change from baseline values. Height and weight will be summarized at baseline.

6.6.4. Electrocardiograms

ECG parameters will be summarized at baseline and EOT visits and will include the observed and change from baseline values.

7. INTERIM ANALYSES

7.1. Introduction

No formal interim analysis will be conducted for this study. As this is an open-label study, the sponsor may conduct reviews of the data during the course of the study for the purpose of safety assessment. Interim safety and efficacy data after approximately 50% of the total sample size has been randomized in the study will be prepared and summarized to support presentation of initial results to India Bureau of Health (BoH). Additional summaries of interim data, including MRI data may be prepared to support presentation of results to regulatory agencies.

7.2. Interim Analyses and Summaries

Data package to be provided to India BoH for presentation of initial results will include derived CDISC Study Data Tabulation Model (SDTM) datasets and data summaries that will include descriptive statistics of collected data on safety and efficacy parameters, as well as demographic and baseline characteristics.

8. APPENDICES

Appendix 1. Summary of Analyses

Endpoint	Population	Data Inclusion and Rules for Handling Intercurrent Events and Missing Data	Analysis Method
Incidence of AEs/SAEs	Safety	All data collected will be included regardless of intercurrent events. Missing data will not be imputed.	Descriptive statistics
Response based on IGA Score	FAS	Missing data due to discontinuation of treatment or from the study or use of rescue medication will be considered as non-response.	Descriptive statistics; point and interval estimation
Response based on EASI-75	FAS	Missing data due to discontinuation of treatment or from the study or use of rescue medication will be considered as non-response.	Descriptive statistics; point and interval estimation
Response based on SCORAD75	FAS	Missing data due to discontinuation of treatment or from the study or use of rescue medication will be considered as non-response.	Descriptive statistics; point and interval estimation
Change from baseline in POEM Score	FAS	Data collected after treatment discontinuation or use of rescue medication will be censored. Missing data will not be imputed.	Descriptive statistics
Change from baseline in ADCT Score	FAS	Data collected after treatment discontinuation or use of rescue medication will be censored. Missing data will not be imputed.	Descriptive statistics

Incidence of bone safety findings in knee MRI	Substudy Analysis Set	All data collected will be included regardless of intercurrent events. Missing data will not be imputed.	Descriptive statistics; point and interval estimation

Appendix 2. Data Derivation Details

Appendix 2.1. Definition and Use of Visit Windows in Reporting

Visit windows will be used for efficacy variables.

Visit Label	Target Day	Definition [Day window]
Baseline	Day 1 (Day of first dose)	Last observation prior to and including day of first dose
Week 2	15	Days 2 to 22
Week 4	29	Days 23 to 43
Week 8	57	Days 44 to 71
Week 12	85	Days 72 to 99
Substudy EOT[a]	366	Days 353 to 380 or observation collected at Early Termination
Substudy Follow-up[b]	394	Days 381 to 408 or observation collected after Early Termination

a. For MRI and HSS Pedi-FABS assessments in the Substudy
b. For HSS Pedi-FABS assessments in the Substudy

If two or more assessments fall into the same window, the assessment closest to the Target Day will be used. If two assessments are equally distant from the Target Day in absolute value, the later assessment will be used.

Appendix 2.2. Investigator's Global Assessment (IGA) Scores

The Investigator's Global Assessment of AD is scored on a 5-point scale (0-4), reflecting a global consideration of the erythema, induration and scaling. The clinical evaluator of AD will perform an assessment of the overall severity of atopic dermatitis and assign an IGA

score and category as described in Table 3. The assessment will be a static evaluation without regard to the score at a previous visit.

Table 3. Investigator's Global Assessment (IGA) Scores

Score	Category	Description
0	Clear	Atopic dermatitis is cleared, except for any residual discoloration (post-inflammatory hyperpigmentation and/or hypopigmentation).
1	Almost Clear	Overall, the atopic dermatitis is not entirely cleared and remaining lesions are light pink (not including post inflammatory hyperpigmentation) and/or; have barely palpable hard thickened skin and/or papules and/or; have barely perceptible lichenification; excoriation and oozing/crusting are absent.
2	Mild	Overall, the atopic dermatitis consists of lesions that are light red; with slight, but definite hard thickened skin and/or papules; with slight, but definite linear or picked scratch marks or penetrating surface injury; with slight, but definite thickened skin, fine skin markings, and lichenoid scale; oozing/crusting is absent.
3	Moderate	Overall, the atopic dermatitis consists of lesions that are red; with easily palpable moderate hard thickened skin and/or papules; with moderate linear or picked scratch marks or penetrating surface injury; with moderate thickened skin, coarse skin markings, and coarse lichenoid scale; with slight oozing/crusting.
4	Severe	Overall, the atopic dermatitis consists of lesions that are deep, dark red; with severe hard thickened skin and/or papules; with severe linear or picked scratch marks or penetrating surface injury; with severe thickened skin with very coarse skin markings and lichenoid scale; with moderate to severe oozing/crusting.

Appendix 2.3. Eczema Area and Severity Index

The EASI quantifies the severity of a participant's AD based on both severity of lesion clinical signs and the percent of BSA affected. EASI is a composite scoring by the AD clinical evaluator of the degree of erythema, induration/papulation, excoriation, and lichenification (each scored separately) for each of four body regions, with adjustment for the percent of BSA involved for each body region and for the proportion of the body region to the whole body.

Lesion Severity by Clinical Signs: The basic characteristics of AD lesions - erythema, induration/papulation, excoriation, and lichenification-provide a means for assessing the

severity of lesions. Assessment of these four main clinical signs is performed separately for four body regions: head and neck, upper limbs, trunk (including axillae and groin) and lower limbs (including buttocks). Average erythema, induration/papulation, excoriation, and lichenification are scored for each body region according to a 4-point scale: 0 = absent; 1 = mild; 2 = moderate; 3 = severe. Morphologic descriptors for each clinical sign severity score are shown in Table 4.

Table 4. Clinical Sign Severity Scoring Criteria for the Eczema Area and Severity Index (EASI)

Score		Description
Erythema (E)		
0	Absent	None; may have residual discoloration (post-inflammatory hyperpigmentation and/or hypopigmentation).
1	Mild	Light pink to light red
2	Moderate	Red
3	Severe	Deep, dark red
Induration/Papulation (I)		
0	Absent	None
1	Mild	Barely palpable to slight, but definite hard thickened skin and/or papules
2	Moderate	Easily palpable moderate hard thickened skin and/or papules
3	Severe	Severe hard thickened skin and/or papules
Excoriation (Ex)		
0	Absent	None
1	Mild	Slight, but definite linear or picked scratch marks or penetrating surface injury
2	Moderate	Moderate linear or picked scratch marks or penetrating surface injury
3	Severe	Severe linear or picked scratch marks or penetrating surface injury
Lichenification (L)		
0	Absent	None
1	Mild	Barely perceptible to slight, but definite thickened skin, fine skin markings, and lichenoid scale
2	Moderate	Moderate thickened skin, coarse skin markings, and coarse lichenoid scale
3	Severe	Severe thickened skin with very coarse skin markings and lichenoid scale

Percent BSA with Atopic Dermatitis: The number of handprints of skin afflicted with AD in a body region can be used to determine the extent (%) to which a body region is involved

with AD (Table 5). When measuring, the handprint unit refers to the size of each individual participant's hand with fingers in a closed position.

Table 5. Handprint Determination of Body Surface Area (BSA)

Body Region	Total Number of Handprints in Body Region*	Surface Area of Body Region Equivalent of One Handprint*
Head and Neck	10	10%
Upper Limbs	20	5%
Trunk (including axillae and groin/genitals)	30	3.33%
Lower Limbs (including buttocks)	40	2.5%

*Handprint refers to the hand size of each individual participant.

The extent (%) to which each of the four body regions is involved with AD is categorized to a numerical Area Score using a non-linear scaling method according to the following BSA scoring criteria (Table 6).

Table 6. Eczema Area and Severity Index (EASI) Area Score Criteria

Percent BSA with Atopic Dermatitis in a Body Region	Area Score
0%	0
>0 - <10%	1
10 - <30%	2
30 - <50%	3
50 - <70%	4
70 - <90%	5
90 - 100%	6

Body Region Weighting: Each body region is weighted according to its approximate percentage of the whole body (Table 7).

Table 7. Eczema Area and Severity Index (EASI) Body Region Weighting

Body Region	Body Region Weighting
Head and Neck	0.1
Upper Limbs	0.2
Trunk (including axillae and groin/genitals)	0.3
Lower Limbs (including buttocks)	0.4

In each body region, the sum of the Clinical Signs Severity Scores for erythema, induration/papulation, excoriation, and lichenification is multiplied by the Area Score and by the Body Region Weighting to provide a body region value, which is then summed across all four body regions resulting in an EASI score as described in Equation 3.

Equation 3: EASI = 0.1Ah(Eh+Ih+Exh+Lh) + 0.2Au(Eu+Iu+ExU+Lu) + 0.3At(Et+It+Ext+Lt) + 0.4Al(El+Il+Exl+Ll)

A = Area Score; E = erythema; I = induration/papulation; Ex = excoriation; L = lichenification; h = head and neck; u = upper limbs; t = trunk; l = lower limbs

The EASI score can vary in increments of 0.1 and range from 0.0 to 72.0, with higher scores representing greater severity of AD.

The EASI Clinical Sign Severity Scores for each of the 4 body regions and the involvement by total number of handprints for each of the 4 body regions will be recorded on the CRF.

Appendix 2.4. Scoring Atopic Dermatitis

SCORAD is a validated scoring index for atopic dermatitis, which combines extent (0-100), severity (0-18), and subjective symptoms (0-20) based on pruritus and sleep loss, each scored using a visual analog scale (0-10).

Extent (A, maximum score of 100%)

To determine extent of AD, rule of 9 is used to calculate body surface area affected by AD as a percentage of the whole-body surface area. Body surface area as percentage of total body surface area for each body region is as follows:

- Head and neck 9%;
- Upper limbs 9% each;
- Lower limbs 18% each;
- Anterior trunk 18%;
- Back 18%;
- 1% for genitals.

The score for each body region is added up to determine the BSA affected by AD (A), which has a possible maximum score of 100%.

Severity (B, maximum score of 18)

A representative area of AD is selected. In this area, the severity of each of the following signs is assessed as none (0), mild (1), moderate (2) or severe (3).

- Erythema (reddening);
- Edema (swelling)/papulation;
- Oozing/crusting;
- Excoriation (scratch marks);
- Skin thickening (lichenification);
- Xerosis (dryness) (this is assessed in an area where there is no inflammation).

The severity scores are added together to give 'B' (maximum score of 18).

Subjective Symptoms (C, maximum score of 20)

Subjective symptoms, ie, itch and sleep loss, are each scored by the participant or caregiver using a visual analog scale (VAS) where “0” is no itch (or no sleep loss) and “10” is the worst imaginable itch (or sleep loss). The value for each should reflect the average on a 10-point scale for the last 3 days/nights. These scores are added to give 'C' (maximum score of 20).

SCORAD Total Score

The SCORAD for an individual is calculated by the formula: A/5 + 7B/2 + C (can range from 0 to 103).

Appendix 2.5. HSS Pedi-FABS

The Hospital for Special Surgery Pediatric Functional Activity Brief Scale (HSS Pedi-FABS) is a validated 8-item instrument designed to quantify the activity level of children between 10 and 18 years of age.[6,7] Two versions of the instrument based on the recall period will be implemented (1 month recall and 1 year recall). Items regarding running, cutting, decelerating, pivoting, duration, and endurance are scored 0 to 4, while items regarding completion and supervision are scored 0 to 3. The HSS Pedi-FABS score is derived as the sum of scores from all 8 items. The range of HSS Pedi-FABS score is 0 to 30, with higher

scores indicating more physical activity. If one item is unanswered, the HSS Pedi-FABS score is set to missing.

Summary statistics for HSS Pedi-FABS score will be presented for all scheduled time points, for each treatment arm.

Appendix 3. List of Abbreviations

Abbreviation	Term
Abs	absolute
AD	Atopic Dermatitis
ADCT	Atopic Dermatitis Control Tool
AE	adverse event
ALT	alanine aminotransferase
APTT	Activated Partial Thromboplastin Time
AST	aspartate aminotransferase
β-hCG	beta-human chorionic gonadotropin
BoH	Bureau of Health
BSA	body surface area
CaPS	CDISC and Pfizer Standards
CDISC	Clinical Data Interchange Standards Consortium
CI	confidence interval
CRF	Case Report Form
EASI	Eczema Area and Severity Index
EASI-75	≥75% improvement from baseline in the Eczema Area and Severity Index total score
ECG	Electrocardiogram
EOT	End of Treatment
FAS	Full analysis set
FSH	follicle-stimulating hormone
HBsAg	hepatitis B surface antigen
HCV	hepatitis C virus
HDL	high-density lipoprotein
HIV	human immunodeficiency virus
HSS Pedi-FABS	Hospital for Special Surgery Pediatric Functional Activity Brief Scale
IGA	Investigator's Global Assessment
INR	international normalized ratio
LDL	low-density lipoprotein
MedDRA	Medical Dictionary for Regulatory Activities
MRI	magnetic resonance imaging
POEM	Patient-Oriented Eczema Measure

Abbreviation	Term
PR	pulse rate
PRO	patient-report outcome
PT	prothrombin time
QD	once daily
QTcF	corrected QT (Fridericia method)
SAE	serious adverse event
SAP	statistical analysis plan
SAS	Substudy Analysis Set
SCORAD	SCORing Atopic Dermatitis
SCORAD75	≥75% improvement from baseline in the Scoring Atopic Dermatitis total score
sCr	serum creatinine
sCysC	serum Cystatin C
SD	standard deviation
SDTM	Study Data Tabulation Model
TB	tuberculosis
TBili	total bilirubin
VAS	Visual Analogue Scale
VLDL	very-low-density lipoprotein
WBC	white blood cell
WOCBP	woman/women of childbearing potential

Appendix 4. References

1. Charman C, Venn A, 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(12)(Dec):1513-9.

2. Charman CR, Venn AJ, Ravenscroft JC, et al.. Translating Patient-Oriented Eczema Measure (POEM) scores into clinical practice by suggesting severity strata derived using anchor-based methods. Br J Dermatol 2013;169(6)(Dec):1326-32.

3. Spuls PI, Gerbens LAA, Simpson E, et al. Patient-Oriented Eczema Measure (POEM), a core instrument to measure symptoms in clinical trials: a Harmonising Outcome Measures for Eczema (HOME) statement. Br J Dermatol 2017;176(4)(Apr):979-84.

4. Pariser DM, Simpson EL, Gadkari A, et al. Evaluating patient-perceived control of atopic dermatitis: design, validation, and scoring of the Atopic Dermatitis Control Tool (ADCT). Curr Med Res Opin. 2020;36(3):367-376.

5. Simpson E, Eckert L, Gadkari A, et al. Validation of the Atopic Dermatitis Control Tool (ADCT©) using a longitudinal survey of biologic-treated patients with atopic dermatitis. BMC Dermatol. 2019 Nov 6;19(1):15.

6. Fabricant PD, Suryavanshi JR, Calcei JG, et al. The Hospital for Special Surgery Pediatric Functional Activity Brief Scale (HSS Pedi-FABS): Normative Data. Am J Sports Med. 2018 Apr;46(5):1228-34.

7. Fabricant PD, Robles A, Downey-Zayas T, et al. Development and Validation of a Pediatric Sports Activity Rating Scale: The Hospital for Special Surgery Pediatric Functional Activity Brief Scale (HSS Pedi-FABS). The American Journal of Sports Medicine. 2013;41(10):2421-9.