I4V-MC-JAIP Statistical Analysis Plan v3

# A Phase 3, Multicenter, Randomized, Double-blind, Placebo-controlled, Parallel-group, Outpatient Study Evaluating the Pharmacokinetics, Efficacy, and Safety of Baricitinib in Pediatric Patients with Moderate-to-Severe Atopic Dermatitis

NCT03952559

Approval Date: 01-Jun-2022

# 1. Statistical Analysis Plan: I4V-MC-JAIP(b): A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Outpatient Study Evaluating the Pharmacokinetics, Efficacy, and Safety of Baricitinib in Pediatric Patients with Moderate to Severe Atopic Dermatitis

# BREEZE-AD-PEDS



Baricitinib (LY3009104) Atopic Dermatitis

Study I4V-MC-JAIP (JAIP) is a Phase 3, multicenter, randomized, double-blind, placebo-controlled, parallel-group, outpatient, 131-week study designed to evaluate the pharmacokinetics, efficacy, and safety of baricitinib 1 mg, 2 mg, and 4 mg in pediatric patients with moderate to severe atopic dermatitis.


Eli Lilly and Company
Indianapolis, Indiana USA 46285


Protocol I4V-MC-JAIP
Phase 3

Statistical Analysis Plan V1 electronically signed and approved by Lilly: 14-Feb-2020
Statistical Analysis Plan V2 electronically signed and approved by Lilly: 25-May-2022
Statistical Analysis Plan V3 electronically signed and approved by Lilly on date provided below.

Approval Date: 01-Jun-2022 GMT

# 2. Table of Contents

**Section** **Page**

## Table of Contents

## Table of Contents

**Figure** **Page**

## Table of Contents

**Appendix** **Page**

# 3. Revision History

Statistical Analysis Plan (SAP) Version 1 is based on Protocol I4V-MC-JAIP(a) and was approved prior to the production transfer for the first data monitoring committee (DMC) meeting.

Statistical Analysis Plan (SAP) Version 2 is based on Protocol I4V-MC-JAIP(b) and was approved prior to the Week 16 primary outcome data base lock.

Statistical Analysis Plan (SAP) Version 3 is based on Protocol I4V-MC-JAIP(b) and was approved prior to the Week 16 primary outcome data base lock. The summary of changes between Version 2 and Version 3 are as follows:

| Section | Summary of Changes |
|---|---|
| **Table JAIP.6.2** | Updated the baseline definition and 16-week analysis window for imaging data |
| **Table JAIP.6.6** | Updated total EASI score algorithm for patients 0 to <8 years old |

# 4. Study Objectives

Table JAIP.4.1 shows the objectives and endpoints of the study.

**Table JAIP.4.1. Objectives and Endpoints**

| Objectives | Endpoints |
|---|---|
| **Primary** | |
| Primary Objective for Double-blind Treatment Period<br>• To demonstrate the superiority of each dose of baricitinib versus placebo in the treatment of patients with moderate-to-severe AD. | • Proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement at Week 16 |
| Primary Objective for PK lead-in Period<br>• To assess whether baricitinib exposure in pediatric patients receiving baricitinib high dose once daily is comparable to the exposure in adults receiving baricitinib 4 mg once daily. | • Comparability will be assessed using non-compartmental methods (e.g., AUC and $C_{max}$) |
| **Key Secondary**<br>*These are prespecified objectives that will be adjusted for multiplicity* | |
| To compare the efficacy of baricitinib high, medium, or low dose to placebo in AD during the 16-week double-blind placebo-controlled treatment period as measured by improvement in signs and symptoms of AD. | • Proportion of patients achieving EASI75 at 16 weeks<br>• Proportion of patients achieving EASI90 at 16 weeks<br>• Mean change from baseline in EASI score at 16 weeks<br>• Proportion of patients achieving SCORAD75 at 16 weeks |
| To compare the efficacy of baricitinib high, medium, or low dose to placebo in AD during the 16-week double-blind placebo-controlled treatment period as assessed by patient-reported outcome measures. | • Proportions of patients achieving a 4-point improvement in Itch NRS at 1 week, 2 weeks, 4 weeks, and 16 weeks for patients ≥10 years old |
| **Other Secondary**<br>*These are prespecified objectives that will not be adjusted for multiplicity.* | |
| To compare the efficacy of baricitinib high, medium, or low dose to placebo in AD during the 16-week double-blind placebo-controlled period as measured by physician-assessed signs and symptoms of AD. | • Proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement at Week 4<br>• Proportion of patients achieving EASI50 at 16 weeks<br>• Proportion of patients achieving IGA of 0 at 16 weeks<br>• Mean change from baseline in SCORAD at 16 weeks<br>• Mean percent change from baseline in SCORAD at 16 weeks<br>• Proportion of patients achieving SCORAD90 at 16 weeks<br>• Mean percent change from baseline in EASI score at 16 weeks |

| Objectives | Endpoints |
|---|---|
| | • Mean change from baseline in BSA affected at 16 weeks<br>• Proportion of patients developing skin infections requiring antibiotic treatment by Week 16 |
| To compare the efficacy of baricitinib high, medium, or low dose to placebo in AD during the 16-week, double-blind, placebo-controlled treatment period as assessed by patient-reported outcome/QoL measures. | • Mean number of days without use of background TCS over 16 weeks<br>• Mean gram quantity of TCS used over 16 weeks (tube weights)<br>• Mean change from baseline in Itch NRS at 1 week, 4 weeks and 16 weeks for patients ≥10 years old<br>• Mean percent change from baseline in Itch NRS at 1 week, 4 weeks and 16 weeks for patients ≥10 years old<br>• Mean change in the PRISM at 1 week, 2 weeks, 4 weeks, and 16 weeks for patients <10 years old<br>• Mean change from baseline in the total score of the POEM at 16 weeks<br>• Mean change in PGI-S-AD scores at 16 weeks<br>• Mean change from baseline in the PROMIS-pediatric depression at 16 weeks<br>• Mean change from baseline in the PROMIS-pediatric anxiety at 16 weeks<br>• Mean change from baseline in DFI at 16 weeks<br>• Mean change in CDLQI score at 16 weeks<br>• Mean change in IDQOL score at 16 weeks<br>• Mean change in WPAI-AD-CG scores at 16 weeks<br>• Mean change in EQ-5D-Y scores at 16 weeks<br>• Mean change from baseline in the score of Item 2 of the ADSS at 1 week and 16 weeks for patients ≥10 years old<br>• Mean change from baseline in Skin Pain NRS at 16 weeks for patients ≥10 years old |
| To assess the patient acceptability and palatability of baricitinib tablets and oral suspension. | • Assessment of tablet or oral suspension product acceptability and palatability during the Open-label PK Lead-in period |
| To characterize the pharmacokinetic profile of the baricitinib in pediatric patients with AD. | • Population PK Analysis based on sparse sampling over 16 weeks (Study Period 3) with secondary endpoints including $C_{max}$, AUC, and $t_{1/2}$ |
| To evaluate the potential effects of baricitinib on the cellular and humoral immune system. | • Change of IgG titers from pre-vaccination to 4 weeks and 12 weeks post vaccination in patients eligible for vaccination with tetanus, diphtheria, and acellular pertussis (TDaP) |

| Objectives | Endpoints |
|---|---|
| | and/or pneumococcal conjugate vaccine according to local guidelines |
| To assess efficacy of baricitinib during longer-term treatment. | • Proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement at 1 and 2 years<br>• Proportion of patients achieving EASI75 at 1 year<br>• Proportion of patients achieving SCORAD75 at 1 year |
| To assess growth and bone safety of baricitinib during longer-term treatment. | • Mean changes in growth (height and weight), growth velocity, and bone age over the course of treatment during the long-term extension treatment period. |
| **Tertiary/Exploratory** | |
| • Frequency of patient-reported "no itch" (Itch NRS score = 0) days from daily diaries from Week 12 to Week 16<br>• Frequency of patient-reported "no pain" (Skin Pain NRS score = 0) days from daily diaries from Week 12 to Week 16<br>• Time to achieve a 4-point improvement in Itch NRS<br>• Mean change from baseline in the score of Item 1 of the ADSS at 1 week and 16 weeks<br>• Mean change from baseline in the score of Item 3 of the ADSS at 1 week and 16 weeks<br>• To evaluate changes from baseline in IgE levels during the study<br>• To evaluate changes from baseline in eosinophil levels during the study<br>• To characterize baricitinib pharmacokinetics in the AD population and explore relationships between baricitinib exposure and study endpoints<br>• Assessment of efficacy outcomes in patients who choose to take a voluntary drug interruption (drug holiday) during Study Period 4<br>• Number of patients able to maintain control of AD signs and symptoms without use of TCS<br>• Proportion of patients achieving IGA of 0 or 1 with a ⩾2-point improvement at 2, 3, and 4 years during long-term extension<br>• Proportion of patients achieving EASI75 at 2, 3, and 4 years during long-term extension | |

Abbreviations: AD = atopic dermatitis; ADSS = Atopic Dermatitis Sleep Scale; AUC = area under the drug plasma concentration versus time curve; BSA = body surface area; CDLQI = Children's Dermatology Life Quality Index; $C_{max}$ = maximum concentration; DFI = Dermatology Family Impact; EASI75/90 = 75%/90% response rate on the Eczema Area and Severity Index; EQ-5D-Y = the European Quality of Life–5 Dimensions–Youth; IDQOL = Infant's Dermatitis Quality of Life Index; IGA = Investigator's Global Assessment; IgE = immunoglobulin E; IgG = immunoglobulin G; NRS = numeric rating scale; PK = pharmacokinetic; QoL = quality of life; PGI-S-AD = Patient Global Impression of Severity–Atopic Dermatitis; POEM = Patient-Oriented Eczema Measure; PRISM = Parent-Reported Itch Severity Measure; PROMIS = Patient-Reported Outcomes Measurement Information System; $t_{1/2}$ = half-life; TCS = topical corticosteroids; TDaP = tetanus, diphtheria toxoids, and acellular pertussis (vaccine); SCORAD75 = 75% response rate on the SCORing Atopic Dermatitis; WPAI-AD-CG = Work Productivity and Activity Impairment: Atopic Dermatitis – Caregiver.

# 5. Study Design

## 5.1. Summary of Study Design

Study I4V-MC-JAIP (JAIP) is a Phase 3, multicenter, randomized, double-blind, placebo-controlled, parallel-group, outpatient study evaluating the pharmacokinetics (PK), efficacy, and safety of baricitinib compared with placebo in pediatric patients with moderate-to-severe atopic dermatitis (AD). The study is divided into 5 periods: a 5-week Screening period; a 2-week Open-label PK lead-in period; a 16-week Double-Blind Treatment period; a 4-year Long-term Extension period; and a 4-week Post-treatment Follow-up period.

**Period 1: Screening Period:** between 8 and 35 days prior to Week 0 (Visit 2)

**Period 2: Open-label PK Lead-in Period:** 2 week period to evaluate comparability of exposure to baricitinib high dose between pediatric patients and adults.

**Period 3: Double-blind, Placebo-controlled Treatment Period**: from Week 0 (Visit 2) through Week 16 (Visit 8). Patients participating in Study Period 3 will continue using Sponsor-provided low- and medium-potency topical corticosteroids (TCS) for use as clinically indicated and determined by the investigator.

**Period 4: Long-term Extension Treatment Period:**

a. PK Lead-in Patients:

   After completing the PK lead-in (Visit 4), patients may proceed to the Long-term Extension period (Visit 9) 2 weeks after completing Visit 4 and continue to receive open-label baricitinib at the same dose they received in the PK lead-in. Patients may continue in the long-term extension for an additional 4 years.

b. Non-PK Lead-in Patients

   Patients who participate in Study Period 3 and complete through Week 16 (Visit 8) will be eligible to continue in the Long-term Extension period for up to 4 additional years of treatment. At Visit 8 (primary endpoint and end of the double-blind period), patients will be transitioned into the long-term extension treatment period as follows:

   - Responders (Investigator's Global Assessment [IGA] 0/1/2) who have not required rescue with topical treatments or systemic treatments during Study Period 3 will continue on the Double-blind Treatment to which they were randomized at Visit 2 (Week 0).
   - Nonresponders (IGA 3/4) or those patients requiring rescue with topical treatments or systemic treatments during Study Period 3 will be transitioned to open-label baricitinib at the high dose for their age group.

During Study Period 4, patients are allowed to use TCS (all potencies), topical calcineurin inhibitors (TCNI), and/or a phosphodiesterase 4 (PDE-4) inhibitor as

background treatments with investigative product (IP) if they experience worsening or lack of control of their AD according to the clinical protocol.

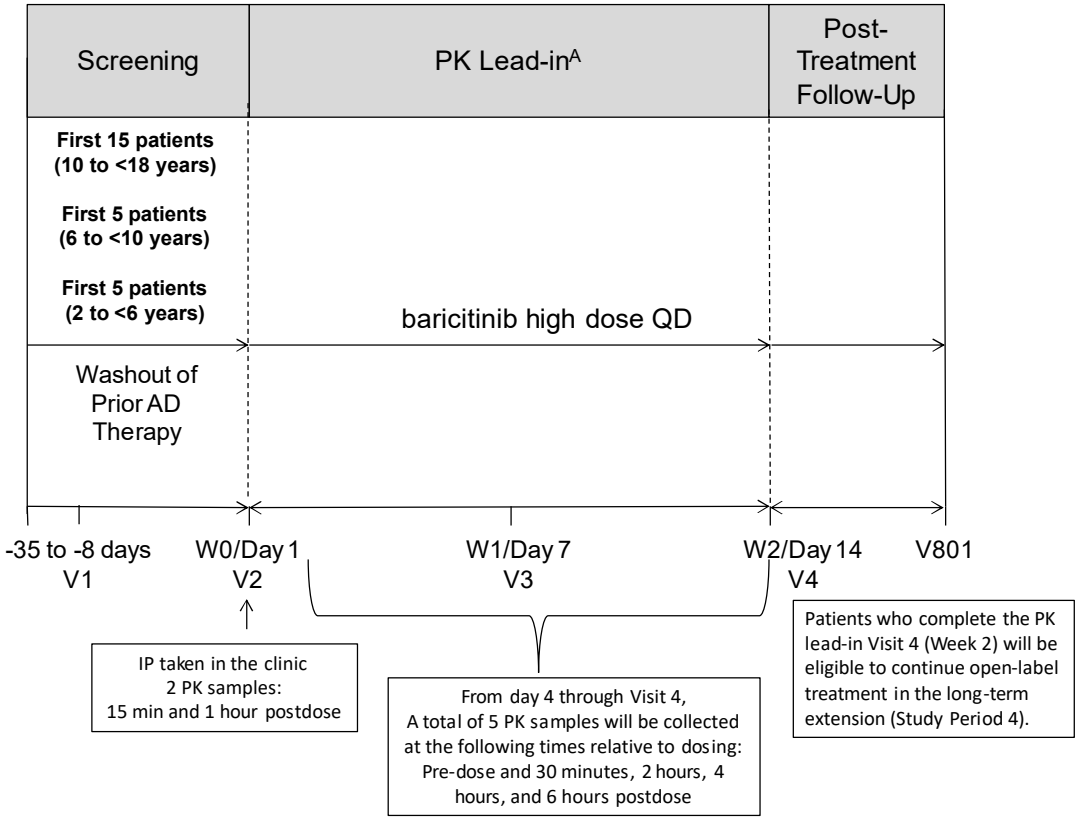
During Study Period 4, for patients on double-blind treatment whose IGA worsens to 3 or 4 and who are unable to recapture an IGA response of 0, 1, or 2, despite the use of emollients and TCS, the patient may be transitioned at the discretion of the investigator to open-label baricitinib at the high dose for their age group.

After the first year of the extension treatment period, treatment and transition to open-label baricitinib will continue as in the first year; however, patients will be allowed to voluntarily interrupt IP treatment after Visit 15 provided they continue to complete all other study visit procedures per protocol.

**Period 5: Post-Treatment Follow-Up Period:** from last treatment period visit or early termination visit (ETV) to approximately 28 days after the last dose of investigational product for patients who have completed either Study Period 2 or 3 and who do not enter the long-term extension period (Study Period 4).

Approximately 465 patients 2 to <18 years of age who have responded inadequately to or who are intolerant to topical therapy will be enrolled into the study, of which, approximately 25 patients will be enrolled into the Open-label PK Lead-in period (Study Period 2). The remaining patients (at least 440 patients) will be randomized at a 1:1:1:1 ratio to receive placebo once daily (QD), baricitinib low dose QD, baricitinib medium dose QD, or baricitinib high dose QD (110 patients in each treatment group with at least 320 patients 10 to <18 years old and at least 120 patients 2 to <10 years old).

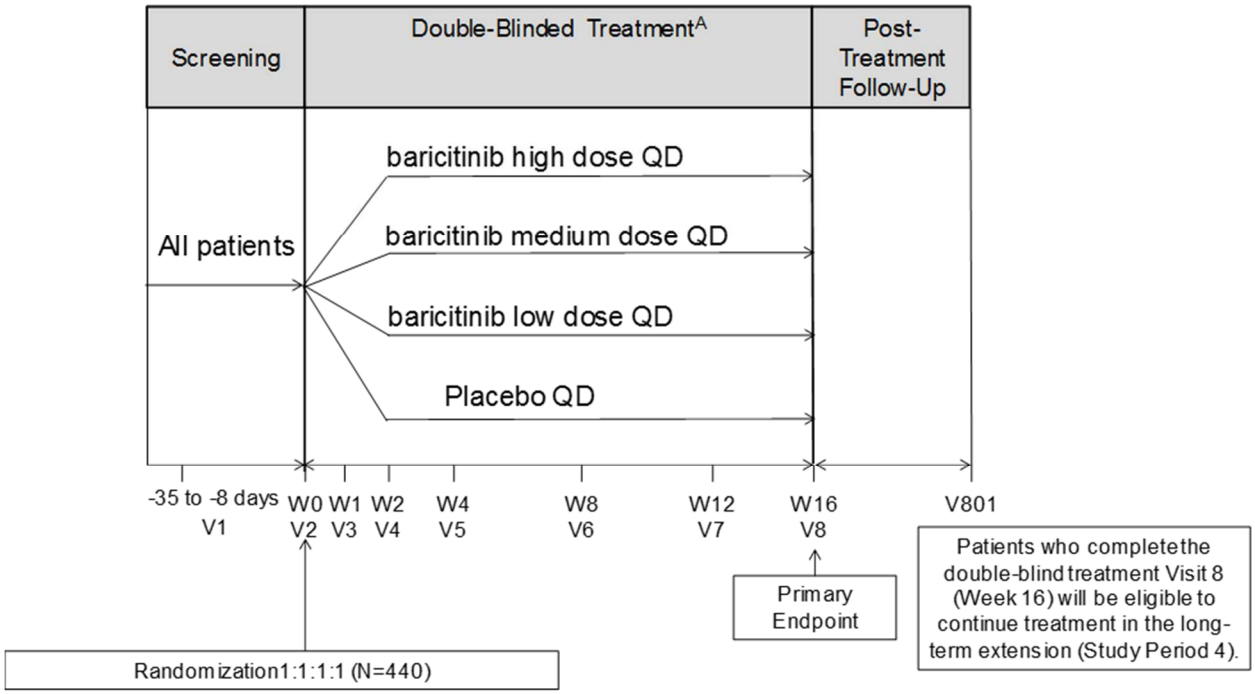
Figure JAIP.5.1, Figure JAIP.5.2, and Figure JAIP.5.3 illustrate the study design. The blinding procedure is described in the Protocol.




Abbreviations: AD = atopic dermatitis; PBPK = Physiologically Based Pharmacokinetic; PK = pharmacokinetic; QD = once daily; V = visit; W = week.

A Based on PBPK modelling results, the high dose for patients 10 to <18 years old is 4 mg, and the high dose for patients 2 to <10 years is 2 mg.

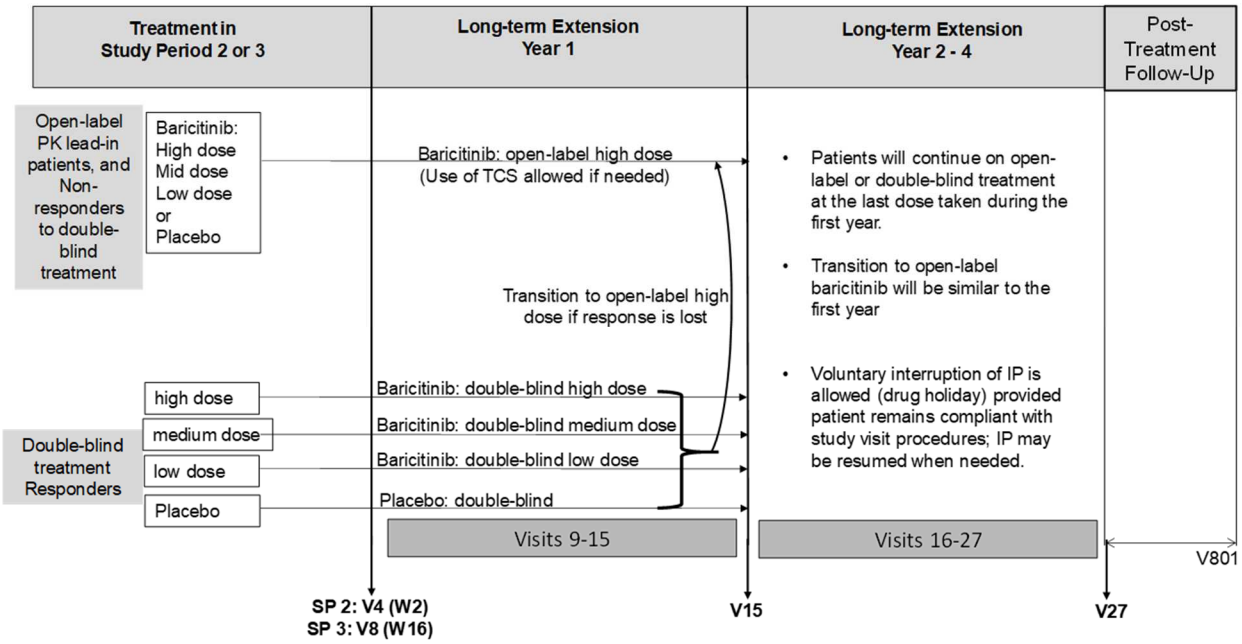
**Figure JAIP.5.1. Illustration of study design for Clinical Protocol I4V-MC-JAIP (PK Lead-in, Study Period 2).**




Abbreviations: PK = pharmacokinetic; QD = once daily; V = visit; W = week.

[A] Dosing for the Double-blind Treatment period will be confirmed through the PK lead-in. The anticipated daily baricitinib doses for patients 10 to <18 years old are 4 mg, 2 mg, and 1 mg. The anticipated daily baricitinib doses for patients 2 to <10 years old are 2 mg, 1 mg, and 0.5 mg.

**Figure JAIP.5.2. Illustration of study design for Clinical Protocol I4V-MC-JAIP (Double-Blinded Treatment, Study Period 3).**




Abbreviations: IP = investigational product; PK = pharmacokinetic; SP = study period; TCS = topical corticosteroids; V = visit; W = Week.

**Figure JAIP.5.3. Illustration of study design for Clinical Protocol I4V-MC-JAIP (Long-term Extension, Study Period 4).**

## 5.2. Method of Assignment to Treatment

Patients participating in the PK lead-in who meet all criteria for enrollment will receive open label baricitinib at the high dose for their age group beginning at Visit 2. Patients not participating in the PK lead-in who meet all criteria for enrollment will be randomized in a 1:1:1:1 ratio (placebo, baricitinib low dose; baricitinib medium dose; baricitinib high dose) to double-blind treatment at Visit 2. Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive web-response system (IWRS). The IWRS will be used to assign blister packs or bottles containing double-blind IP to each patient according to the schedule of activities. Site personnel will confirm that they have located the correct blister packs or bottles by entering a confirmation number found on the blister packs or bottles into the IWRS. Patients will be stratified at randomization according to disease severity (IGA 3 versus 4) and geographic region.

This study will be conducted internationally in multiple sites. Table JAIP.5.1 describes how regions will be defined for stratification. Regions may be combined for statistical analyses. The two region strata with the least number of patients may be pooled. Analysis methods for maximized extended enrollment (ME2) patients will follow the methods described in Section 6.

**Table JAIP.5.1. Geographic Regions for Stratification**

| Region | Country |
|---|---|
| Europe (EU) | Austria, Czech Republic, France, Germany, Hungary, Poland, Spain, Switzerland |
| Japan (JP) | Japan |
| Rest of World (ROW) | Argentina, Australia, Brazil, Israel, Mexico, Russia, Taiwan, United Kingdom, India |

# 6. A Priori Statistical Methods

## 6.1. Determination of Sample Size

Approximately 25 patients will be enrolled into the Open-label PK Lead-in (Study Period 2) and may continue on open-label treatment during the long-term extension (Study Period 4). Data from patients participating in the PK lead-in will be analyzed separately from patients randomized into the Double-blind treatment (Study Period 3).

Study JAIP will aim to enroll at least 440 patients 2 to <18 years of age into the Double-blind Treatment period (Study Period 3), which includes at least 320 older pediatric patients (10 to <18 years) and at least 120 younger pediatric patients (2 to <10 years). The proposed sample size (N=440) will ensure a >90% power to detect any difference between the baricitinib high dose and placebo treatment groups or the baricitinib medium dose and placebo treatment groups, each using a 2-sided alpha of 0.05, assuming a 10% placebo, 25% medium dose, and 30% high dose response rate for the primary endpoint using a chi-squared test. The assumptions are based on what was observed in the Phase 2 study in adults with AD (Study I4V-MC-JAHG [JAHG]). The proposed end point of IGA 0 or 1 represents patients whose AD is clear or almost clear from a baseline of moderate or severe disease. The anticipated effect size represents 3 times more patients achieving this benefit compared to placebo which, in discussion with therapeutic experts, is of a magnitude that is considered to be clinically relevant.

Furthermore, in older pediatric patients the sample size of 320 is sufficient to detect that the baricitinib high or medium dose is superior to placebo at least 80% of the time. Similarly, in the younger pediatric patients, the planned sample size of 120 patients has >80% simulated power using the Bayesian approach described in Section 6.17.3 to detect any difference between the baricitinib high dose or medium dose and placebo treatment groups with a probability threshold of 0.95.

Sample size estimates were calculated using nQuery® Advisor 7.0 for the older subgroup of patients, and power estimates were obtained from R 3.5.0 and JAGS 4.2.0 for the younger subgroup of patients.

## 6.2. General Considerations

This plan describes *a priori* statistical analyses for efficacy, health outcomes, and safety that will be performed.

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly). The primary and key secondary statistical analyses will be performed using SAS® Version 9.4 or higher.

Not all displays described in this SAP will necessarily be included in the Clinical Study Report (CSR). Not all displays will necessarily be created as a “static” display. Some may be incorporated into interactive display tools instead of or in addition to a static display

Statistical tests of treatment effects and CIs will be performed at a 2-sided significance level of 0.05, unless otherwise stated (e.g., graphical multiple testing strategy in Section 6.3).

Data collected at early termination visits will be mapped to the next scheduled visit number for that patient if it falls within the visit window as discussed in Section 6.2.2. For by-visit summaries, only visits in which a measure was scheduled to be collected will be summarized. Applicable unscheduled visit data will be included in patient-level listings, and the data will still also be used in other analyses, including shift analyses for safety analytes, change from baseline to endpoint using modified last observation carried forward (mLOCF) for efficacy analyses, and other categorical analyses including safety.

### *6.2.1. Analysis Populations*

Patients participating in the PK lead-in (PKLI) period will be analyzed separately from patients participating in the double blind treatment period. PK lead-in patients will not be included in primary efficacy analyses. For efficacy analyses in the long-term extension period, data from ITT population defined below will be summarized.

For long-term safety analyses, PK lead-in patients and patients participating in the double blind treatment period will be pooled and analyzed. Safety analysis will be made using Extended Bari and All Bari population defined as the following.

The following major analysis populations will be used.

**PK lead-in (PKLI) population**: all patients who received at least 1 dose of IP in the PKLI period (Study period 2).

**Intent-to-treat (ITT) population**: all randomized patients in the double-blind treatment period (Study period 3).

**Per-protocol Set (PPS):** The PPS of the ITT population analysis set will include those patients who do not have any important protocol violations. Qualifications for and identification of significant or important protocol violations will be determined while the study remains blinded, prior to database lock.

**Follow-up population:** The follow-up population is defined as patients who entered the follow-up period and who do not enter the long-term extension study.

Unless otherwise specified, the efficacy analyses will be conducted on the ITT population during the double blind period (Period 3) (Gillings and Koch 1991), which seeks to preserve the benefits of randomization and avoid the issue of selection bias. Patients will be analyzed according to the treatment to which they were randomized. In addition, the primary and key secondary analyses will be repeated using the PPS population.

**Week 16 Responders and Partial Responders:** randomized patients who have achieved a response of IGA 0, 1, or 2 at Visit 8 (Week 16) without requiring rescue with topical treatments or systemic treatments during Study Period 3, continued on the Double-blind Treatment to which

they were randomized at Visit 2 (Week 0) and received at least 1 dose of IP in the long-term extension period.

**Week 16 Non-responders:** randomized patients who have not achieved a response (i.e., IGA of ≥3) or who have required rescue with topical treatments (i.e., high-/ultra-high-potency TCS) or systemic treatments during Study Period 3, transitioned to open-label baricitinib at the high dose for their age group and received at least 1 dose of IP in the long-term extension period.

The long-term efficacy analyses will be made using W16 Responders and Partial Responders, and non-responders. The following are the treatment groups and analysis period.

**Table JAIP.6.1. Definition of Analysis Population for Long-term Efficacy Analysis**

| Analysis Population | Analysis Period | Treatment Group | Description | Censoring Rule | Inferential Comparisons |
|---|---|---|---|---|---|
| W16 Responders and Partial Responders | Study Period 4 | Placebo | Randomized to placebo at Week 0 | Data censored at dose switch or permanent drug discontinuation | Descriptive statistics without formal comparison |
| | | Low dose | Randomized to low dose at Week 0 | | |
| | | Medium dose | Randomized to medium dose at Week 0 | | |
| | | High dose | Randomized to high dose at Week 0 | | |
| | Study Period 4 while patients are on open-label high dose Bari | Placebo to high Dose | Randomized to placebo at Week 0 and switch Bari high dose in SP4 | Data censored at permanent drug discontinuation | |
| | | Low dose to high dose | Randomized to Bari low dose at Week 0 and switch Bari high dose in SP4 | | |
| | | Medium dose to high dose | Randomized to Bari medium dose at Week 0 and switch Bari high dose in SP4 | | |
| | | High dose to high dose | Randomized to Bari high dose at Week 0 and switch Bari high dose in SP4 | | |
| W16 Non-responders | Study Period 4 | Placebo to high Dose | Randomized to placebo at Week 0 and switch Bari high dose in SP4 | Data censored at dose switch or permanent drug discontinuation | |
| | | Low dose to high dose | Randomized to Bari low dose at Week 0 and switch Bari high dose in SP4 | | |
| | | Medium dose to high dose | Randomized to Bari medium dose at Week 0 and switch Bari high dose in SP4 | | |
| | | High dose to high dose | Randomized to Bari high dose at Week 0 and switch Bari high dose in SP4 | | |

**Safety population**: all randomized patients who receive at least 1 dose of IP and who did not discontinue from the study for the reason 'Lost to Follow-up' at the first postbaseline visit. This definition excludes patients with no safety assessments postbaseline so that incidence percentages are not underestimated.

**All Bari Population:** all patients (including PK lead-in patients) who received at least 1 dose of Baricitinib at any time during the study, either during Period 2 or Period 3, or after transitioning from PBO to baricitinib in Period 4.

**Extended Bari Population:** all patients who were randomized and received at least 1 dose of Baricitinib (low, medium, high) or PBO originally assigned in Period 3.

For Japan submission, All Bari Population and Extended Bari Population will exclude patients who were randomized to Bari low dose at week 0, other populations will be subset to Japan subpopulation.

For the double-blind treatment period safety analyses will be performed using the safety population. Patients will be analyzed according to the dosing regimen to which they were assigned. Analyses of the safety endpoints, many of which are incidence based, will include all patients in the safety population, unless specifically stated otherwise. In the rare situation where a patient is Lost to Follow-up at the first postbaseline visit, but some safety data exists (e.g., unscheduled laboratory assessments) after first dose of study drug, a listing of the data or a patient profile will be provided, when requested. The following are the treatment groups and analysis for the safety analysis in double-blind treatment period based on safety population:

**Table JAIP.6.2. Analysis Population and Analysis for Safety Analysis in Study Period 3**

| Population and Treatment Groups | Topics | Baseline Time Period | Postbaseline Time Period | Inferential Comparisons |
|---|---|---|---|---|
| Safety Population: Placebo, Bari low dose, Bari medium dose and Bari high dose, Pooled Bari (Randomized to low dose, medium dose, or high dose and received at least one dose of study drug) | TEAEs including events of special topics | From screening to just before first study drug administration | First day of study drug administration up to and including 16 weeks of treatment (16-week visit) for ongoing patients or up to 30 days after the last dose date of the study drug (not exceeding the Week 16 visit) for patients who discontinued drug prior to the 16-week visit | All possible comparisons between Bari high, medium, low and PBO. |
| | AEs leading to study drug discontinuation, AEs leading to study drug interruption, and SAEs | Not applicable | | |
| | Treatment-emergent (TE) low and high summaries in labs and vital signs, CTCAE and NCEP shift summaries in labs, TE/shift in C-SSRS and HADS, and change from minimum or maximum baseline to minimum or maximum postbaseline measurements in labs and vital signs | The entire period (including scheduled "planned" and unscheduled "unplanned" measurements) up to date of first study drug administration (including baseline pre-dose measurements) | Post-first dose of study drug measurements up to and including 16 weeks of treatment (16-week visit) for ongoing patients or up to 30 days after the last dose date of the study drug (not exceeding the Week 16 visit) for patients who discontinued drug prior to the 16-week visit. | |
| | Analyses by time point, change from last baseline to last postbaseline observation in labs, vital signs and growth parameters. | Last scheduled "planned" measurements prior to or at date of first study drug dose administration<br><br>For imaging data, last scheduled "planned" measurements prior to or at date of first study drug dose administration + 7 days | Post-first dose of study drug "planned" measurements up to and including the 16-week visit excluding the follow-up visit. The early termination visit is considered a "planned" visit.<br><br>For imaging data on 16-week visit, analysis window will be extended to include measurements up to 30 days prior to 16-week visit date. | |

The long-term safety analysis will be performed using All BARI and Extended Bari population. Dose comparisons will be made between Bari high, medium, low in the Extended Bari population.

## Table JAIP.6.3. Analysis Population and Analysis for Long-term Safety Analysis

| Analysis Population and Treatment Groups | Description | Topic | Baseline Time Period | Postbaseline Time Period |
|---|---|---|---|---|
| All BARI(including all doses).<br><br>No censoring of data at dose change to open-label high dose. (i.e. for double-blind patients who are randomized to low, medium or high dose at Week 0, if they transition to open-label high dose, the safety data after dose transition will still be included into safety analysis) | Includes all patients who received at least 1 dose of baricitinib at any time during the study, either during Period 2 or 3, or after transitioning from PBO to baricitinib in Period 4. | TEAEs including events in safety topics of special interest | Randomized/ enrolled to receive BARI: The entire period up to date of first BARI administration,<br><br>Randomized to PBO: The period prior to initiation of BARI (it means any onset after start of BARI or ongoing events that increased in severity after start of BARI are TEAEs) | First day of BARI administration up to data cut date or up to 30 days after the last dose date of BARI (not exceeding the data cut date) |
| | | SAEs | Not applicable | |
| | | TE low and high summaries in labs and vital signs | Randomized/ enrolled to receive BARI: The entire period including “planned” and “unplanned” measurements up to date of first BARI administration (including baseline pre-dose measurements) Randomized to PBO: The last non-missing “planned” or “unplanned” measurement prior to initiation of BARI | Post-first BARI dose measurements (including “planned” and “unplanned” measurements) up to data cut date or up to 30 days after the last dose date of BARI (not exceeding the data cut date) |
| Extended Bari<br><br>(including all doses)<br><br>Censoring will occur at dose switching to open-label high dose Bari. | Includes all patients who were randomized and received at least 1 dose of Baricitinib (low, medium, high) or PBO originally assigned in Period 3 IP in the long-term extension period and have never changed their randomized treatment throughout the study. | TEAEs including events in safety topics of special interest | The entire period up to date of IP administration<br>TEAE: Any onset after start of IP or ongoing events that increased in severity after start of Bari | First day of IP administration up to data cut date or up to 30 days after the last dose date of IP (not exceeding the data cut date) or up to dose change, whichever occurs the first. |
| | | SAEs | Not applicable | |
| | | TE low and high summaries in labs | The entire period including “planned” and “unplanned” measurements up to date of first IP administration | Post-first IP dose measurements (including “planned” and “unplanned” measurements) up to data cut date or up to 30 days after the last dose date of IP (not exceeding the data cut date) or up to dose change, whichever occurs the first. |

### 6.2.2. Definition of Baseline and Postbaseline Measures

*Baseline*
The baseline value for efficacy variables measured at scheduled visits is defined as the last non-missing measurement on or prior to the date of first study drug administration (expected at Week 0, Visit 2).

The baseline value for the daily diary assessments (Patient Global Impression of Severity – Atopic Dermatitis [PGI-S-AD], Itch Numeric Rating Scale [NRS], Skin Pain NRS, Atopic Dermatitis Sleep Scale [ADSS], Parent-Reported Itch Severity Measure [PRISM], missed school days, and TCS use) is the mean of the non-missing assessments in the 7 days prior to the date of first study drug administration (expected at Week 0, Visit 2). Criteria for derivation of the baseline score requires there be at least 4 non-missing measurements in the 7 days indicated; otherwise, an expanded window of up to 14 days prior to first study drug administration, if available, may be utilized in order to obtain the most recent 4 non-missing measurements prior to first study drug administration. If there are not at least 4 non-missing measurements collected prior to the date of the first study drug administration using the aforementioned method, then the baseline will be designated as missing. Baseline for the safety analyses is defined in Section 6.2.1 for different populations.

*Postbaseline*
Postbaseline measurements are collected after study drug administration through Week 16 (Visit 8) or early discontinuation visit. Non-missing efficacy data collected at scheduled visits (e.g., electronic clinical outcome assessment [eCOA], clinical-reported outcome [ClinRO]) will be used for analyses. If data for a scheduled visit are missing, data from proximal unscheduled visits, if available, will be used if they fall within visit windows as follows: a ±2-day window is applied to Visits 3 through 5 (Weeks 1, 2, 4), a ±4-day window is applied to Visits 6 through 8 (Weeks 8, 12, 16). If there is more than 1 unscheduled visit within the defined visit window and no scheduled visit is available, the unscheduled visit closest to the scheduled visit date will be used. If two unscheduled visits of equal distance are available, then the later of the two visits will be used. If there is no non-missing measure collected at the scheduled visit or at an unscheduled visit falling within the specified visit window, the measure is considered missing for that scheduled visit.

Note that during exceptional circumstances, during Study Period 3 (post-randomization and up to but not including the primary endpoint [Visit 8]), visit windows may be extended up to a total of 28 days.

Participants should complete primary endpoint visit (Visit 8/Study Period 3) and the final study endpoint visit (Visit 27/Study Period 4) as per original schedule whenever possible and safe to do so, at the investigator's discretion. However, in order to maximize the ability for such on-site visits, minimize missing data, and preserve the intended conduct of the study, the visit windows may be brought forward no sooner than 14 days or extended up to 28 days, upon specific guidance from the sponsor.

Postbaseline daily diary endpoints will be the mean of weekly visit windows (diary windows) anchored on day of first dose (Day 1) and day of Week 16 scheduled visit. Weeks 1-14 are defined as follows:

| Week | Days |
|---|---|
| 1 | 1-7 |
| 2 | 8-14 |
| 3 | 15-21 |
| 4 | 22-28 |
| 5 | 29-35 |
| 6 | 36-42 |
| 7 | 43-49 |
| 8 | 50-56 |
| 9 | 57-63 |
| 10 | 64-70 |
| 11 | 71-77 |
| 12 | 78-84 |
| 13 | 85-91 |
| 14 | 92-98 |

**Week 16 Daily Diary Window Construction**

The following sequential steps will be used to determine the Week 16 diary window. The general goal is to anchor on the scheduled Week 16 visit (or a proximal unscheduled visit) if such a visit exists or to use an interval based on days in study for cases where a scheduled Week 16 or a proximal surrogate does not exist.

Step 1: If the Week 16 scheduled visit exists, the Week 16 diary interval is the 7 days prior to the Week 16 date provided that window has at least 4 non-missing observations. If there are less than 4 non-missing observations, the diary window's lower bound will be extended 1 day at a time (up to day 99) to a maximum of 14 days prior to the Week 16 date until 4 non-missing observations are obtained. If, after extending this diary window's lower bound to 14 days, there are less than 4 non-missing observations then go to Step 3.

Step 2: If the Week 16 scheduled visit does not exist, the 7 days prior to the last visit (scheduled or unscheduled) occurring after Day 105, will constitute the Week 16 diary window provided that window contains at least 4 non-missing observations. If there are less than 4 non-missing

observations, the diary window's lower bound will be extended 1 day at a time (up to Day 99) to a maximum of 14 days prior to the unscheduled visit date until 4 non-missing observations are obtained. If, after extending this diary window's lower bound to 14 days, there are less than 4 non-missing observations then go to Step 3.

Step 3: If neither a Week 16 scheduled visit is available nor an unscheduled visit to act as a surrogate for the Week 16 diary window, then the Week 16 window will be Day 106 to Day 112. If there are less than 4 non-missing observations, the dairy window's lower bound will be extended 1 day at a time up to Day 99 until 4 non-missing observations are obtained.

If the steps above do not detect a window with at least 4 non-missing observations, then the Week 16 window is 7 days from either the Week 16 visit, the surrogate visit or Days 106 through 112 and the mean is missing and subject to imputation rules.

**Week 15 Daily Diary Window Construction**

The lower bound of Week 15 diary window is defined as Day 99. The upper bound of the Week 15 diary window is the minimum of either Day 105 or the lower bound of the Week 16 diary window -1. Consequently, Week 15 may be less than 4 days if the Week 16 scheduled visit is before Day 112. Moreover, as Week 15 diary window cannot exceed 7 days, there could be daily assessments between Weeks 15 and 16 diary windows that do not fall into a diary window. If after constructing the diary windows, there are fewer than 4 non-missing values the mean for that particular window is missing and subject to imputation rules.

### *6.2.3. Analysis Methods and Covariate Adjustment*

In Protocol I4V-MC-JAIP(b) Section 10.3.1 General Statistical Considerations, it is specified that

- Treatment comparison will be made using a logistic model with region, disease severity, age, treatment group, and treatment group-by-age interaction as covariates
- Treatment-by-age interaction will be added to the logistic regression model of the primary and key secondary variables as a sensitivity analysis.

The wording in Section 10.3.1 needs clarification. The intent of the wording above and also specifically the wording in Section 10.3.3.1 Primary Analysis, was that the main analysis method of categorical efficacy variables will use a logistic regression analysis with region, baseline disease severity (IGA), age cohort (older [10 to <18 years] and younger [2 to <10 years] cohort), treatment group in the model. The treatment-by-age interaction term will be used in additional models as sensitivity analyses (to estimate the treatment within each age cohort and also to check for interaction of treatment effect and the 2 age cohorts), see details in Section 6.17.

Firth's correction will be used in order to accommodate (potential) sparse response rates. The p-value for the odds ratio from the logistic regression model will be used for statistical inference, unless Firth's correction still results in quasi-separation. In that case, Fisher's exact test will be used for statistical inference. The percentages, difference in percentages, and 95% CI of the difference in percentages using the Newcombe-Wilson method without continuity correction will

be reported. The p-value from the Fisher's exact test will also be produced as a secondary analysis.

The main analysis method for all continuous efficacy variables will use mixed model repeated measures (MMRM) analysis. The MMRM model will use a restricted maximum likelihood (REML) estimation. The model will include treatment, age cohort (older [10 to <18 years] and younger [2 to <10 years] cohort), region, baseline disease severity (IGA), visit, treatment-by-visit-interaction, and treatment-by-age cohort interaction as fixed categorical effects and baseline score and baseline score-by-visit-interaction as fixed continuous effects. For daily diary assessments, the model for analyses up to Week 16 will include all weekly assessments. An unstructured covariance structure will be used to model the between- and within-patient errors. If this analysis fails to converge, the heterogeneous autoregressive [ARH(1)], followed by the heterogeneous compound symmetry (CSH), followed by the heterogeneous Toeplitz (TOEPH) , followed by autoregressive [AR(1)], followed by compound symmetry (CS) will be used. The Kenward-Roger method will be used to estimate the degrees of freedom. Treatment least squares means (LSM) will be estimated within the framework of the MMRM using type 3 sums of squares. Differences in LSM between each dose of BARI and placebo (and associated p-values, standard errors and 95% confidence intervals [CI]) will be used for statistical inference. The LSM difference, standard error, p-value and 95% CI will be reported.

Treatment comparisons for continuous efficacy variables may also be made using analysis of covariance (ANCOVA) for key secondary objectives. When an ANCOVA model is used, the model includes region, baseline disease severity, treatment group, age cohort (older [10 to <18 years] and younger [2 to <10 years] cohort), and baseline value. Inclusion of baseline in the ANCOVA models ensures treatment LSMs are estimated at the same value. Treatment LSM will be estimated within the framework of the ANCOVA using type 3 sums of squares. Reported differences in LSM and associated p-values, standard errors and 95% CI will be used for statistical inference.

For safety analysis in Study Period 3, Fisher's exact test will be used to test for differences between each baricitinib dose and placebo in proportions of patients experiencing adverse events (AEs), discontinuation from study drug, and for other categorical safety data. Continuous vital signs, body weight, and other continuous safety variables, including laboratory variables will be analyzed by an ANCOVA with treatment group, age cohort, and baseline value in the model. The significance of within-treatment group changes from baseline will be evaluated by testing whether or not the treatment group LSM changes from baseline are different from zero; the standard error for the LSM change will also be displayed. Differences in LSM will be displayed, with the p-value associated with the LSM comparison to placebo and a 95% CI on the LSM difference also provided. In addition to the LSMs for each group, the within-group p-value for the change from baseline will be displayed.

For long term safety analysis beyond Study Period 3 and additional safety analysis in Study Period 3 to support CSS, exposure adjusted incident rate (EAIR) will be reported. The IR evaluating the incidence of a first event per 100 PYR are provided. Exposure was calculated based on the time interval defined as the treatment period plus up to 30 days off-drug follow-up

time. Time at risk for a patient with an event was terminated at the time of the event. Time at risk for a patient without an event was terminated at the end of the analysis interval. A Poisson distribution 95% CI was calculated for IR.

For Study Period 3 safety analysis to support CSS only, Mantel-Haenszel odds ratios (OR) with investigated treatment as the numerator and reference treatment as the denominator will also be reported.

Specifically for long term safety analysis, following statistical methods will be used.

Exposure-adjusted IR:

- incidence rate difference (IRD) together with its 95% CI; the Mantel-Haenszel method was used for the IRD and its 95% CI calculation, and
- the incidence rate ratio (IRR) and its 95% CI derived from a Poisson regression model fitting the treatment as explanatory factors using time to event as an offset term may also be provided.

For laboratory analytes, vital signs measurements, and physical characteristics, IRs based on 100 PYE are provided to account for different durations of exposure.

**Covariate adjustment**

The randomization to treatment groups at Week 0 (Visit 2) is stratified by disease severity (IGA 3 vs 4) and geographic region as described in Section 5.2. Unless otherwise specified, the statistical analysis models will adjust for these stratification variables. The covariates used in the logistic model for categorical data will include the parameter value at baseline. The covariates used in the ANCOVA model for continuous data will include the parameter value at baseline. Inclusion of baseline in the model ensures treatment LSM are estimated at the same baseline value. When an MMRM analysis is performed, baseline value and baseline-by-visit interactions will be included as covariates.

### *6.2.4. Derived Data*

- Age (year), derived using the date of Visit 1 as the reference start date and the date of birth and kept in one decimal place. If only the birth month and year are collected, impute the 15$^{th}$ of the month for the date of birth. If birth month and date is missing while only birth year was collected, impute to July 1$^{st}$ as birth date and month.
- Age group (2 to <10, 10 to <18 years old)
- Age group (2 to <6, 6 to <10, 10 to <18 years old)
- Body mass index (BMI) (kg/m$^2$) = Weight (kg)/([Height (cm)/100]$^2$)
- BMI category (<14 kg/m$^2$, ≥14 to <20 kg/m$^2$, ≥20 to <25 kg/m$^2$, ≥25 to <30 kg/m$^2$, ≥30 kg/m$^2$)
- The duration of AD from diagnosis (years) = ([Date of informed consent – Date of AD diagnosis] + 1)/365.25.
    - If year of onset is missing, duration of AD will be set as missing. If diagnosis month and date are missing and diagnosis year is the same as birth year, birth

month and date will be taken. If diagnosis date is missing and diagnosis month and year are the same as birth month and year, the date of birth will be taken. Otherwise, unknown month will be taken as January, and unknown day will be taken as 01. The duration of AD will be rounded to 1 decimal place.

- Duration of AD (years) category (6 months to <1 year, 1 year to <2 years, 2 to <5 years, 5 to <10 years, ≥10 to <20 years)
- Diagnosis age (years), derived using diagnosis date as the reference start date and the date of birth and truncated to a whole-integer age. If only the birth month and year are collected, impute the 15$^{th}$ of the month for the date of birth.
- Diagnosis age group (<10, ≥10 years old)
- Change from baseline = postbaseline measurement at Visit x – baseline measurement.
    - If a baseline value is missing, it will not be imputed and the change from baseline will not be calculated.
- Percent change from baseline at Visit x:
  ([Postbaseline measurement at Visit x - Baseline measurement]/Baseline measurement)*100.
    - If a baseline value is missing, it will not be imputed and percent change from baseline will not be calculated.
- Weight (kg) = weight (lbs) * 0.454.
- Weight category (<10kg, ≥10 kg to <20 kg, ≥20 kg to <30 kg, ≥30 kg to <40 kg, ≥40 kg to <50 kg, ≥50 kg to <60 kg, ≥60 to <100 kg, ≥100 kg)
- Height (cm) = height (in) * 2.54.
- Prior TCNI use
    - Yes
    - No. (Reasons of not using the medication including: Physician decision, concern about side effects, unfavorable benefit risk, contraindication, insurance coverage/cost issues, patient/caregiver decision, not considered and unknown)
- TCNI inadequate response (yes, no)
    - Set **yes** if patient had prior TCNI use and the reason for discontinuation is inadequate response or to enter this trial.
- TCNI intolerance (yes, no)
    - Set **yes** if patient had prior TCNI use and the reasons for discontinuation are: intolerance to medication or contraindication (Physician indicated TCNI was used and a contraindication was noted).
- Prior TCNI use
    - Yes
    - No. (Reasons of not using the medication including: Physician decision, concern about side effects, unfavorable benefit risk, contraindication, insurance coverage/cost issues, patient/caregiver decision, not considered and unknown)
- Prior TCS use
    - Yes.

  - No. (Reasons of not using the medication including: Physician decision, concern about side effects, unfavorable benefit risk, contraindication, insurance coverage/cost issues, patient/caregiver decision, not considered and unknown)
- TCS inadequate response (yes, no)
  - Set **yes** if patient had prior TCS use and the reason for discontinuation is inadequate response or to enter this trial.
- TCS intolerance (yes, no)
  - Set **yes** if patient had prior TCS use and the reasons for discontinuation are: intolerance to medication or contraindication.

## 6.3. Handling of Dropouts or Missing Data

Intercurrent events (International Council on Harmonisation [ICH] E9 R1) are events which occur after randomization such that subsequent data (collected after the intercurrent event) are difficult to interpret.

Depending on the estimand being addressed, different methods will be used to handle missing data as a result of intercurrent events. Intercurrent events can occur through the following:

- application of one of the censoring rules (including after permanent study drug discontinuation or after rescue therapy)
- discontinuation
- missing an intermediate visit prior to discontinuation or rescue
- lost to follow-up.

Note that as efficacy data can accrue after a patient permanently discontinues study drug or begins rescue therapy, specific general censoring rules to the data will be applied to all efficacy observations subsequent to these events depending on the estimand being addressed. These specific censoring rules are described below.

The *primary censoring rule* in Study Period 3 will censor efficacy data after permanent study drug discontinuation or after rescue therapy. This censoring rule will be applied to all continuous and categorical efficacy endpoints. This censoring rule is equivalent to using all the data up to rescue.

A *secondary censoring rule* in Study Period 3 will only censor efficacy data after permanent study drug discontinuation. This sensitivity analysis will include all observed values up to study drug discontinuation. The secondary censoring rule will be applied to primary and key secondary efficacy endpoints as sensitivity analyses.

A *tertiary censoring rule* in Study Period 3 will censor efficacy data permanent study drug discontinuation or after starting concomitant medications that could be considered “rescue” but was used to treat other conditions. The tertiary censoring rule will be applied to selected primary and secondary categorial endpoints (IGA (0,1), EASI75, SCORAD75, Itch-NRS 4-point

improvement and PRISM 2-point improvement), as sensitivity analysis. The definition of the medications is provided in Appendix 2.

Table JAIP.6.4 describes the planned imputation methods for efficacy endpoints with associated censoring rules in Study Period 3. Sections 6.3.1 through 6.3.4 summarize the methodology of each imputation rule.

**Table JAIP.6.4. Imputation Techniques for Various Variables**

| Efficacy Endpoints | Imputation Method |
|---|---|
| IGA(0,1), EASI75, 4-point Itch NRS improvement, EASI90, SCORAD75 | NRI[abc], pMI[a] |
| EASI (percent) change from baseline | MMRM[ab], mLOCF[a], pMI[a] |
| All remaining categorical measures | NRI[a] |
| All remaining continuous efficacy measures | MMRM[a], mLOCF[a] |

Abbreviations: AD = atopic dermatitis; EASI = Eczema Area and Severity Index score EASI75/90 = 75%/90% response rate on EASI score; IGA = Investigator's Global Assessment for AD; mLOCF = modified last observation carried forward; MMRM = mixed model repeated measures; NRI = nonresponder imputation, NRS = Numeric Rating Scale; pMI = placebo multiple imputation; SCORAD75 = 75% response rate on SCORing Atopic Dermatitis.

[a] Analyses utilizing the primary censoring rule.

[b] Analyses utilizing the secondary censoring rule.

[c] Analyses utilizing the tertiary censoring rule.

### 6.3.1. Nonresponder Imputation

A nonresponder imputation (NRI) method imputes missing values as non-responses and can be justified based on the composite strategy for handling intercurrent events (ICH E9 R1). This imputation procedure assumes the effects of treatments disappear after the occurrence of an intercurrent event.

All categorical endpoints will utilize the NRI method after applying the primary censoring rule to patients who permanently discontinued study drug or were rescued (described in Section 6.3). Additionally, all primary and key secondary categorical endpoints will utilize NRI after applying the secondary censoring rule as sensitivity analyses. For analyses which utilize either of the censoring methods, randomized patients without at least 1 postbaseline observation will be defined as nonresponders for all visits. As well, patients who are missing a value prior to discontinuation or rescue (if censoring on rescue) (i.e., the patient is missing an intermediate visit) will be imputed as nonresponders, at that visit only.

### 6.3.2. Mixed Model for Repeated Measures

Mixed model for repeated measures analyses will be performed on continuous endpoints to mitigate the impact of missing data. This approach assumes missing observations are missing-at-random (missingness is related to observed data) and borrows information from patients in the same treatment arm taking into account both the missingness of data through the correlation of the repeated measurements.

Essentially MMRM estimates the treatment effects had all patients remained on their initial treatment throughout the study. For this reason, the MMRM imputation implies a different estimand (hypothetical strategy [ICH E9 R1]) than the one used for NRI on categorical outcomes.

All continuous endpoints will utilize MMRM after applying the primary censoring rule. As sensitivity analyses, all secondary continuous endpoints will also utilize MMRM after applying the secondary censoring rule (Table JAIP.6.4).

### 6.3.3. Modified Last Observation Carried Forward

A modified last observation carried forward (mLOCF) is performed by carrying forward the last postbaseline assessment for the continuous measures, assuming that effects of treatments remain the same after the occurrence of the intercurrent event (after application of the primary censoring rule). After mLOCF imputation, data from patients with nonmissing baseline and at least 1 postbaseline observation will be included in the analyses. These mLOCF analyses help ensure the maximum number of randomized patients who were assessed postbaseline will be included in the analyses.

For patients who experience any intercurrent event at any time, the last nonmissing postbaseline observation on or prior to this event will be carried forward to subsequent time points for evaluation. If a patient does not have a nonmissing observed record (or one imputed by other means) for a postbaseline visit prior to discontinuation or rescue, the last postbaseline record prior to the missed visit will be used for the visit.

All continuous efficacy endpoints will use mLOCF imputation methodology with an ANCOVA as sensitivity analyses to the MMRM analyses.

### 6.3.4. Placebo Multiple Imputation

The placebo multiple imputation (pMI) methodology will be used as a sensitivity analysis for the analysis of the primary efficacy endpoint (IGA 0 or 1 at Week 16) as well as the key secondary endpoints at Week 16. In these sensitivity analyses the primary censoring rule will be applied.

The pMI assumes that the statistical behavior of drug- and placebo- treated patients after the occurrence of intercurrent events will be the same as if patients were treated with placebo. Thus, in the effectiveness context, pMI assumes no pharmacological benefit of the drug after the occurrence of intercurrent events but is a more conservative approach than mLOCF because it accounts for uncertainty of imputation, and therefore does not underestimate standard errors, and it limits bias. In the efficacy context pMI is a specific form of a missing not at random analysis and expected to yield a conservative estimate of efficacy.

In the pMI analysis, multiple imputations are used to replace missing outcomes for drug- and placebo-treated patients who have an intercurrent event using multiple draws from the posterior predictive distribution estimated from the placebo arm. The binary outcomes will then be derived from the imputed data.

Data are processed sequentially by repeatedly calling SAS® PROC MI to impute missing outcomes at visits $t$=1,.., T.

1. *Initialization:* Set $t$=0 (baseline visit)
2. *Iteration:* Set $t$=$t$+1. Create a data set combining records from drug- and placebo-treated patients with columns for covariates **X** and outcomes at visits 1,..,$t$ with outcomes for all drug-treated patients set to missing at visit $t$ and set to observed or imputed values at visits 1,..,$t$-1.
3. *Imputation:* Run Bayesian regression in SAS® PROC MI on this data to impute missing values for visit $t$ using previous outcomes for visits 1 to $t$-1 and baseline covariates. Note that only placebo data will be used to estimate the imputation model since no outcome is available for drug-treated patients at visit $t$.
4. Replace imputed data for all drug-treated patients at visit $t$ with their observed values, whenever available up to permanent study drug discontinuation and/or rescue (if censoring on rescue). If $t$ < T then go to Step 2, otherwise proceed to Step 5.
5. Repeat steps 1-4, $m$ times with different seed values to create $m$ imputed complete data sets.

*Analysis:* For each completed data set, use the model that would have been applied had the data been complete for the continuous outcome. For the binary primary and secondary key efficacy endpoints (IGA [0,1], EASI75, EASI90, SCORAD75, and 4-point improvement from baseline in Itch NRS), the binary outcomes will be derived from the imputed underlying continuous outcome for each patient before fitting the logistic regression model.

The number of imputed data sets will be $m$=100 and a 6-digit seed value will be pre-specified for each analysis. Within the program, the seed will be used to generate the $m$ seeds needed for imputation. The initial seed values are given in Table JAIP.6.5.

**Table JAIP.6.5. Seed Values for Multiple Imputation**

| Analysis | Seed value |
|---|---|
| Proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16 using the primary censoring rule | 123450 |
| Change and percent change from baseline in EASI score at 16 weeks using the primary censoring rule. EASI75 and EASI90 will leverage imputation from EASI and therefore do not need a new seed number. | 123451 |
| Proportion of patients achieving SCORAD75 at 16 week using the primary censoring rule, with data up to rescue | 123452 |
| Proportions of patients achieving a 4-point improvement from baseline in Itch NRS at Week 16 using the primary censoring rule | 123453 |

Abbreviations: EASI = Eczema Area and Severity Index score; EASI75/90 = 75%/90% response rate on EASI score; IGA = Investigator's Global Assessment for AD; NRS = Numeric Rating Scale; SCORAD75 = 75% response rate on SCORing Atopic Dermatitis.

The final inference on treatment difference is conducted from the multiple datasets using Rubin's combining rules, as implemented in SAS® PROC MIANALYZE.

## 6.4. Multicenter Studies

This study will be conducted by multiple investigators at multiple sites internationally. The countries will be grouped into geographic regions, as described in Section 5.2.

For the analysis of the primary endpoint, treatment-by-region interaction will be added to the logistic regression model as a subgroup analysis and results from this model will be compared to the primary model (without the interaction effect). If the treatment-by-region interaction is significant at a 2-sided α level of 0.1, the nature of this interaction will be inspected as to whether it is quantitative (i.e., the treatment effect is consistent in direction across all regions but not in size of treatment effect) or qualitative (the treatment is beneficial in some but not all regions). If the treatment-by-region interaction effect is found to be quantitative, results from the primary model will be presented. If the treatment-by-region interaction effect is found to be qualitative, further inspection will be used to identify in which regions baricitinib is found to be more beneficial.

## 6.5. Multiple Comparisons/Multiplicity

The primary and key secondary endpoints will be adjusted for multiplicity in order to control the overall family-wise Type I error rate at a 2-sided alpha level of 0.05. The graphical multiple testing procedure described in Bretz et al. (2011) will be used. The graphical approach is a closed testing procedure; hence, it strongly controls the familywise error rate across all endpoints (Alosh et al. 2014).

The following is a list of primary and key secondary endpoints to be tested. The subscript for **H** denotes dose (**H**igh, **M**id, **L**ow), the numeric identifier of the endpoint within the dose, and the type of hypothesis (0 for null, 1 for alternative), respectively. The High, Mid, and Low baricitinib doses will be 4 mg, 2 mg, and 1 mg for patients ages 10 to <18 years old and 2 mg, 1 mg, and 0.5 mg for patients ages 2 to <10 years old.

Primary Null Hypotheses:

- $\mathbf{H_{H,1,0}}$: Proportion of baricitinib high dose patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16 is less than or equal to the proportion of placebo patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16 (IGA0-1)
- $\mathbf{H_{M,1,0}}$: Proportion of baricitinib mid dose patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16 is less than or equal to the proportion of placebo patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16 (IGA0-1)
- $\mathbf{H_{L,1,0}}$: Proportion of baricitinib low dose patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16 is less than or equal to the proportion of placebo patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16 (IGA0-1)

Key Secondary Null Hypotheses

- $\mathbf{H_{H,2,0}}$: Proportion of baricitinib high dose patients achieving EASI75 is less than or equal to the proportion of placebo patients achieving EASI75 at Week 16 (EASI75)
- $\mathbf{H_{H,3,0}}$: Proportion of baricitinib high dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 16 (ITCH W16)
- $\mathbf{H_{H.4,0}}$: Mean change from baseline in EASI score for baricitinib high dose patients is greater than or equal to the mean change from baseline in EASI score for placebo patients at Week 16 (EASI PCFB)
- $\mathbf{H_{H,5,0}}$: Proportion of baricitinib high dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 4 (ITCH W4)
- $\mathbf{H_{H,6,0}}$: Proportion of baricitinib high dose patients achieving SCORAD75 is less than or equal to the proportion of placebo patients achieving SCORAD75 at Week 16 (SCORAD75)
- $\mathbf{H_{H,7,0}}$: Proportion of baricitinib high dose patients achieving EASI90 is less than or equal to the proportion of placebo patients achieving EASI90 at Week 16 (EASI 90)
- $\mathbf{H_{H,8,0}}$: Proportion of baricitinib high dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 2 (ITCH W2)
- $\mathbf{H_{H,9,0}}$: Proportion of baricitinib high dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 1(ITCH W1)
- $\mathbf{H_{M,2,0}}$: Proportion of baricitinib mid dose patients achieving EASI75 is less than or equal to the proportion of placebo patients achieving EASI75 at Week 16 (EASI75)
- $\mathbf{H_{M,3,0}}$: Proportion of baricitinib mid dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 16 (ITCH W16)
- $\mathbf{H_{M.4,0}}$: Mean change from baseline in EASI score for baricitinib mid dose patients is greater than or equal to the mean change from baseline in EASI score for placebo patients at Week 16 (EASI PCFB)
- $\mathbf{H_{M,5,0}}$: Proportion of baricitinib mid dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 4 (ITCH W4)
- $\mathbf{H_{M,6,0}}$: Proportion of baricitinib mid dose patients achieving SCORAD75 is less than or equal to the proportion of placebo patients achieving SCORAD75 at Week 16 (SCORAD75)
- $\mathbf{H_{M,7,0}}$: Proportion of baricitinib mid dose patients achieving EASI90 is less than or equal to the proportion of placebo patients achieving EASI90 at Week 16 (EASI 90)
- $\mathbf{H_{M,8,0}}$: Proportion of baricitinib mid dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 2 (ITCH W2)
- $\mathbf{H_{M,9,0}}$: Proportion of baricitinib mid dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 1(ITCH W1)

- $H_{L,2,0}$: Proportion of baricitinib low dose patients achieving EASI75 is less than or equal to the proportion of placebo patients achieving EASI75 at Week 16 (EASI75)
- $H_{L,3,0}$: Proportion of baricitinib low dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 16 (ITCH W16)
- $H_{L.4,0}$: Mean change from baseline in EASI score for baricitinib low dose patients is greater than or equal to the mean change from baseline in EASI score for placebo patients at Week 16 (EASI PCFB)
- $H_{L,5,0}$: Proportion of baricitinib low dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 4 (ITCH W4)
- $H_{L,6,0}$: Proportion of baricitinib low dose patients achieving SCORAD75 is less than or equal to the proportion of placebo patients achieving SCORAD75 at Week 16 (SCORAD75)
- $H_{L,7,0}$: Proportion of baricitinib low dose patients achieving EASI90 is less than or equal to the proportion of placebo patients achieving EASI90 at Week 16 (EASI 90)
- $H_{L,8,0}$: Proportion of baricitinib low dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 2 (ITCH W2)
- $H_{L,9,0}$: Proportion of baricitinib low dose patients achieving a 4-point improvement in Itch NRS is less than or equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 1 (ITCH W1)

The multiple testing strategy for the primary and the major secondary endpoints will be implemented through the graphical testing procedure depicted by Figure JAIP.6.1.

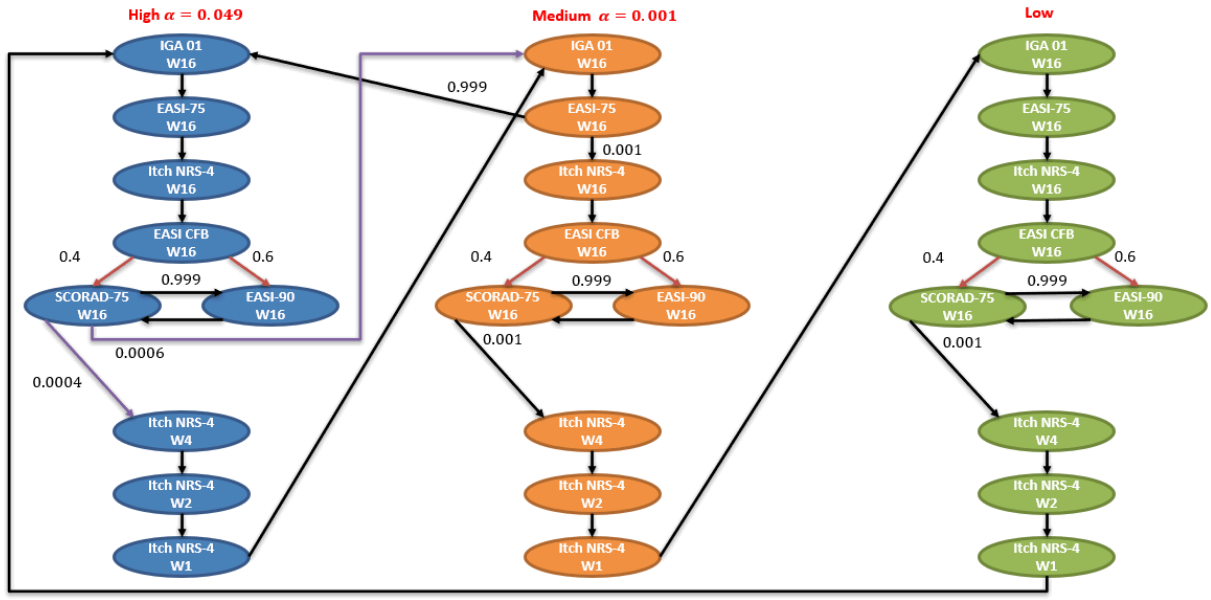



**Figure JAIP.6.1. Illustration of graphical multiple testing procedure with initial α allocation and weights.**

## 6.6. Patient Disposition

An overview of patient populations will be summarized by treatment group. Frequency counts and percentages of patients excluded prior to randomization by primary reason for exclusion will be provided for patients who failed to meet study entry requirements during screening. A listing of patient disposition will be provided for all randomized patients, with treatment assignment, the extent of their participation in the study, and the reason for discontinuation.

Patient disposition through Week 16 will be summarized using the ITT population. Frequency counts and percentages of patients who complete the study treatment visits or discontinue early from the study along with whether they completed follow-up, did not complete follow-up or enrolled into the extension will be summarized separately by treatment group, along with their reason for study discontinuation. Frequency counts and percentages of patients who complete the treatment or discontinue treatment early will also be summarized separately by treatment group, along with their reason for treatment discontinuation.

## 6.7. Patient Characteristics

Patient characteristics including demographics and baseline characteristics will be summarized descriptively by treatment group and by age group for the ITT population as well as for PK lead-in population. Historical illnesses and pre-existing conditions will be summarized descriptively by treatment group for the ITT population. No formal statistical comparisons will be made among treatment groups unless otherwise stated.

### *6.7.1. Demographics*

Patient demographics will be summarized as described above. The following demographic information will be included:

- Age
- Age group (2 to <10, 10 to <18)
- Age group (2 to <6, 6 to <10, 10 to <18)
- Gender (male, female)
- Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Multiple)
- Region (as defined in Table JAIP.5.1)
- Country
- Weight (kg)
- Weight category <20 kg, ≥20 kg to <60 kg, ≥60 kg)
- Height (cm)
- BMI ($kg/m^2$)
- BMI category (<20 $kg/m^2$, ≥20 to <30 $kg/m^2$, ≥30 $kg/m^2$)

A listing of patient demographics will also be provided for the ITT population.

### *6.7.2. Baseline Disease Characteristics*

The following baseline disease information will be categorized and presented for baseline AD clinical characteristics, baseline health outcome measures, and other baseline demographic and disease characteristics as described above:

- Duration since AD diagnosis (years) = ([Date of informed consent – Date of AD diagnosis] +1)/365.25.
- Duration since AD diagnosis category (<6 months, 6 months to <1 year, 1 year to <2 years, 2 to <5 years, 5 to <10 years, 10 to <15 years, ≥15 years)
- Age at Diagnosis (years)
- Age Group at Diagnosis (<10 years, ≥10 years)
- Substance Use for patients ≥10 years (Alcohol: Never, Current, Former; Tobacco: Never, Current, Former)
- Skin Infections treated with a pharmacological agent within past year (yes, no, unknown; number if yes)
- Atopic Dermatitis Flares within past year (yes, no, unknown; number if yes)
- vIGA-AD
- EASI score
- SCORAD
- BSA affected by AD
- POEM
- PROMIS – Depression
- PROMIS – Anxiety
- PRISM
- Itch NRS
- Skin Pain NRS
- ADSS Item 2
- Children's Dermatology Life Quality Index (CDLQI)
- Infant's Dermatitis Quality of Life Index (IDQOL)
- PGI-S-AD
- Prior therapy (topical therapy only; systemic therapy)
- Prior use of TCNI
    - Yes
        - TCNI inadequate response (yes, no)
        - TCNI intolerance (yes, no)
    - No
- Prior use of TCS
    - Yes
        - TCS inadequate response (yes, no)
        - TCS intolerance (yes, no)
    - No

- Vaccine (yes, no)
- Baseline renal function status: estimated glomerular filtration rate (eGFR) <100 mL/min/1.73 $m^2$ or eGFR ≥100 mL/min/1.73 $m^2$.
- Immunoglobulin E (IgE)

### *6.7.3. Historical Illness and Pre-existing Conditions*

Historical illnesses are defined as those conditions recorded in the Pre-existing Conditions and Medical History electronic case report form (eCRF) or from the Prespecified Medical History: Comorbidities eCRF with an end date prior to the informed consent date. The number and percentage of patients with historical diagnoses will be summarized using MedDRA Preferred Term nested within System Organ Class, for the ITT population.

Pre-existing conditions are defined as those conditions recorded in the Pre-existing Conditions and Medical History eCRF, the Prespecified Medical History: Comorbidities eCRF, or the Adverse Events eCRF with a start date prior to the first dose of study treatment (or randomization date for those who don't ever receive a dose of study drug) and an end date at or after informed consent or ongoing. For events occurring on the day of the first dose of study treatment, the date and time of the onset of the event will both be used to determine if the event was pre-existing. Conditions with a partial or missing start date (or time if needed) will be assumed to be 'not pre-existing' unless there is evidence, through comparison of partial dates, to suggest otherwise. The number and percentage of patients with pre-existing conditions will be summarized using MedDRA Preferred Term nested within System Organ Class, for the ITT population.

## 6.8. Treatment Compliance

Patient compliance with study medication will be assessed from Week 0 (Visit 2) to Week 16 (Visit 8) or Early Termination using the ITT population.

Double-blinded older patient should take 3 tablets per day whatever the dose is. One blister has 27 tablets. (one blister card is dispensed during 1-week interval; 4 week interval has 4 blister card; the same number of blister cards are delivered across placebo and BARI low/med/high treatment groups.)

Patient compliance may be assessed for PK population for combined period 2 and 4. Each open-label bottle contains 36 tablets and PK patient takes 1 tablet per day.

A patient is considered noncompliant if he or she misses >20% of the prescribed doses during the study, unless the patient's study drug is withheld by the investigator (refer to clinical protocol for daily doses). Refer to the Treatments section of the clinical protocol for daily treatment regimen. For patients who had their treatment temporarily interrupted by the investigator, the period of time that dose was withheld will be taken into account in the compliance calculation.

Compliance for tablets in the period of interest up to Visit $x$ will be calculated as follows:

$$\text{Compliance} = \frac{\text{total number of tablets dispensed} - \text{total number of tablets returned}}{\text{expected number of total tablets}}$$

where

- Total number of tablets dispensed: sum of tablets dispensed in the period of interest prior to Visit $x$;
- Total number of tablets returned: sum of the tablets returned in the period of interest prior to and including Visit $x$;
- Expected number of tablets: number of days in the period of interest*number of tablets taken per day = ([date of visit $x$ – date of first dose + 1] – number of days of temporary drug interruption)*number of tablets taken per day.

Compliance for suspension in the period of interest up to Visit $x$ will be calculated as follows:

$$\text{Compliance} = \frac{\text{weight of suspension dispensed (g)} - \text{weight of suspension returned (g)}}{1.05 * \text{Days of exposure} * \text{Volume of daily dose (mL)}} * CF$$

where

- Weight of suspension dispensed: weight of suspension dispensed at the visit prior to Visit $x$;
- Weight of suspension returned: weight of suspension returned at Visit $x$. Note that 1.05 in the denominator is a conversion factor included in the calculation to convert weight measurements to an equivalent volume;
- Days of exposure: number of days in the period of interest = (date of visit $x$ – date of first dose + 1) – number of days of temporary drug interruption;
- CF is a correction factor that accounts for the 0.05 mL of suspension that is left over in the tip of the syringe after each dose. The CF for each dose is as follows:
  - For the 2-mg dose: CF = 100/120
  - For the 1-mg dose: CF = 100/110
  - For the 0.5-mg dose: CF = 100/105

If patients have missing initial unused tube weight, it will be imputed using 144 g.

Patients who are significantly noncompliant (compliance <80%) through Week 16 will be excluded from the PPS population.

For <u>any</u> patient taking placebo suspension liquid during Week 0-16, the compliance for Week 0 through Week 16 will not be calculated because the volume dispensed is not recorded in IWRS data or INFORM data. And such patients will not be excluded from the PPS population due to the reason that compliance is not calculated.

Descriptive statistics for percent compliance and non-compliance rate will be summarized for the ITT population by treatment group for Week 0 through Week 16. The number of expected doses, doses dispensed, doses returned, and percent compliance will be listed by patient for Week 0 through Week 16.

### *6.8.1. Rescue Treatment*

Refer to the clinical protocol for detailed description of rescue therapies by study period. Rescue treatments are broadly summarized by study period as follows:

Study Period 3: Topical treatment with high or ultra-high potency TCS will be considered rescue therapy. Any systemic treatment will be considered rescue therapy. The initial rescue therapy is defined as the first non-missing record with medication start date before the last dose date in study period 3 in the following categories collected from the CRF page *Concomitant Therapy: Atopic Dermatitis Therapy (CM_SI*): HIGH OR ULTRA POTENCY TOPICAL CORTICOSTEROID, SYSTETEMIC CONVENTIAL THERAPY, SYSTEMIC BIOLOGIC THERAPY, any therapy in OTHER category including the text of "phototherapy" or "UV".

Study Period 4: Systemic conventional therapy, systemic biologic therapy and phototherapy will be considered as rescue therapy in Period 4.

A summary of the initial rescue therapy and the reason for requiring initial rescue will be produced, as well as a summary of the proportion of patients initially rescued at each study visit. A summary of all rescue medications will be provided.

### *6.8.2. Background Therapy*

The use of low- and medium-potency TCS is permitted as background therapy throughout Study Period 3. The use of TCS (all potencies) is permitted as background therapy throughout Study Period 4.

**Secondary endpoints for background therapy:**

- Mean number of days <u>without</u> use of background TCS over 16 weeks

The following analyses will be performed: The total number of days that patients did not use background TCS will be summarized throughout study period 3.

The main analysis applies the secondary censoring rule, assuming that background TCS was applied each day afterwards for the remainder of the study time. Before censoring, daily diary entries will be used to calculate days on background TCS (any, i.e., mild to medium potency). In case of missing values in the daily diary, it will be assumed that background TCS has been used.

Data will be summarized descriptively and presented overall and in monthly (28-day) intervals. Analysis will be done via analysis of variance (ANOVA), with geographic region, baseline disease severity, age cohort and treatment as factors in the model. These analyses will be done for 16.

A secondary analysis will apply the primary censoring rule, with the same assumptions for missing values as described above. This analysis will only be performed for 16 weeks.

Descriptive statistics of proportion of the time that background TCS is not used will also be presented for 16 weeks.

- Mean gram quantity of background TCS used over 16 weeks (tube weights)

Descriptive statistics for drug accountability of topical low and moderate potency background medication provided by the sponsor will be presented, including the amount utilized throughout the treatment period from Week 0 through Week 16. The dispensed sponsor-provided TCS tubes were weighed with cap (without the carton) to determine the dispensed amount of TCS in grams. Returned tubes were weighed with cap (without the carton) to determine the amount of TCS in grams (g) used at each visit.

The total amount in grams for low and moderate potency, as well as the sum of both potencies will be summarized between visits (Week 0 through Week 1, Week 1 through Week 2, Week 2 through Week 4, Week 4 through Week 8, Week 8 through 12, Week 12 through Week 16), as well as throughout the treatment period from Week 0 through Week 16. If a returned tube is not weighed in grams then the tube can be classified as partially used, fully used, unused, or unknown. Partially used background medication tubes will be considered to be 50% used whereas Fully used and Unused will be considered as 100% and 0% used, respectively. When drug accountability is not performed for a particular tube of background medication or an answer of Unknown is given for a tube which is not returned, that particular tube will not be included in the analysis.

The analysis approach will be similar to the analyses described above for days without background TCS use, and will be performed using the sum of both low and mid potency background TCS. The analyses will apply secondary censoring rule as well as primary censoring rule with the same assumption as described above.

Note: As rescue TCS (high/ultra-high potency) is not weighed, an analysis similar to the main analysis described above for the mean gram quantity without background TCS cannot be performed.

## 6.9. Previous and Concomitant Therapy

Summaries of previous and concomitant medications will be based on the ITT population.

At screening, previous and current AD treatments are recorded for each patient. Concomitant therapy for the treatment period is defined as therapy that starts before or during the treatment period and ends during the treatment period or is ongoing (has no end date or ends after the treatment period). Should there be insufficient data to make this comparison (e.g., the concomitant therapy stop year is the same as the treatment start year, but the concomitant therapy stop month and day are missing), the medication will be considered as concomitant for the treatment period.

Summaries of previous medications will be as follows:

- Previous AD therapies

Summaries of concomitant medications will be as follows:

- Concomitant medications excluding rescue medicine

## 6.10. Efficacy Analyses

The general methods used to summarize efficacy data, including the definition of baseline value for assessments are described in Section 6.2. The censoring rules applied to data as well as imputation methods are described in Section 6.3.

Table JAIP.6.6 provides the descriptions and derivations of the primary, secondary, and exploratory efficacy outcomes.

Table JAIP.6.7 provides the detailed analyses including analysis type, method and imputation, population, time point, and comparisons for efficacy analyses.

**Table JAIP.6.6. Description and Derivation of Primary, Secondary and Exploratory Efficacy Outcomes**

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| Validated Investigator's Global Assessment for AD (IGA) | The validated Investigator's global assessment of the patient's overall severity of their AD, based on a static, numeric 5-point scale from 0 (clear) to 4 (severe). The score is based on an overall assessment of the degree of erythema, papulation/induration, oozing/crusting, and lichenification. | IGA score | Single item. Range: 0 to 4<br>0 represents "clear"<br>4 represents "severe" | Single item, missing if missing. |
| | | Change from baseline in IGA score | Change from baseline: observed IGA score – baseline IGA score | Missing if baseline or observed value is missing. |
| | | ▪ IGA [0,1] with ≥2-point improvement<br>▪ IGA [0] | ▪ Observed score of 0 or 1 and change from baseline ≤-2<br>▪ Observed score of 0 | ▪ Missing if baseline or observed value is missing.<br>▪ Single item, missing if missing. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| Eczema Area and Severity Index (EASI) | The EASI assesses objective physician estimates of 2 dimensions of atopic dermatitis – disease extent and clinical signs (Hanifin et al. 2001) – by scoring the extent of disease (percentage of skin affected: 0 = 0%; 1 = 1-9%; 2 = 10-29%; 3 = 30-49%; 4 = 50-69%; 5 = 70-89%; 6 = 90-100%) and the severity of 4 clinical signs (erythema, edema/papulation, excoriation, and lichenification) each on a scale of 0 to 3 (0 = none, absent; 1 = mild; 2 = moderate; 3 = severe) at 4 body sites (head and neck, trunk, upper limbs, and lower limbs). Half scores are allowed between severities 1, 2 and 3. Each body site will have a score that ranges from 0 to 72, and the final EASI score will be obtained by weight-averaging these 4 scores. Hence, the final EASI score will range from 0 to 72 for each time point. | EASI score | Derive EASI region score for each of head and neck, trunk, upper limbs, and lower limbs as follows:<br>$EASI_{region}$ = (Erythema + edema/papulation + Excoriation + Lichenification) *(value from percentage involvement), where erythema, edema/papulation, excoriation, and lichenification are evaluated on a scale of 0 to 3 and value from percentage involvement is on a scale of 0 to 6.<br><br>Then total EASI score is as follows:<br>$EASI = 0.1*EASI_{head\ and\ neck} + 0.3*EASI_{trunk} + 0.2*EASI_{upper\ limbs} + 0.4*EASI_{lower\ limbs}$<br><br>For patients 0 to <8 years old, total EASI score is as follows:<br>$EASI = 0.2*EASI_{head\ and\ neck} + 0.3*EASI_{trunk} + 0.2*EASI_{upper\ limbs} + 0.3*EASI_{lower\ limbs}$<br><br>For total EASI score on or before Visit 2, baseline age will be used in the derivation. For all postbaseline visits, to ensure the scientific merit of the measure, age at the assessment will be used in calculation.<br><br>Note that sites will also collect EASI score under some circumstances, for which they will continue to use baseline age for postbaseline visits. In those circumstances, discrepancy between statistical calculation and site report will be expected. | N/A – partial assessments cannot be saved. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| | | ▪ Change from baseline in EASI score<br>▪ Percent change from baseline EASI score | Change from baseline: observed EASI score – baseline EASI score<br>% change from baseline: $100 \times \frac{Observed\ score - Baseline}{Baseline}$ | Missing if baseline or observed value is missing. |
| | | EASI50 | % Improvement in EASI score from baseline ≥50%:<br>% change from baseline ≤-50 | Missing if baseline or observed value is missing. |
| | | EASI75 | % Improvement in EASI score from baseline ≥75%:<br>% change from baseline ≤-75 | Missing if baseline or observed value is missing. |
| | | EASI90 | % Improvement in EASI score from baseline ≥90%:<br>% change from baseline ≤-90 | Missing if baseline or observed value is missing. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| Body Surface Area (BSA) Affected by AD | Body surface area affected by AD will be assessed for 4 separate body regions and is collected as part of the EASI assessment: head and neck, trunk (including genital region), upper extremities, and lower extremities (including the buttocks). Each body region will be assessed for disease extent ranging from 0% to 100% involvement. The overall total percentage will be reported based off of all 4 body regions combined, after applying specific multipliers to the different body regions to account for the percent of the total BSA represented by each of the 4 regions. | BSA score | Use the percentage of skin affected for each region (0 to 100%) in EASI as follows:<br>BSA Total = $0.1*BSA_{\text{head and neck}} + 0.3*BSA_{\text{trunk}} + 0.2*BSA_{\text{upper limbs}} + 0.4*BSA_{\text{lower limbs}}$ | N/A – partial assessments cannot be saved. |
| | | Change from baseline in BSA score | Change from baseline: observed BSA score – baseline BSA score | Missing if baseline or observed value is missing. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| SCORing Atopic Dermatitis (SCORAD) | | SCORAD score | SCORAD = A/5 + 7B/2 + C, where<br>A is extent of disease, range 0-100<br>B is disease severity, range 0-18<br>C is subjective symptoms, range 0-20 | Missing if components A and B are missing or if component C is missing. Partial assessments performed by physician cannot be saved and partial assessments performed by subject cannot be saved. |
| | | ▪ Change from baseline in SCORAD score<br>▪ Percent change from baseline in SCORAD score | Change from baseline: observed SCORAD score – baseline SCORAD score<br><br>% change from baseline:<br>$100 \times \frac{Observed\ score - Baseline}{Baseline}$ | Missing if baseline or observed value is missing. |
| | | SCORAD75 | % Improvement in SCORAD from baseline ≥75%:<br>% change from baseline ≤-75 | Missing if baseline or observed value is missing. |

| | | | | |
|---|---|---|---|---|
| | The SCORing Atopic Dermatitis (SCORAD) index uses the rule of nines to assess disease extent (head and neck 9%; upper limbs 9% each; lower limbs 18% each; anterior trunk 18%; back 18%; and genitals 1%). It evaluates 6 clinical characteristics to determine disease severity: (1) erythema, (2) edema/papulation, (3) oozing/crusts, (4) excoriation, (5) lichenification, and (6) dryness on a scale of 0 to 3 (0=absence, 1=mild, 2=moderate, 3=severe). The SCORAD index also assesses subjective symptoms of pruritus and sleep loss in the last 72 hours on visual analogue scales (VAS) of 0 to 10 where 0 is no itch or sleep loss and 10 is worst imaginable itch or sleep loss. These 3 aspects: extent of disease, disease severity, and subjective symptoms combine to give a maximum possible score of 103 (Stalder et al. 1993; Kunz et al. 1997; Schram et al. 2012). | SCORAD90 | % Improvement in SCORAD from baseline ≥90%:<br>% change from baseline ≤-90 | Missing if baseline or observed value is missing. |

**Table JAIP.6.7. Description of Primary, Secondary, and Exploratory Efficacy Analyses***

| Measure | Variable | Analysis Method (Section 6.2.3) | Population (Section 6.2.1) | Comparison/Time Point | Analysis Type |
|---|---|---|---|---|---|
| Validated Investigator's Global Assessment for AD (IGA) | Proportion of patients achieving IGA [0,1] with a ≥2-point improvement | Logistic regression using NRI | ITT | Bari high, mid, or low dose vs. PBO; Week 16 | Primary analysis |
| | | | | Combined Bari high and mid dose vs. PBO; Week 16 | Exploratory analysis |
| | | | PPS | Bari high, mid, or low dose vs. PBO; Week 16 | Sensitivity analysis |
| | | Logistic regression using pMI | ITT | Bari high, mid, or low dose vs. PBO; Week 16 | Sensitivity analysis |
| | | Descriptive statistics using NRI and observed | ITT- Week 16 responder and partial responder | No statistical comparison; Week 52 | Secondary analysis |
| | | | | No statistical comparison; 2,3,4 year | Exploratory analysis |
| | | | ITT- Week 16 Non- responder | No statistical comparison; Week 52 | Secondary analysis |
| | | | | No statistical comparison; 2,3,4 year during long term extension | Exploratory analysis |
| | Proportion of patients achieving IGA [0] | Logistic regression using NRI | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Secondary analysis |
| Eczema Area and Severity Index (EASI) | • EASI score<br>• Change from baseline in EASI score<br>• Percent change from baseline in EASI score | MMRM | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Key secondary analysis |
| | | | PPS | Bari high or mid or low dose vs. PBO; Week 16 | Sensitivity analysis |
| | | ANCOVA using mLOCF | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Sensitivity analysis |
| | | pMI | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Sensitivity analysis |
| | • Proportion of patients achieving EASI50 | Logistic regression using NRI | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Secondary analysis |

| Measure | Variable | Analysis Method (Section 6.2.3) | Population (Section 6.2.1) | Comparison/Time Point | Analysis Type |
|---|---|---|---|---|---|
| | • Proportion of patients achieving EASI75<br>• Proportion of patients achieving EASI90 | Logistic regression using NRI | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Key secondary analysis |
| | | | | Combined Bari high and mid dose vs. PBO; Week 16 | Exploratory analysis |
| | | | PPS | Bari high or mid or low dose vs. PBO; Week 16 | Sensitivity analysis |
| | | | ITT- Week 16 responder and partial responder | No statistical comparison; Week 52 | Secondary analysis |
| | | | | No statistical comparison; 2,3,4 year during long term extension | Exploratory analysis |
| | | Descriptive statistics using NRI and observed | ITT- Week 16 Non-responder | No statistical comparison; Week 52 | Secondary analysis |
| | | | | No statistical comparison; 2,3,4 year during long term extension | Exploratory analysis |
| | | pMI | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Sensitivity analysis |
| Body Surface Area (BSA) Affected by AD | • BSA score<br>• Mean change from baseline in BSA score | MMRM | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Secondary analysis |
| | | ANCOVA using mLOCF | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Sensitivity analysis |
| SCORing Atopic Dermatitis (SCORAD) | • SCORAD score<br>• Change from baseline in SCORAD score<br>• Percent change from baseline in SCORAD score | MMRM | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Secondary analysis |
| | | ANCOVA using mLOCF | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Sensitivity analysis |
| | | Logistic regression using NRI | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Key secondary analysis |

| Measure | Variable | Analysis Method (Section 6.2.3) | Population (Section 6.2.1) | Comparison/Time Point | Analysis Type |
|---|---|---|---|---|---|
| | Proportion of patients achieving SCORAD75 | | | Combined Bari high and mid dose vs. PBO; Week 16 | Exploratory analysis |
| | | | PPS | Bari high or mid or low dose vs. PBO; Week 16 | Sensitivity analysis |
| | | Logistic regression using pMI | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Sensitivity analysis |
| | | Descriptive statistics using NRI and observed | ITT- Week 16 responder and partial responder | No statistical comparison; Week 52 | Secondary analysis |
| | | | | No statistical comparison; 2,3,4 year during long term extension | Exploratory analysis |
| | | | ITT- Week 16 Non-responder | No statistical comparison; Week 52 | Secondary analysis |
| | | | | No statistical comparison; 2,3,4 year during long term extension | Exploratory analysis |
| | Proportion of patients achieving SCORAD90 | Logistic regression using NRI | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Secondary analysis |
| Skin Infections | Proportion of patients developing skin infections requiring antibiotic treatment | Fisher's exact | ITT | Bari high or mid or low dose vs. PBO; Week 16 | Secondary analysis |

Abbreviations: AD = atopic dermatitis; ANCOVA = analysis of covariance; Bari = baricitinib; EASI50/75/90 = 50%/75%/90% response rate on the Eczema Area and Severity Index; ITT = intent-to-treat; mLOCF = modified last observation carried forward; MMRM = mixed model repeated measures; NRI = nonresponder imputation; PBO = placebo; pMI=placebo multiple imputation; PPS = per protocol set; SCORAD75/90 = 75%/90% response rate on SCORing Atopic Dermatitis; vs. = versus.

* Written in general dose terminology. Note that Bari high, mid, and low doses depend on the patients age cohort (i.e., doses for patients in the 10- to <18-year-old cohort are 4 mg, 2 mg, 1 mg, while doses for patients in the 2- to <10-year-old cohort are 2 mg, 1 mg, and 0.5 mg).

### 6.10.1. Primary Outcome and Methodology

The validated IGA for AD uses the clinical characteristics of erythema, papulation/induration, oozing/crusting and lichenification to produce a single-item score ranging from 0 to 4.

The primary objective of this study is to test the hypotheses that each dose of baricitinib (high, mid, and low dose) is superior to placebo in the treatment of patients with moderate-to-severe AD, as measured by the proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16 using the ITT population, and assuming the treatment response disappears after patients are rescued or permanently discontinue from treatment. This will serve as the primary estimand. In this estimand, missing data due to the application of the primary censoring rule and the occurrence of other non-censor intercurrent events will be imputed using the NRI method described in Section 6.3.1.

A supplemental estimand is to test the hypotheses that each dose of baricitinib (high, mid, and low dose) is superior to placebo when evaluating the proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16 using the ITT population, assuming the treatment response disappears after patients permanently discontinue from treatment. In this supplemental estimand, missing data due to the application of the secondary censoring rule and the occurrence of other non-censor intercurrent events will be imputed using the NRI method described in Section 6.3.1.

A logistic regression analysis as described in Section 6.2.3 will be used for the comparisons. The odds ratio, the corresponding 95% CIs and p-value, as well as the treatment differences and the corresponding 95% CIs, will be reported.

Multiplicity controlled analyses will be performed on the primary and key secondary (see Section 4) objectives in order to control the overall Type I error rate at a 2-sided alpha level of 0.05. A graphical approach will be used to perform the multiplicity controlled analyses as described in Section 6.5. There will be no adjustment for multiple comparisons for any other analyses.

### 6.10.2. Secondary and Exploratory Efficacy Analyses

For secondary analysis, the hypothesis is that each dose of baricitinib (high, mid, and low dose) is superior to placebo in the ITT population. These analyses assume treatment response disappears after patients are rescued or permanently discontinued from treatment and will serve as the primary estimand. In this estimand, missing data due to the application of the primary censoring rule and the occurrence of other non-censor intercurrent events will be imputed using the method described in Table JAIP.6.4.

A supplemental estimand for secondary endpoints is to test hypothesis that each dose of baricitinib (high, mid, and low dose) is superior to placebo in the ITT population. These analyses assumes the treatment response disappears after patients permanently discontinue from treatment. In this supplemental estimand, missing data due to the application of the secondary censoring rule and the occurrence of other non-censor intercurrent events will be imputed using the method described in Table JAIP.6.4.

A list of exploratory endpoints are provided in Section 4. There will be no adjustment for multiple comparisons for exploratory endpoints. The secondary and exploratory efficacy endpoints are detailed in Table JAIP.6.6 and Table JAIP.6.8 and analyses are provided in Table JAIP.6.7 and Table JAIP.6.9. In addition to the exploratory analysis included in Table JAIP.6.7 efficacy outcomes (IGA and EASI total scores) in patients who choose to take a voluntary drug interruption (drug holiday) during Study Period 4 will be summarized by dose at the following time points: visit prior to voluntary drug interruption, visit at which IP is resumed, and the last visit in Study Period 4. The number of patients who are able to maintain control of AD signs and symptoms (where control is defined as and IGA ≤2) without the use of TCS will be summarized in two groups: those patients who are able to stop TCS and never use TCS again in the study will be summarized by treatment group, and those who were able to stop TCS, but needed to resume use during the study will be summarized by treatment group.

### *6.10.3. Sensitivity Analyses*

Sensitivity analyses are included to demonstrate robustness of analyses methods using different missing data imputations, censoring rules, populations and analyses assumptions. Sensitivity analyses for select outcomes have been previously described and include the following:

- Analyses of key endpoints using the per-protocol analysis set (Section 6.2.1)
- Analyses of key endpoints using the secondary censoring rule (Section 6.3)
- Analyses of key endpoints using the tertiary censoring rule (Section 6.3)
- Placebo multiple imputation (Section 6.3.4)
- Analysis of continuous outcomes with ANCOVA (Section 6.2.3), with missing data imputed using mLOCF (Section 6.3.3).

## 6.11. Exploratory Analyses

Table JAIP.4.1 includes exploratory analysis specified in protocol. Additional exploratory efficacy analysis are presented in Table JAIP.6.7 and Table JAIP.6.9. Additional exploratory safety analysis for Study Period 3 will be using safety population with pooled Bari high and medium dose.

## 6.12. Acceptability and Palatability Analysis

Acceptability and palatability data will be collected and analyzed to address secondary objectives of this study. Baricitinib tablet or oral suspension product acceptability and palatability during the PK lead-in period will be summarized categorically (frequency and percentage) by age group, for each visit separately and in aggregate.

## 6.13. Health Outcomes/Quality-of-Life Analyses

The general methods used to summarize health outcomes and quality-of-life measures, including the definition of baseline value for assessments are described in Section 6.2.

Health outcomes and quality-of-life measures will generally be analyzed according to the formats discussed in Section 6.2.

Table JAIP.6.7 includes the descriptions and derivations of the health outcomes and quality-of-life measures.

Table JAIP.6.9 provides the detailed analyses including analysis type, method and imputation, population, time point, and comparisons for health outcomes and quality-of-life measures.

The number of missed school days for school age children will be reported via daily diary by the parent/caregiver or patient, and will be summarized by treatment group during Study Period 3.

Additional psychometric analyses will be performed by Global Patient Outcomes Real World Evidence at Lilly and documented in a separate analysis plan.

**Table JAIP.6.8. Description and Derivation of Health Outcomes and Quality-of-Life Measures**

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| Itch Numeric Rating Scale (NRS) | The Itch NRS is a patient-administered, 11-point horizontal scale anchored at 0 and 10, with 0 representing "no itch" and 10 representing "worst itch imaginable." Overall severity of a patient's itching is indicated by selecting the number that best describes the worst level of itching in the past 24 hours (Naegeli et al. 2015; Kimball et al. 2016). Refer to Section 6.2.2 for details on how to calculate the weekly score which will be used in the continuous analysis. | Itch NRS score | Single item; range 0-10. Refer to Section 6.2.2 on how to derive the visit score. | Refer to Section 6.2.2 on how to derive the weekly visit score. |
| | | ▪ Change from baseline in Itch NRS<br>▪ Percent change from baseline in Itch NRS | Change from baseline: observed Itch score – baseline Itch score<br>% change from baseline:<br>$100 \times \frac{Observed\ score - Baseline}{Baseline}$ | Missing if baseline or observed value is missing. |
| | | 4-point Itch improvement in subgroup of patients with baseline Itch NRS ≥4 | Change from baseline ≤-4 and baseline ≥4 | Missing if baseline is missing or <4 or observed value is missing. |
| | | Itch-free days (Itch NRS = 0) | Count of observed value = 0 for 28-day (4-week) intervals starting on the day of the first study drug administration. This will be calculated for the following visit intervals: baseline to Week 4, Week 4 to Week 8, Week 8 to Week 12 and Week 12 to Week 16. Day 1 is defined as the day of first study drug administration, therefore, the baseline to Week 4 assessment is based on Day 1 to Day 28, Week 4 to Week 8 is based on Day 29 to Day 56, etc. | If patients do not have at least 16 non-missing assessments in each 4-week interval, the score is missing. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| | | Time to reaching Itch NRS 4-pt improvement (primary censoring rule) | First time reaching Itch NRS 4-point improvement as Event, excluding data after rescue, treatment discontinuation. | Use observed value if no event happened data will be censored at rescue, treatment discontinuation or Week 16 visit date, whichever occurs first |
| Skin Pain Numeric Rating Scale (NRS) | Skin Pain NRS is a patient-administered, 11-point horizontal scale anchored at 0 and 10, with 0 representing "no pain" and 10 representing "worst pain imaginable." Overall severity of a patient's skin pain is indicated by selecting the number that best describes the worst level of skin pain in the past 24 hours Refer to Section 6.2.2 for details on how to calculate the weekly score which will be used in the continuous analysis. | Skin Pain NRS score | Single item; range 0 to 10. Refer to Section 6.2.2 on how to derive the visit score. | Refer to Section 6.2.2 on how to derive the visit score. |
| | | Change from baseline in Skin Pain NRS | Change from baseline: observed skin pain score – baseline skin pain score | Missing if baseline or observed value is missing. |
| | | Pain-free days (Skin pain NRS = 0) | Count of observed value = 0 for 28-day (4-week) intervals starting on the day of the first study drug administration. This will be calculated for the following visit intervals: baseline to Week 4, Week 4 to Week 8, Week 8 to Week 12 and Week 12 to Week 16. Thus, if Day 1 is defined as the day of first study drug administration, the baseline to Week 4 assessment is based on Day 1 to Day 28, Week 4 to Week 8 is based on Day 29 to Day 56, etc. | If patients do not have at least 16 non-missing assessments in each 4-week interval, the score is missing. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| Atopic Dermatitis Sleep Scale (ADSS) | The ADSS is a 3-item, patient-administered questionnaire developed to assess the impact of itch on sleep including difficulty falling asleep, frequency of waking, and difficulty getting back to sleep last night. Patient's rate their difficulty falling asleep and difficulty getting back to sleep, items 1 and 3, respectively, using a 5-point Likert-type scale with response options ranging from 0 "not at all" to 4 "very difficult." Patients report their frequency of waking last night, item 2, by selecting the number of times they woke up each night, ranging from 0 to 29 times. The ADSS is designed to be completed each day with respondents thinking about sleep "last night." Each item is scored individually. | ▪ Item 1 score of ADSS<br>▪ Item 2 score of ADSS<br>▪ Item 3 score of ADSS | Single items: Item 1, range 0 to 4; Item 2, range 0 to 29; Item 3, range 0 to 4. Refer to Section 6.2.2 on how to derive the visit score. | Refer to Section 6.2.2 on how to derive the weekly visit score. |
| | | ▪ Change from baseline in score of Item 1 of ADSS<br>▪ Change from baseline in score of Item 2 of ADSS<br>▪ Change from baseline in score of Item 3 of ADSS | Change from baseline: observed ADSS item score – baseline ADSS item score | Missing if baseline or observed value is missing. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| Patient- Oriented Eczema Measure (POEM) | The POEM is a simple, 7-item, patient-administered scale that assesses disease severity in children and adults. Patients respond to questions about the frequency of 7 symptoms (itching, sleep disturbance, bleeding, weeping/oozing, cracking, flaking, and dryness/roughness) over the last week. Response categories include "No days," "1-2 days," "3-4 days," "5-6 days," and "Every day" with corresponding scores of 0, 1, 2, 3, and 4, respectively. Scores range from 0-28 with higher total scores indicating greater disease severity (Charmin et al. 2004). | POEM score | POEM total score: sum of questions 1 to 7, Range 0 to 28. | If a single question is left unanswered, then that question is scored as 0. If more than 1 question is left unanswered, then the tool is not scored. If more than 1 response is selected, then the response with the highest score is used. |
| | | Change from baseline in POEM score | Change from baseline: observed POEM score – baseline POEM score | Missing if baseline or observed value is missing. |
| Patient Global Impression of Severity–Atopic Dermatitis (PGI-S-AD) | The PGI-S-AD is a single-item question asking the patient how they would rate their overall AD symptoms over the past 24 hours. The 5 categories of responses range from "no symptoms" to "severe." | PGI-S-AD score | Single item. Range 1 to 5. Refer to Section 6.2.2 on how to derive the visit score. | Refer to Section 6.2.2 on how to derive the visit score. |
| | | Change from baseline in PGI-S-AD | Change from baseline: observed PGI-S-AD score – baseline PGI-S-AD score | Missing if baseline or observed value is missing. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| Patient-Reported Outcomes Measurement Information System (PROMIS) | PROMIS is a set of person centered measures that evaluates and monitors physical, mental, and social health in adults and children. Both the anxiety short form and the depression short form are available in a pediatric self-report (ages 8 to <18 years) and for parents/caregivers serving as proxy reporters for their children (youth ages ≥5 years). Children aged <5 years will not complete this assessment. Both pediatric self-report and proxy-report versions assess depression or anxiety "in the past seven days." Response options range from 1 = Never; 2 = Rarely; 3 = Sometimes; 4 = Often; to 5 = Almost always. Total raw scores are converted to T-Scores with higher scores representing greater depression or anxiety. | PROMIS Anxiety domain score<br><br>PROMIS Depression domain score | Anxiety domain score and depression domain score will be derived from an Item Response Theory (IRT) model. The model and scoring program are provided by Northwestern University. The resulting T-scores will have population mean and standard deviation of 50 and 10, with higher score indicating greater depression or anxiety. | N/A – partial assessments cannot be saved. |
| | | Change from baseline in PROMIS domain | Change from baseline: observed PROMIS domain score – baseline PROMIS domain score | Missing if baseline or observed value is missing. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| Infant's Dermatitis Quality of Life Index (IDQOL) | IDQOL is a simple, caregiver-administered, 11-question, validated, quality-of-life questionnaire designed for use in pediatric patients <4 years old with AD (Lewis-Jones et al. 2001; Basra et al. 2013). It covers 2 domains, including Dermatitis Severity and Life Quality Index. The recall period is over the "last week." Response categories for the Dermatitis Severity domain include "None," "Fairly good," "Average," "Severe," and "Extremely severe" with corresponding scores of 0, 1, 2, 3, and 4, respectively. The Life Quality Index domain response categories vary for individual questions; however, each question has 4 response options with corresponding scores ranging from 0-3. Total scores of the Life Quality Index range from 0-30 with higher scores indicating greater impairment of quality of life. The Dermatitis Severity is scored separately and can be correlated with the IDQOL. An IDQOL total score of 0 to 1 is considered as having no effect on a child's life. | IDQOL Score for Dermatitis Severity and Life Quality Index | Dermatitis Severity domain score is the result of a single question, range 0 to 4;<br>Life quality index score is sum of the 10 life quality questions, range 0 to 30. | N/A – partial assessments cannot be saved. |
| | Change from baseline in IDQOL Life Quality Index | Change from baseline: observed IDQOL Life Quality Index score – baseline IDQOL Life Quality Index score | Missing if baseline or observed value is missing. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| Children's Dermatology Life Quality Index (CDLQI) | The Children's Dermatology Life Quality Index (CDLQI) is a simple, patient-administered, 10-question, validated, quality-of-life questionnaire that is designed for use in children ≥4 years old that covers 6 domains including symptoms and feelings, leisure, school or holidays, personal relationships, sleep, and treatment (Lewis-Jones and Finlay1995). The recall period is over the "last week." Response categories include "not at all," "only a little," "quite a lot," and "very much," with corresponding scores of 0, 1, 2, and 3, respectively, with unanswered ("not relevant") responses scored as 0 and "Prevented School" scored as 3. Scores range from 0 to 30 with higher scores indicating greater impairment of quality of life. A CDLQI total score of 0 to 1 is considered as having no effect on a child's life (Waters et al. 2010). | Symptoms and feelings domain | Sum of questions 1 and 2, range 0 to 6. | N/A – partial assessments cannot be saved. |
| | | Leisure domain | Sum of questions 4, 5, and 6, range 0 to 9. | N/A – partial assessments cannot be saved. |
| | | School or holidays domain | Question 7, range 0 to 3. | N/A – partial assessments cannot be saved. |
| | | Personal relationships | Sum of questions 3 and 8, range 0 to 6. | N/A – partial assessments cannot be saved. |
| | | Sleep | Question 9, range 0 to 3. | N/A – partial assessments cannot be saved. |
| | | Treatment domain | Question 10, range 0 to 3. | N/A – partial assessments cannot be saved. |
| | | CDLQI total score | CDLQI total score: sum of all CDLQI domain scores, range 0 to 30. | N/A – partial assessments cannot be saved. |
| | | Change from baseline in CDLQI total score and domain scores | Change from baseline: observed CDLQI score – baseline CDLQI score | Missing if baseline or observed value is missing. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| Work Productivity and Activity Impairment: Atopic Dermatitis - Caregiver (WPAI-AD-CG) | The WPAI-AD-CG records impairment due to AD during the past 7 days. The WPAI-AD-CG consists of 6 items grouped into 4 domains: absenteeism (work time missed), presenteeism (impairment at work/reduced on-the-job effectiveness), work productivity loss (overall work impairment/ absenteeism plus presenteeism), and activity impairment. Scores are calculated as impairment percentages (Reilly et al. 1993, 1996), with higher scores indicating greater impairment and less productivity. | Employment status | Question (Q)1 | Single item, missing if missing. |
| | | Change in employment status | Employed at baseline and remained employed: Q1 = 1 at postbaseline visit and at baseline visit.<br>Not employed at baseline and remain unemployed: Q1 = 0 at postbaseline visit and at baseline visit. | Missing if baseline or observed value is missing. |
| | | Percentage of absenteeism | Percent work time missed due to problem: (Q2/(Q2 + Q4))*100 | If Q2 or Q4 is missing, then missing. |
| | | Change from baseline in absenteeism | Change from baseline: observed absenteeism – baseline absenteeism | Missing if baseline or observed value is missing. |
| | | Percentage of presenteeism | Percent impairment (reduced productivity while at work) while working due to problem: (Q5/10)*100 | If Q5 is missing, then missing. |
| | | Change from baseline in presenteeism | Change from baseline: observed presenteeism – baseline absenteeism | Missing if baseline or observed value is missing. |
| | | Overall work impairment | Percent overall work impairment (combines absenteeism and presenteeism) due to problem: (Q2/(Q2+Q4) + [(1-Q2/(Q2+Q4))*(Q5/10)])*100 | If Q2, Q4, or Q5 is missing, then missing. |
| | | Change from baseline in work impairment | Change from baseline: observed work impairment – baseline work impairment | Missing if baseline or observed value is missing. |
| | | Percentage of impairment in activities | Percent activity impairment (performed outside of work) due to problem: (Q6/10)*100 | If Q6 is missing, then missing. |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| | | Change from baseline in impairment in activities | Change from baseline: observed impairment in activities – baseline impairment in activities | Missing if baseline or observed value is missing. |
| European Quality of Life–5 Dimensions – Youth (EQ-5D-Y) | | ▪ EQ-5D-Y mobility<br>▪ EQ-5D-Y self-care<br>▪ EQ-5D-Y usual activities<br>▪ EQ-5D-Y pain/ discomfort<br>▪ EQ-5D-Y anxiety/ depression | Five health profile dimensions, each dimension has 3 levels:<br>1 = no problems<br>2 = some problems<br>3 = a lot of problems<br>It should be noted that the numerals 1 to 3 have no arithmetic properties and should not be used as a primary score. | Each dimension is a single item, missing if missing. |
| | | EQ-5D-Y VAS | Single item. Range 0 to 100.<br>0 represents “worst health you can imagine”<br>100 represents “best health you can imagine” | Single item, missing if missing. |
| | | Change from baseline in EQ-5D-Y VAS | Change from baseline: observed EQ-5D-Y VAS score – baseline EQ-5D-Y VAS score | Missing if baseline or observed value is missing. |

| | | | | |
|---|---|---|---|---|
| | The European Quality of Life-5 Dimensions-Youth version (EQ-5D-Y) is a widely used, generic questionnaire that assesses health status "today" (EuroQol 2014). The EQ-5D-Y is self-completed for pediatric patients ≥8 years old and is completed by parents/ caregivers (proxy) for children 4 to <8 years old. This assessment will not be completed for children <4 years old per developer recommendation. The questionnaire consists of 2 parts: the first part assesses 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) that have 3 possible levels of response (no problems, some problems, or a lot of problems). This part of the EQ-5D-Y can be used to generate a health state index score, which is often used to compute QALY for utilization in health economic analyses. The health state index score is calculated based on the responses to the 3 dimensions, providing a single value on a scale from less than 0 (where zero is a health state equivalent to death; negative values are valued as worse than dead) to 1 (perfect health), with | | | |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| | higher scores indicating better health utility. The second part of the questionnaire consists of a visual analog scale on which the patient rates their perceived health state from 0 ("the worst health you can imagine") to 100 ("the best health you can imagine"). Published studies by EuroQol Group members showed preliminary evidence of the instrument's feasibility, reliability, and validity (Ravens-Sieberer et al. 2010) | | | |
| Parent-Reported Itch Severity Measure (PRISM) | The Parent-Reported Itch Severity Measure (PRISM) is a single-item, parent/caregiver administered scale that reports the overall severity of their child's itching. Parent/Caregiver's report the overall severity of their child's itching based on observed actions of the child in the past 24 hours. Response options range include "No Itch," "Mild," "Moderate," "Severe," and "Very Severe." The PRISM will be completed for patients <10 years old by the parent/caregiver. | PRISM Score | Single question 1 to 5 | N/A – partial assessments cannot be saved. |
| | | Change from baseline in PRISM Score | Change from baseline: observed PRISM score – baseline PRISM score | Missing if baseline or observed value is missing. |
| | | 2-point improvement in subgroup of patients with baseline PRISM≥3 | Change from baseline ≤-2 and baseline ≥3 | Missing if baseline is missing or <3 or observed value is missing. |
| | | Itch-free days (PRISM score=1) | Count of observed value = 0 for 28-day (4-week) intervals starting on the day of the first study drug administration. This will be calculated for the following visit intervals: baseline to Week 4, Week 4 to Week 8, Week 8 to Week 12 and Week 12 to Week 16. Day 1 is defined as the day of first study drug administration, | If patients do not have at least 16 non-missing assessments in each 4-week |

| Measure | Description | Variable | Derivation / Comment | Imputation Approach if Missing Components |
|---|---|---|---|---|
| | | | therefore, the baseline to Week 4 assessment is based on Day 1 to Day 28, Week 4 to Week 8 is based on Day 29 to Day 56, etc. | interval, the score is missing. |
| Dermatitis Family Impact Questionnaire (DFI) | The Dermatitis Family Impact (DFI) questionnaire is a simple, caregiver-administered, 10 question, validated, quality-of-life questionnaire that is designed to assess the impact of AD on the quality of life of the parents and family members of children with AD (Lawson et al. 1998; Dodington et al. 2013). The recall period is over the "last week." Response options include "Not at all," "A little," "A lot," and "Very much," with corresponding scores of 0, 1, 2, and 3, respectively. Scores range from 0 to 30 with higher scores indicating greater impairment of quality of life. | DFI score | DFI score is sum of the ten quality-of-life questions, range 0 to 30. | N/A – partial assessments cannot be saved. |
| | | Change from baseline in DFI | Change from baseline: observed DFI score – baseline DFI score | Missing if baseline or observed value is missing |
| Missed school days | Missed school day will be recorded in daily dairy. For patients 10 to <18 years old, the question is patient rated. For patients 2 to 10 years old, the question is parent/caregiver rated | Percentage of missed school days | Percentage of missed school days is calculated as total missed school day/ expected school days *100% | Single item, missing if missing. |

**Table JAIP.6.9. Description of Health Outcomes and Quality-of-Life Measures Analyses**

| Measure | Variable | Analysis Method (Section 6.2.3) | Population (Section 6.2.1) | Comparison/Time Point | Analysis Type |
|---|---|---|---|---|---|
| Itch Numeric Rating Scale (NRS) | • Itch NRS score<br>• Change from baseline in Itch NRS score<br>• Percent change from baseline Itch score | MMRM | ITT- Patients ≥10 years old | Bari high or Bari mid or Bari low dose vs. PBO; Week 1, 4, 16 | Secondary Analysis |
| | | ANCOVA using mLOCF | ITT- Patients ≥10 years old | Bari high or Bari mid or Bari low dose vs. PBO; Week 1, 4, 16 | Sensitivity Analysis |
| | Proportion of patients achieving a 4-point improvement in Itch NRS | Logistic regression using NRI | ITT – Patients ≥10 years old with baseline Itch NRS≥4 | Bari high or Bari mid or Bari low dose vs. PBO; Week 1, 2, 4, 16 | Key Secondary Analysis |
| | | | | Combined Bari high and mid dose vs. PBO; Week 16 | Exploratory analysis |
| | | | PPS | Bari high or Bari mid or Bari low dose vs. PBO; Week 1,2,4, 16 | Sensitivity analysis |
| | | Logistic regression using pMI | ITT– Patients ≥10 years old with baseline Itch NRS≥4 | Bari high or Bari mid or Bari low dose vs. PBO; Week 1,2,4,16 | Sensitivity analysis |
| | Number of Itch-free (Itch NRS = 0) Days | Descriptive statistics | ITT | No comparisons; Week 12 to 16 | Exploratory Analysis |
| | Time to 4-point reduction in Itch NRS (in the subset of patients who had baseline Itch NRS ≥4 | Kaplan-Meier Curves (primary censoring rule) | ITT – Patients ≥10 years old with baseline Itch NRS≥4 | Bari high or Bari mid or Bari low dose vs. PBO; | Exploratory Analysis |
| Skin Pain Numeric Rating Scale (NRS) | • Skin Pain NRS score | MMRM | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Secondary Analysis |

| Measure | Variable | Analysis Method (Section 6.2.3) | Population (Section 6.2.1) | Comparison/Time Point | Analysis Type |
|---|---|---|---|---|---|
| | • Change from baseline in Skin Pain NRS score | ANCOVA using mLOCF | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Sensitivity Analysis |
| | Number of Skin Pain-free (Skin pain NRS = 0) Days | Descriptive statistics | ITT | No comparisons; Week 12 to 16 | Exploratory Analysis |
| Atopic Dermatitis Sleep Scale (ADSS) | • ADSS item scores<br>• Change from baseline in ADSS item scores | MMRM | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 1, 16 | Secondary Analysis |
| | | ANCOVA using mLOCF | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 1, 16 | Sensitivity Analysis |
| Patient-Oriented Eczema Measure (POEM) | • POEM score<br>• Change from baseline in POEM score | MMRM | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Secondary Analysis |
| | | ANCOVA using mLOCF | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Sensitivity Analysis |
| Patient Global Impression of Severity–Atopic Dermatitis (PGI-S-AD) | • PGI-S-AD score<br>• Change from baseline in PGI-S-AD score | MMRM | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Secondary Analysis |
| | | ANCOVA using mLOCF | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Sensitivity Analysis |
| Patient-Reported Itch Severity Measure (PRISM) | • PRISM scores<br>• Change from baseline in PRISM score | MMRM | ITT – patients < 10 years old and baseline PRISM score ≥ 3 | Bari high or Bari mid or Bari low dose vs. PBO; Weeks 1, 2, 4, and 16 | Secondary Analysis |
| | | ANCOVA using mLOCF | ITT – patients < 10 years old and baseline PRISM score ≥ 3 | Bari high or Bari mid or Bari low dose vs. PBO; Weeks 1, 2, 4, and 16 | Sensitivity Analysis |

| Measure | Variable | Analysis Method (Section 6.2.3) | Population (Section 6.2.1) | Comparison/Time Point | Analysis Type |
|---|---|---|---|---|---|
| | Proportion of patients achieving a 2-point improvement in PRISM | Logistic regression using NRI | ITT- patients < 10 years old and baseline PRISM score ≥ 3 | Bari high or Bari mid or Bari low dose vs. PBO; Weeks 1, 2, 4, and 16 | Secondary analysis |
| Children's Dermatology Life Quality Index (CDLQI) | • CDLQI total score<br>• Change from baseline in total CDLQI scores<br>• Observed and change from baseline in domain scores<br>-Symptoms and feelings<br>-Leisure<br>-School or holidays<br>-Personal relationships<br>-Sleep<br>-Treatment | MMRM | ITT – patients ≥4 years old | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Secondary Analysis |
| | | ANCOVA using mLOCF | ITT – patients ≥4 years old | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Sensitivity Analysis |
| Infants' Dermatitis Quality of Life Index (IDQOL) | • IDQOL total score<br>• Change from baseline in IDQOL total scores | MMRM | ITT - – patients <4 years old | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Secondary Analysis |
| | | ANCOVA using mLOCF | ITT - – patients <4 years old | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Sensitivity Analysis |
| Work Productivity and Activity Impairment: Atopic Dermatitis – Caregiver (WPAI-AD-CG) | Observed and Change from baseline in:<br>• absenteeism<br>• presenteeism<br>• overall work impairment<br>• impairment in activities | MMRM | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Secondary Analysis |
| | | ANCOVA using mLOCF | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Sensitivity Analysis |

| Measure | Variable | Analysis Method (Section 6.2.3) | Population (Section 6.2.1) | Comparison/Time Point | Analysis Type |
|---|---|---|---|---|---|
| European Quality of Life-5 Dimensions–Youth (EQ-5D-Y) | Observed level as "no problems" in<br>• EQ-5D-Y mobility<br>• EQ-5D-Yself-care<br>• EQ-5D-Y usual activities<br>• EQ-5D-Y pain/ discomfort<br>• EQ-5D-Y anxiety/ depression | Logistic Regression using NRI | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Exploratory Analysis |
| | Observed and Change from baseline in<br>• EQ-5D-Y VAS | MMRM | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Secondary Analysis |
| | | ANCOVA using mLOCF | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Sensitivity Analysis |
| Dermatitis Family Impact Questionnaire (DFI) | • DFI score<br>• Change from baseline in DFI score | MMRM | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Secondary Analysis |
| | | ANCOVA using mLOCF | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Sensitivity Analysis |
| Patient-Reported Outcomes Measurement Information System (PROMIS) | • PROMIS domain scores<br>• Change from baseline in PROMIS domain | MMRM | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Secondary Analysis |
| | | ANCOVA using mLOCF | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Sensitivity Analysis |
| Missed school days | Percentage of missed school days | Descriptive statistics | ITT | Bari high or Bari mid or Bari low dose vs. PBO; Week 16 | Secondary Analysis |

Abbreviations: ANCOVA = analysis of covariance; Bari = baricitinib; CPH = Cox proportional hazard; ITT = intent-to-treat; mLOCF = modified last observation carried forward; MMRM = mixed model repeated measures; NRI = nonresponder imputation; PBO = placebo; pMI=placebo multiple imputation; PPS = per protocol set; vs. = versus.

## 6.14. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

Pharmacokinetic (PK), pharmacodynamic (PD), and biomarker analyses to address secondary and exploratory objectives of this study will be described by Lilly in separate PK/PD and biomarker analysis plans.

## 6.15. Evaluation of Immunological Measures

Patients who are immunized with tetanus, diphtheria toxoids, and acellular pertussis (TDaP) or pneumococcal conjugate vaccines will have their immunoglobulin G (IgG) antibody titers to the antigens evaluated preimmunization and at 4 and 12 weeks postimmunization. Change of IgG titers from pre-vaccination will be evaluated and summarized using descriptive statistics. A primary immune response will be assessed in patients who have never received TDaP or pneumococcal conjugate vaccines previously, and secondary/booster responses will be assessed if the patients have previously received the vaccines.

## 6.16. Safety Analyses

The general methods used to summarize safety data, including the definition of baseline value are described in Section 6.2.

Safety analyses during the first 16 weeks (Study Period 3) will include data before and after rescue, unless otherwise stated, and patients will be analyzed according to the investigational product to which they were randomized at Visit 2.

For long-term safety analysis, All Bari population and Extended Bari population will be used for long-term safety analysis. In addition, the Exposure-Adjusted Incidence Rate (EAIR) for adverse event, shift tables and treatment-emergent summary of laboratory evaluations will be reported for both All Bari and Extended Bari populations. Statistical comparison will be conducted between Bari doses in Extended Bari populations.

To support CSS only, the Exposure-Adjusted Incidence Rate (EAIR) for adverse events, shift tables and treatment-emergent summary of laboratory evaluations will be reported for safety population during Study Period 3.

Safety topics that will be addressed include the following: AEs, clinical laboratory evaluations, vital signs and physical characteristics, Columbia Suicide Severity Rating Scale (C-SSRS), the Self-Harm Supplement Form, safety in special groups and circumstances, including adverse events of special interest (AESI) (see Section 6.16.7), and investigational product interruptions.

Long term safety analysis topics will include AEs (including AEs of special interest), clinical laboratory evaluations, vital signs and physical characteristics and growth including imaging (x-ray).

Unless otherwise specified, by-visit summaries will only include planned on-treatment visits. For tables that summarize events (such as AEs, categorical lab abnormalities, shift to maximum value), post-last dose follow-up data will be included. Follow-up data is defined as all data occurring up to 30 days (planned maximum follow-up time) after last dose of treatment, regardless of study period. Applicable listings will include all safety data from all periods.

For selected safety assessments other than events, descriptive statistics may be presented for the last measure observed during post-treatment follow-up (up to 30 days after the last dose of treatment including rescue, regardless of study period).

### *6.16.1. Criteria for Notable Patients*

Below additional criteria will be used to identify notable patients in addition to compound level PSAP Section 5.2.2.

- ALT or AST >8× ULN
- ALT or AST >5× ULN for more than 2 weeks
- ALT or AST >3× ULN and total bilirubin level (TBL) >2× ULN or international normalized ratio (INR) >1.5
- ALT or AST >3× ULN with the appearance of fatigue, nausea, vomiting, right upper quadrant pain or tenderness, fever, and/or rash
- ALP >3× ULN (unless allowed after discussion with Sponsor)
- ALP >2.5× ULN and TBL >2× ULN
- ALP >2.5× ULN with the appearance of fatigue, nausea, vomiting, right quadrant pain or tenderness, fever, and/or rash
- white blood cell count <1000 cells/ µ L (1.00x103/ µ L or 1.00 GI/L)
- ANC <500 cells/ µ L (0.50x103/ µ L or 0.50 GI/L)
- lymphocyte count <200 cells/ µ L (0.20x103/ µ L or 0.20 GI/L)
- hemoglobin <6.5 g/dL (<65.0 g/L).

### *6.16.2. Extent of Exposure*

Duration of exposure (in days) will be calculated as follows:

- Duration of exposure to investigational product (including exposure after the initiation of rescue therapy): *date of last dose of study drug (including rescue )– date of first dose of study drug + 1*.

Last dose of study drug including rescue is calculated as last date on study drug. For patients discontinuing study drug or study due to the reason "Lost to Follow-up," the duration of exposure is calculated as *date of second to last visit - date of first dose of study drug +1*.

Total patient-years (PY) of exposure (PYE) to study drug will be reported for overall duration of exposure. Descriptive statistics will be provided for patient-days of exposure and the frequency of patients falling into different exposure ranges in addition to cumulative exposures will be summarized.

Exposure ranges up to Study Period 4 will be summarized as follows:

- ≥28 days, ≥56 days, ≥84 days, and ≥112 days
- >0 to <28 days, ≥28 days to <56 days, ≥56 days to <84 days, ≥84 days to <112 days, and ≥112 days

Exposure ranges for the long term extension (Study Period 4) will be specified as below

- ≥ 4 weeks, ≥ 16 weeks, ≥ 24 weeks, ≥32 weeks, ≥ 52 weeks, then every 24 weeks
- >0 to <4 weeks, ≥4 weeks to <16 weeks, ≥16 weeks to <24 weeks, ≥24 weeks to <32 weeks, ≥32 weeks to <52 weeks, then 24-week intervals after 52 weeks.

Overall exposure for a treatment group will be summarized in total PY which is calculated according to the following formula:

- *Exposure in PYE = sum of duration of exposure in days/365.25*

### 6.16.3. Exposure-Adjusted Incidence Rate

For long-term safety analysis, the Exposure-Adjusted Incidence Rate (EAIR) will be employed for both All Bari and Extended Bari populations. The EAIR evaluating the incidence of a first event (patients with at least 1 event) per 100 patient-years at risk (PYR). Exposure will be calculated based on time at risk during the analysis period, which is defined as the treatment period plus 30 days off-drug follow-up time. Exposure time for a patient with an event will be terminated at the time of the first event. Exposure time for a patient without an event will be followed until the end of the analysis period.

### 6.16.4. Adverse Events

Adverse events are recorded in the eCRFs. The planned summaries are provided in Table JAIP.6.10 and are described more fully in compound-level safety standards.

The MedDRA Lowest Level Term (LLT) will be used in defining which events are treatment emergent. The maximum severity for each LLT during the baseline period up to first dose of the study medication will be used as baseline. If an event with missing severity is preexisting during the baseline period and persists during the treatment period, then the baseline severity will be considered mild for determining treatment-emergence. If an event occurring postbaseline has a missing severity rating, then the event is considered treatment-emergent unless the baseline rating is severe, in which case the event is not a treatment-emergent. For studies in which time is collected and where onset is on the day of the first dose of study treatment, the day and time for events will both be used to distinguish between pretreatment and posttreatment in order to derive treatment emergence.

**Table JAIP.6.10. Summary Tables Related to Adverse Events**

| Analysis |
|---|
| An overview table, with the number and percentage of patients in the safety set with death, an SAE, any TEAE, any TEAE by severity, discontinuation from the study due to an AE, and permanent discontinuation from study drug due to an AE |
| The percentages of patients with TEAEs will be summarized using MedDRA Preferred Term nested within System Organ Class. |
| The percentages of patients with TEAEs will be summarized using MedDRA Preferred Term (regardless of System Organ Class). |
| The percentages of patients with TEAEs will be summarized using MedDRA Preferred Term for the common TEAEs (occurring in any group ≥2%, before rounding, of treated patients). |
| The percentages of patients with TEAEs by maximum severity will be summarized using MedDRA Preferred Term for TEAEs. Only counts and percentages will be included for the TEAEs by maximum severity. |
| A listing of all deaths will be provided. Additional deaths that are reported outside of the treatment period will be obtained from the Lilly Safety System (LSS). |
| The number and percentage of patients who experienced an SAE (including deaths and SAEs temporally associated or preceding deaths) will be summarized using MedDRA Preferred Term nested within System Organ Class. |
| A listing of SAEs will be provided. |
| The number and percentage of patients who permanently discontinued from study drug due to an AE (including AEs that led to death) will be summarized using MedDRA Preferred Term nested within System Organ Class. |
| The number and percentage of patients who temporarily interrupted study drug due to an AE will be summarized using MedDRA Preferred Term nested within System Organ Class. |
| For CSS only, number and percentage of patients with TEAEs using MedDRA PT nested within pre-defined event clusters. |

Abbreviations: AE = adverse event; MedDRA = Medical Dictionary for Regulatory Activities; SAE = serious adverse event; TEAE = treatment emergent adverse event.

### *6.16.5. Clinical Laboratory Evaluation*

The planned summaries for clinical laboratory evaluations other than those included in AE of special interest are provided in Table JAIP.6.11 and are described more fully in compound-level safety standards. Analysis of laboratory evaluation pertinent to AESI will be further addressed in Section 6.16.7.

**Table JAIP.6.11. Summary Tables Related to Clinical Laboratory Evaluations**

| Analysis |
|---|
| Box plots for observed values |
| Box plots for change from baseline values |
| Tables with percentages of patients who are treatment-emergent high/low by treatment group |
| Listing of abnormal findings for laboratory analyte measurements, including qualitative measures |

### *6.16.6. Vital Signs and Other Physical Findings*

The planned summaries for vital signs (systolic blood pressure [BP], diastolic BP, pulse, weight) are provided in Table JAIP.6.12 and are described more fully in compound-level safety standards.

**Table JAIP.6.12. Summary Tables Related to Vital Signs**

| Analysis |
|---|
| Tables with percentages of patients who are treatment-emergent high/low by treatment group. The limits are defined in the compound-level safety standards and are based on literature. |
| Tables and box plots with observed value and change from baseline at each postbaseline visit |

For vital signs and physical characteristics, original-scale data will be analyzed. Mean changes from baseline and as incidence of abnormal values will be summarized. The observed values at each visit and change from baseline to each scheduled visit, respectively, will be displayed in box plots for patients in each treatment period with the corresponding safety population. These box plots will be used to evaluate trends over time and to assess a potential impact of outliers on central tendency summaries. The scheduled visits/measurements will be included. The unscheduled visits and the repeat measurements will be excluded.

The reference limits for blood pressure and pulse/heart rate are included in Table JAIP.6.13.

**Table JAIP.6.13. Categorical Criteria for Abnormal Treatment-Emergent Blood Pressure and Pulse Measurements for Children and Adolescents**

| Age (years) | | Systolic BP, mm Hg (supine or sitting forearm at heart level) | Diastolic BP, mm Hg (supine or sitting forearm at heart level) | Pulse/HR bpm (supine or sitting) |
|---|---|---|---|---|
| Infant <2 | Low | ≤70 (low limit) and decrease from lowest value during baseline ≥15 if >70 at each baseline visit | ≤35 (low limit) and decrease from lowest value during baseline ≥10 if >35 at each baseline visit | <70 (low limit) and decrease from lowest value during baseline ≥25 if ≥70 at each baseline visit |
| | High[a] | ≥108 (high limit) and increase from highest value during baseline ≥15 if <108 at each baseline visit | ≥74 (high limit) and increase from highest value during baseline ≥10 if <74 at each baseline visit | >190 (high limit) and increase from highest value during baseline ≥25 if ≤190 at each baseline visit |
| Child 2-4 | Low | ≤75 (low limit) and decrease from lowest value during baseline ≥15 if >75 at each baseline visit | ≤40 (low limit) and decrease from lowest value during baseline ≥10 if >40 at each baseline visit | <60 (low limit) and decrease from lowest value during baseline ≥25 if ≥60 at each baseline visit |
| | High[a] | ≥110 (high limit) and increase from highest value during baseline ≥15 if <110 at each baseline visit | ≥76 (high limit) and increase from highest value during baseline ≥10 if <76 at each baseline visit | >160 (high limit) and increase from highest value during baseline ≥25 if ≤160 at each baseline visit |
| Child 5-9 | Low | ≤80 (low limit) and decrease from lowest value during baseline ≥15 if >80 at each baseline visit | ≤45 (low limit) and decrease from lowest value during baseline ≥10 if >45 at each baseline visit | <60 (low limit) and decrease from lowest value during baseline ≥25 if ≥60 at each baseline visit |
| | High[a] | ≥119 (high limit) and increase from highest value during baseline ≥15 if <119 at each baseline visit | ≥78 (high limit) and increase from highest value during baseline ≥10 if <78 at each baseline visit | >150 (high limit) and increase from highest value during baseline ≥25 if ≤150 at each baseline visit |
| Child 10-12 | Low | ≤85 (low limit) and decrease from lowest value during baseline ≥20 if >85 at each baseline visit | ≤50 (low limit) and decrease from lowest value during baseline ≥10 if <50 at each baseline visit | <60 (low limit) and decrease from lowest value during baseline ≥25 if ≥60 at each baseline visit |
| | High[a] | ≥126 (high limit) and increase from highest value during baseline ≥20 if <126 at each baseline visit | ≥82 (high limit) and increase from highest value during baseline ≥10 if <82 at each baseline visit | >140 (high limit) and increase from highest value during baseline ≥25 if ≤140 at each baseline visit |
| Adolescent 13 - 17 | Low | ≤90 (low limit) and decrease from lowest value during baseline ≥20 if >90 at each baseline visit | ≤50 (low limit) and decrease from lowest value during baseline ≥10 if >50 at each baseline visit | <50 (low limit) and decrease from lowest value during baseline ≥15 if ≥50 at each baseline visit |
| | High[a] | ≥129 (high limit) and increase from highest value during baseline ≥20 if <129 at each baseline visit | ≥86 (high limit) and increase from highest value during baseline ≥10 if <86 at each baseline visit | >120 (high limit) and increase from highest value during baseline ≥15 if ≤120 at each baseline visit |

Abbreviations: BP = blood pressure; HR = heart rate.

[a] The high limit values shown in this table correspond to 95$^{th}$ percentile for the age group under the 2017 American College of Cardiology/American Heart Association task Force on Clinical Practice Guidelines revised criteria for hypertension. Values higher than 95$^{th}$ percentile are consistent with Stage 1 or Stage 2 hypertension. Under some circumstances it may be appropriate to conduct analyses considering only the change from baseline reference limit.

### 6.16.6.1. Standardized Growth

Weight, height, and BMI data will be merged to the Centers for Disease Control and Prevention (CDC) standard growth data (released in 2000) by age and gender in order to compare patients' growth with the standard. Z-score and standardized percentile of weight, height, and BMI at each visit will be calculated and compared to the 2000 CDC growth charts.

The z-score and percentile calculations are based on algorithms and data provided by the National Center for Health Statistics. The details are provided in the CDC website (https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm).

Baseline, observed mean and mean change from baseline, of actual measure, z-score and standardized percentile of weight, height, and BMI will be summarized for:

- Safety population in double-blind treatment period by treatment
- All BARI population and by quartiles of baseline percentiles
- Patients' mean observed value and mean change from baseline of weight, height, and BMI standardized percentile and z-score will be plotted using box plot versus investigational product exposure time.

By-patient listings of actual measures and change from baseline, z-scores, standardized percentiles in weight, height, and BMI for each visit will be provided.

Observed height velocity by sex and age group will be calculated at week 16 and every 24 weeks starting from week 52 in Study Period 4:

- (current height [cm] – previous height [cm])/interval (in days) between measurements × 365.25

Mean growth velocity will be plotted versus age group.

By-patient listings of head circumference, tibial length will be provided.

The number of patients who experience a treatment emergent loss of weight greater than or equal to 3% from baseline will be summarized by treatment group. The 3% is based on a conservative half of the percentage used to flag treatment emergent weight loss in adult studies. Refer to the compound level safety standard (program safety analysis plan) for references.

### 6.16.6.2. X-ray and Structure Data

For hand X-ray data, following analysis will be performed at Week 16 and Study Period 4.

Descriptive statistics will be used to summarize the skeletal age, and chronological age at baseline for all the patients who have a baseline value available.

The skeletal age can deviate from the chronological age calculated from the date of birth. Results will include a description of the skeletal age, the chronological age, and the difference between skeletal age and chronological age at baseline and post baseline when data are available.

Shift tables from baseline to postbaseline visits will be presented, for the difference between skeletal age and chronological age categories: chronological age - bone age <= -2 years, |chronological age - bone age| <2 years, chronological age - bone age >= 2 years.

Knee X-ray data will only be analyzed for long term safety for extended Bari and all Bari population. Descriptive statistics on age and ratio will be summarized for patients in each growth plate closure status, by location (Distal femur, Proximal tibia) and gender.

#### 6.16.6.3. Menarche Status

Summary of status of menarche will be provided for long term safety analysis. Specifically, descriptive statistics of age will be presented for patients in the following 4 categories: patients with menarche at baseline, patients without menarche at baseline, patients with menarche in postbaseline and patients without menarche in postbaseline.

### *6.16.7. Special Safety Topics, including Adverse Events of Special Interest*

In addition to general safety parameters, safety information on specific topics of special interest will also be presented. Additional special safety topics may be added as warranted. The topics outlined in this section include the protocol-specified AESI.

Safety information on special topics including AEs of special interest (AESIs) will be presented by treatment group for Safety Population in Double-Blind Period, as well as Extended Bari Population and All Bari Population for long term safety analysis.

In general, for topics regarding safety in special groups and circumstances, patient profiles and/or patient listings, where applicable, will be provided when needed to allow medical review of the time course of cases/events, related parameters, patient demographics, study drug treatment and meaningful concomitant medication use. In addition to the safety topics for which provision or review of patient data is specified, these will be provided when summary data are insufficient to permit adequate understanding of the safety topic.

#### 6.16.7.1. Abnormal Hepatic Tests

Analyses for abnormal hepatic tests will involve 4 laboratory analytes: alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin (TBL), and serum alkaline phosphatase (ALP). The number and percentage of patients with abnormal elevations in hepatic laboratory tests at any time will be summarized by treatment groups. Refer to the compound level safety standards for details on ALT, AST, and TBL.

Specifically in JAIP, the percentages of patients with an ALP or TBL measurement ≥2× the central laboratory ULN during the treatment period will be summarized for all patients with a postbaseline value and subset into 5 subsets: patients whose non-missing maximum baseline value is ≤1× ULN, patients whose maximum baseline is >1× ULN but <1.5×ULN, patients whose maximum baseline is >1.5× ULN but <2×ULN, patients whose maximum baseline value is ≥2× ULN, and patients whose baseline values are missing.

Hematologic changes will be defined based on clinical laboratory assessments. Treatment-emergent laboratory abnormalities occurring at any time during the treatment period and shift tables of baseline to maximum grade during the treatment period will be tabulated. Change from baseline by timepoint will also be provided. Refer to the compound level safety standards for details.

### 6.16.7.2. Lipids Effects

Lipids effects will be assessed through analysis of elevated total cholesterol, elevated low-density lipoprotein (LDL) cholesterol, decreased and increased high-density lipoprotein (HDL) cholesterol, and elevated triglycerides. Treatment-emergent laboratory abnormalities related to elevated total cholesterol, elevated triglycerides, elevated LDL cholesterol, and decreased and increased HDL cholesterol occurring at any time during the treatment period will be tabulated. Shift tables will show the number and percentage of patients based on baseline to the least desirable category during the treatment period, with baseline depicted by the least desirable category during the baseline period. Refer to the compound level safety standards for details.

Categorical analyses will be performed using Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents (2011) as shown in Table JAIP.6.14.

**Table JAIP.6.14. Acceptable, Borderline-High, and High Plasma Lipid, Lipoprotein, and Apolipoprotein Concentrations for Children and Adolescents**

| Category | Low (mg/dL)[a] | Acceptable (mg/dL) | Borderline-high (mg/dL)[a] | High (mg/dL)[a] |
|---|---|---|---|---|
| Total cholesterol | -- | <170 | 170 to 199 | ≥200 |
| LDL cholesterol | | <110 | 110 to 129 | ≥130 |
| Non-HDL cholesterol | | <120 | 120 to 144 | ≥145 |
| Apolipoprotein B | | <90 | 90 to 109 | ≥110 |
| Triglycerides | | | | |
| 0 to 9 years of age | | <75 | 75 to 99 | ≥100 |
| 10 to 19 years of age | | <90 | 90 to 129 | ≥130 |
| HDL Cholesterol | <40 | >45 | 40 to 45 | |
| Apolipoprotein A-1 | <115 | >120 | 120 | |

Abbreviations: HDL = high-density lipoprotein; LDL = low-density lipoprotein.

[a] Low cut point for HDL cholesterol and apolipoprotein A-1 represent approximately the 10$^{th}$ percentile. The cut points for high and borderline-high represent approximately the 95$^{th}$ and 75$^{th}$ percentiles, respectively.

### 6.16.7.3. Renal Function Effects

Effects on renal function will be assessed through analysis of elevated creatinine. Common Terminology Criteria for Adverse Events will be applied for laboratory tests related to renal effects. Shift tables will show the number and percentage of patients based on baseline to maximum during the treatment period, with baseline depicted by highest grade during the baseline period.

With each shift table, a shift table summary displaying the number and percentage of patients with maximum postbaseline results will be presented by treatment group for each treatment period. Refer to compound level safety standards for details.

### 6.16.7.4. Elevations in Creatine Phosphokinase (CPK)

Elevations in creatine phosphokinase (CPK) will be displayed in shift tables using CTCAE criteria and treatment-emergent adverse events potentially related to muscle symptoms will be analyzed, based on reported AEs. Refer to the compound level safety standards for details.

### 6.16.7.5. Infections

Infections will be defined using all the preferred terms (PTs) from the MedDRA Infections and Infestations System Organ Class (SOC). The TEAE infections will be further analyzed in terms of potential opportunistic infection, herpes zoster and herpes simplex. Summary of hepatitis B virus (HBV) DNA monitoring results and association between infection and neutropenia/lymphopenia will also be provided in the context of infections. The MedDRA terms used to identify aforementioned specific infections are listed in the compound-level safety standards.

Refer to the compound level safety standards for details.

### 6.16.7.6. Major Adverse Cardiovascular Events (MACE) and Other Cardiovascular Events

Major Adverse Cardiovascular Events (MACE) and other cardiovascular events will be adjudicated by an independent, external adjudication committee. All confirmed events after adjudication will be used for the analysis. The number and percentage of patients with positively-adjudicated MACE, other cardiovascular events, noncardiovascular death, and all-cause death, will be summarized by treatment group. A listing of the events sent for cardiovascular adjudication will be provided to include data concerning the MedDRA PT related to the event, the seriousness of the event, and the event outcome, along with the adjudicated result. Refer to the compound level safety standards for details.

### 6.16.7.7. Venous Thromboembolic (VTE) Events

Venous thromboembolic (VTE) events will be adjudicated by an independent, external adjudication committee. Venous and pulmonary artery thromboembolic events will be classified as deep vein thrombosis (DVT), pulmonary embolism (PE), or other peripheral venous thrombosis. All confirmed events after adjudication will be used for the analysis. The number and percentage of patients with a VTE, DVT/PE, DVT, PE, and other peripheral venous thrombosis, as positively adjudicated, will be summarized by treatment group.

A listing of the VTE events sent for adjudication will be provided to include data concerning the MedDRA PT related to the event, the seriousness of the event, and the event outcome, along with the adjudicated result.

Refer to the compound level safety standards for details.

### 6.16.7.8. Arterial Thromboembolic (ATE) Events

The number and percentage of patients with an ATE, as positively adjudicated, will be summarized.

A listing of the ATE events sent for adjudication will be provided to include data concerning the MedDRA PT related to the event, the seriousness of the event, and the event outcome, along with the adjudicated result.

Refer to the compound level safety standards for details.

### 6.16.7.9. Malignancies

Malignancies will be identified using terms from the Malignant tumors SMQ. Malignancies excluding nonmelanoma skin cancers (NMSC) and NMSC will be reported separately. All the cases identified by the Malignant tumors SMQ will be assessed through medical review to determine confirmed NMSC cases.

The number and percentage of patients with treatment-emergent malignancies excluding NMSC and NMSC will be summarized by treatment group.

Refer to the compound level safety standards for details.

### 6.16.7.10. Allergic Reactions/Hypersensitivities

A search for relevant events related to allergic reaction and hypersensitivity will be performed using the following SMQs:

- Anaphylactic reaction SMQ (20000021)
- Hypersensitivity SMQ (20000214)
- Angioedema SMQ (20000024)

Events that satisfy the queries will be listed, by temporal order within patient ID, and will include SOC, PT, SMQ event categorization including detail on the scope (narrow or broad), reported AE term, AE onset and end dates, severity, seriousness, outcome, etc. Summary of narrow and broad search for anaphylactic reactions, hypersensitivity and angioedema will be provided. The number and percentage of patients with treatment-emergent anaphylactic reactions, hypersensitivity and angioedema will be summarized by treatment group.

Refer to the compound level safety standards for details.

### 6.16.7.11. Gastrointestinal Perforations

Potential gastrointestinal (GI) perforations will be identified using terms from the GI perforations SMQ. Potential GI perforations identified by the SMQ search will be provided as a listing for internal review by the medical safety team. Each case will be assessed to determine whether it is a GI perforation. All confirmed events after medical review will be used for the analysis. A summary table based on medical review will be provided.

Refer to the compound level safety standards for details.

#### 6.16.7.12. Columbia Suicide Severity Rating Scale

The Columbia Suicide Severity Rating Scale (C-SSRS) will only be collected in patients ≥7 years old. A children's version of the C-SSRS will be completed for patients 7 to <12 years old, and an adolescent/adult version of the C-SSRS will be completed for patients 12 to <18 years old. Number and percentage of patients with suicidal ideation, suicidal behavior, and self-injurious behavior during treatment will be summarized by treatment group. Shift tables in C-SSRS from baseline to postbaseline will also be reported.

Refer to the compound level safety standards for details.

#### 6.16.7.13. Self-Harm Supplement Form and Self-Harm Follow-up Form

The Self-Harm Supplement Form is a single question to enter the number of suicidal behavior events, possible suicide behaviors, or nonsuicidal self-injurious behaviors. If the number of behavioral events is greater than zero, it will lead to the completion of the Self-Harm Follow-Up Form. The Self-Harm Follow-Up Form is a series of questions that provides a more detailed description of the behavior cases. A listing of the responses give on the Self-Harm Follow-Up Form will be provided.

#### 6.16.7.14. Covid-19 Adverse Events

Listing of Covid-19 events will be provided. Summary table of Covid-19 events by preferred terms will be provided for long term safety analysis.

## 6.17. Subgroup Analyses

### *6.17.1. Efficacy Subgroup Analyses*

Subgroup analyses comparing each dose of baricitinib to placebo will be performed on the ITT population at Week 16 using the primary censoring rule for the following:

- Proportion of patients achieving IGA 0 or 1
- Proportion of patients achieving EASI75 Response Rate
- Proportion of patients achieving Itch NRS 4-point improvement
- Proportion of patients achieving PRISM 2-point improvement

The following subgroups, categorized into disease-related characteristics and demographic characteristics, will be evaluated:

- Patient Demographic and Characteristics Subgroups:
  - Gender (male, female)
  - Age group (<10, ≥10 years old)
  - Age group (2 to <6, 6 to <10, 10 to <18)
  - Baseline weight: (<20kg, ≥20 to <60 kg, ≥60 kg)
  - Baseline BMI (<20kg/$m^2$, ≥20 to <30 kg/$m^2$, ≥30 kg/$m^2$)
  - Race: (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Multiple)

- Geographic Region Subgroups:
  - Region: (as defined in Table JAIP.5.1)
  - Specific regions (Europe, other)
  - Specific country (Japan, other)
- Previous and Concomitant Therapy Subgroups
  - Prior use of TCNI (yes, no)
    - TCNI inadequate response (yes, no)
    - TCNI intolerance (yes, no)
  - Prior use of TCS (yes, no)
    - TCS inadequate response (yes, no)
    - TCS intolerance (yes, no)
  - Prior systemic therapy use (yes, no)
- Baseline Disease-Related Characteristics Subgroup
  - Baseline disease severity (IGA score): 3, 4
- Other atopic conditions (asthma, allergic asthma, allergic conjunctivitis, food allergy, Allergic rhinitis ) [yes, no]
- Asthma (asthma, allergic asthma) [yes, no]
- Food allergy [yes, no]
- Allergic rhinitis/rhino-conjunctivitis (allergic conjunctivitis, allergic rhinitis) [yes, no]

Descriptive statistics will be provided for each treatment and stratum of a subgroup as outlined, regardless of sample size. If each level of a subgroup comprises ≥10% of the overall sample size, subgroup analyses for categorical outcomes will be performed using logistic regression using Firth's correction to accommodate (potential) sparse response rates. The model will include the categorical outcome as the dependent variable and stratification variables, treatment, subgroup, and treatment-by-subgroup interaction as explanatory variables. If stratification variables coincide with the subgroup variables, the corresponding stratification variable will be omitted. Missing data will be imputed using NRI (Section 6.3.1). The p-value from the logistic regression model will be reported for the interaction test and the subgroup test, unless the model did not converge. Response counts and percentages will be summarized by treatment for each subgroup category. The difference in percentages and 95% CI of the difference in percentages using the Newcombe-Wilson without continuity correction will be reported. The corresponding p-value from the Fisher's exact test will also be produced.

Additional subgroup analyses on efficacy may be performed as deemed appropriate and necessary.

### *6.17.2. Safety Subgroup Analysis*

Safety subgroup analysis for common TEAEs and AE overview will be summarized for Safety Population in Period 3. The common TEAEs will be presented by decreasing frequency of PT within SOC.

A logistic regression model with treatment, subgroup, and the interaction of subgroup-by-treatment included as factors will be used. The p-value from the logistic regression model will be

reported for the interaction test and the subgroup test, unless the model did not converge. The response variable will be each AE. Treatment group differences will be evaluated within each category of the subgroup using Fisher's exact test, regardless of whether the interaction is statistically significant.

The following subgroups will be analyzed:

- Age group (<10, ≥10 years old)
- Age group (2 to <6, 6 to <10, 10 to <18)
- Sex
- Race

For long term safety analysis, AE overview will be summarized by age group (<10, ≥10 years old). Analysis method will follow long term safety analysis described in Section 6.2.3.

### *6.17.3. Bayesian Analysis*

The primary analysis will include data from the entire study population 2 to <18 years old, and a frequentist approach will be used for the analysis. However, in the event that the older subgroup of patients (10 to <18 years old) completes the Double-blind Treatment period (Study Period 3) more than 6 months earlier than anticipated for the younger subgroup of patients (2 to <10 years old), each age subgroup is adequately powered and can be individually unblinded and analyzed as described in this section. This is to ensure that potential delay in completing the enrollment of the younger patients will not delay analysis and submission of the data from the older patients.

In the younger subgroup, the primary analysis will leverage results observed in the older subgroup through a Bayesian approach for the primary efficacy outcome at 16 weeks. In this approach, the posterior distribution of the treatment response of baricitinib 4 mg in older pediatric patients will be used in the construction of a proper prior distribution for baricitinib 2 mg in younger pediatric patients. A similar approach will be used for the baricitinib 1 and 0.5 mg in the younger subgroup, respectively, given that these doses are expected to provide exposures that are similar to 2- and 1-mg doses in older pediatric patients. In this type of prior, leveraging of information is done by carefully employing adaptive borrowing on the treatment and placebo responses, that is, baricitinib information will be borrowed from JAIP adolescents patients (age 10-18) to increase precision/power if the treatment effect in younger pediatric patients (age 2-10) is consistent with that in adolescents; Contrarily, if substantial differences are observed, no borrowing from adolescents cohort will be employed so that the prior distributions degenerate to a vague prior. Hence the prior has the characteristic that is robust when there is prior-data conflict.

For mathematical and implementation details, please refer to Appendix 1.

## 6.18. Protocol Deviations

Protocol deviations will be tracked by the clinical team, and their importance will be assessed by key team members during protocol deviation review meetings. Of all important protocol

deviations (IPDs) identified, a subset occurring during Period 3 with the potential to affect efficacy analyses will result in exclusion from the PPS population.

Potential examples of deviations include patients who receive excluded concomitant therapy, significant non-compliance with study medication (<80% of assigned doses taken, failure to take study medication and taking incorrect study medication), patients incorrectly enrolled in the study, and patients whose data are questionable due to significant site quality or compliance issues. Refer to a separate document for the important protocol deviations.

The Trial Issue Management Plan includes the categories and subcategories of important protocol deviations and whether or not these deviations will result in the exclusion of patients from per protocol set. The number and percentage of patients having IPD(s) will be summarized within category and subcategory of deviation by treatment group for Period 3 using the ITT population. Individual patient listings of IPDs will be provided. A summary of reasons patients were excluded from the PPS will be provided by treatment group.

## 6.19. Interim Analyses and Data Monitoring

A DMC will monitor the overall safety of this trial. The DMC will consist of members external to Lilly. This DMC will follow the rules defined in the DMC charter, focusing on potential and identified risks for this molecule and for this class of compounds. Data Monitoring Committee membership will include, at a minimum, specialists with expertise in pediatrics, statistics, and other appropriate specialties.

The DMC will be authorized to review unblinded results of analyses by treatment group prior to database lock, including study discontinuation data, AEs including SAEs, clinical laboratory data, vital sign and growth data, etc. The DMC may recommend continuation of the study as designed; temporary suspension of enrollment; or the discontinuation of a particular age cohort, dose regimen or the entire study. While the DMC may request to review efficacy data to investigate the benefit/risk relationship in the context of safety observations for ongoing patients in the study, no information regarding efficacy will be communicated. Moreover, the study will not be stopped for positive efficacy results nor will it be stopped for futility. Hence, no alpha is spent. Details of the DMC, including its operating characteristics, are documented in the DMC charter and further details are given in the Interim Analysis Plan in Section 6.19.1.

Besides DMC members, a limited number of pre-identified individuals may gain access to the limited unblinded data from Study Period 3, as specified in the unblinding plan, prior to the database lock for the primary analysis or final database lock to initiate the final population PK/PD model development processes or for preparation of regulatory documents, respectively. Information that may unblind the study during the analyses will not be reported to study sites or the blinded study team until the study has been unblinded.

Unblinding details will be specified in a separate unblinding plan document.

### *6.19.1. Interim Analysis Plan*

Analyses for the DMC will include listings and/or summaries of the following information:

- patient disposition, demographics, and baseline characteristics
- concomitant medications
- exposure
- AEs, to include the following:
  - TEAEs
  - SAEs, including deaths
  - selected special safety topics
- clinical laboratory results
- vital signs
- assessment of growth (changes in height and weight, BMI, imaging data)
- C-SSRS

Summaries will include TEAEs, SAEs, special topics AEs, and treatment-emergent high and low laboratory and vital signs in terms of counts and percentages where applicable. For continuous analyses, box plots of laboratory analytes will be provided by time point and summaries will include descriptive statistics.

Mean change from baseline of EASI score will be provided to DMC as efficacy data. Further details are given in the DMC charter.

## 6.20. Planned Exploratory Analyses

The planned exploratory analyses are described in Sections 6.10, 6.11, and 6.16. Additional exploratory analyses may be conducted and will be documented in a supplemental SAP. Health Technology Assessment (HTA) toolkit analyses, which may be produced, will also be documented in the supplemental SAP.

## 6.21. Annual Report Analyses

Annual report analyses, such as the Development Safety Update Report (DSUR), will be documented in a separate analysis plan.

## 6.22. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include a summary of AEs, provided as a dataset which will be converted to an XML file. Both SAEs and 'Other' AE are summarized: by treatment group, by MedDRA PT.

- An AE is considered "Serious" whether or not it is a TEAE.
- An AE is considered in the "Other" category if it is both a TEAE and is not serious. For each SAE and "Other" AE, for each term and treatment group, the following are provided:
  - the number of participants at risk of an event
  - the number of participants who experienced each event term
  - the number of events experienced.

- Consistent with www.ClinicalTrials.gov requirements, “Other” AEs that occur in fewer than 5% of patients/subjects in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- AE reporting is consistent with other document disclosures for example, the CSR, manuscripts, and so forth.

Similar methods will be used to satisfy the European Clinical Trials Database (EudraCT) requirements.

# 7. Unblinding Plan

Refer to a separate blinding and unblinding plan document for details.

# 8. References

Alosh M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med.* 2014;33(4):693-713.

Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J.* 2011;53(6):894-913.

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(12):1513-1519.

EuroQol Group. EQ-5D-Y User Guide. Version 2.1. Available at: https://euroqol.org/wp-content/uploads/2019/10/EQ-5D-Y-User-Guide.pdf. Accessed February 2020.

Gillings D, Koch G. The application of the principle of intention-to-treat to the analysis of clinical trials. *Drug Information Journal.* 1991;25(3):411-424.

Hanifin JM, Thurston M, Omoto M, Cherill R, Tofte SJ, Graeber M. The Eczema Area and Severity Index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. *Exp Dermatol*. 2001;10(1):11-18.

[ICH E9 R1] Estimands and Sensitivity Analysis in Clinical Trials. Available at: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf. Accessed February 2020.

Kimball AB, Naegeli AN, Edson-Heredia E, Lin CY, Gaich C, Nikaï E, Wyrwich K, Yosipovitch G. Psychometric properties of the Itch Numeric Rating Scale in patients with moderate-to-severe plaque psoriasis. *Br J Dermatol.* 2016;175(1):157-162.

Kunz B, Oranje AP, Labrèze L, Stalder JF, Ring J, Taïeb A. Clinical validation and guidelines for the SCORAD index: consensus report of the European Task Force Atopic Dermatitis. *Dermatology*. 1997;195(1):10-19.

Lewis-Jones MS, Finlay AY. The Children's Dermatology Life Quality Index (CDLQI): initial validation and practical use. *Br J Dermatol.* 1995;132(6):942-949.

Lewis-Jones MS, Finlay AY, Dykes PJ. The Infants' Dermatitis Quality of Life Index. *Br J Dermatol.* 2001;144(1):104-110.

Naegeli AN, Flood E, Tucker J, Devlen J, Edson-Heredia E. The Worst Itch Numeric Rating Scale for patients with moderate to severe plaque psoriasis or psoriatic arthritis. *Int J Dermatol.* 2015;54(6):715-722.

Program Safety Analysis Plan: Baricitinib (LY3009104) Version 7. Report on file, Eli Lilly and Company.

Protocol I4V-MC-JAIP(a). A Phase 3, Multicenter, Randomized, Double-blind, Placebo-controlled, Parallel-group, Outpatient Study Evaluating the Pharmacokinetics, Efficacy, and Safety of Baricitinib in Pediatric Patients with Moderate-to-Severe Atopic Dermatitis. Report on file, Eli Lilly and Company.

Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, Egmar AC, Gusi N, Herdman M, Jelsma J, Kind P, Olivares PR, Scalone L, Greiner W. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res*. 2010;19(6):887-97.

Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics*. 1993;4(5):353-365.

Schram ME, Spuls PI, Leeflang MM, Lindeboom R, Bos JD, Schmitt J. EASI, (objective) SCORAD and POEM for atopic eczema: responsiveness and minimal clinically important difference. *Allergy.* 2012;67(1):99-106.

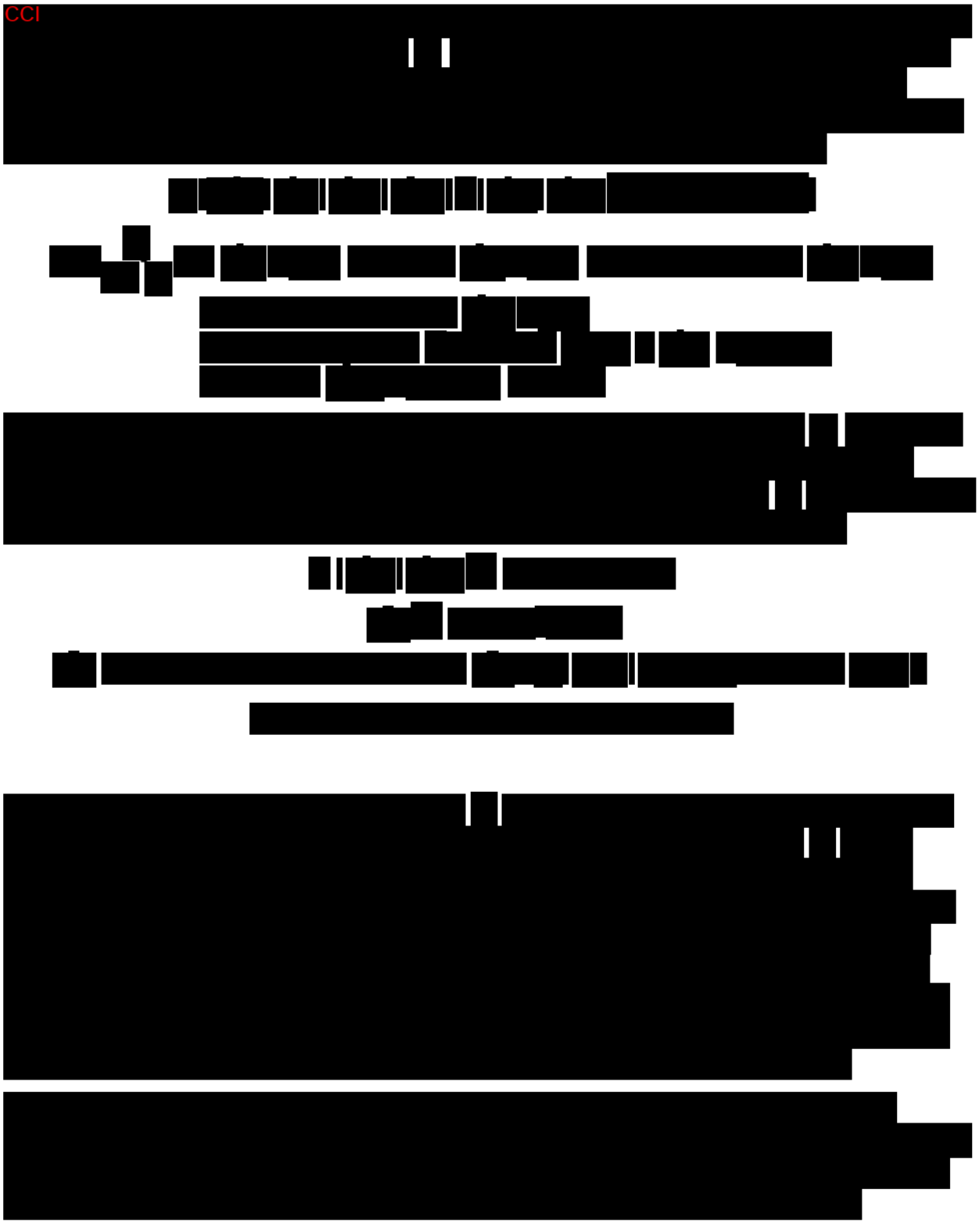
Stalder J, Taieb A, Atherton D, Bieber P, Bonifazi E, Broberg A, Calza A, Coleman R, De Prost Y, Stalder J, Gelmetti C, Cuannetti A, Harper J, Kunz B, Lachapelle J, Langeland T, Lever R, Oranje A, Oueille-Roussel C, Revuz J, Ring J, Roujeau J, Saurat J, Song M, Tennstedt D, Van Neste D, Vieluf D, Poncet M. Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. *Dermatology*. 1993;186(1):23-31.

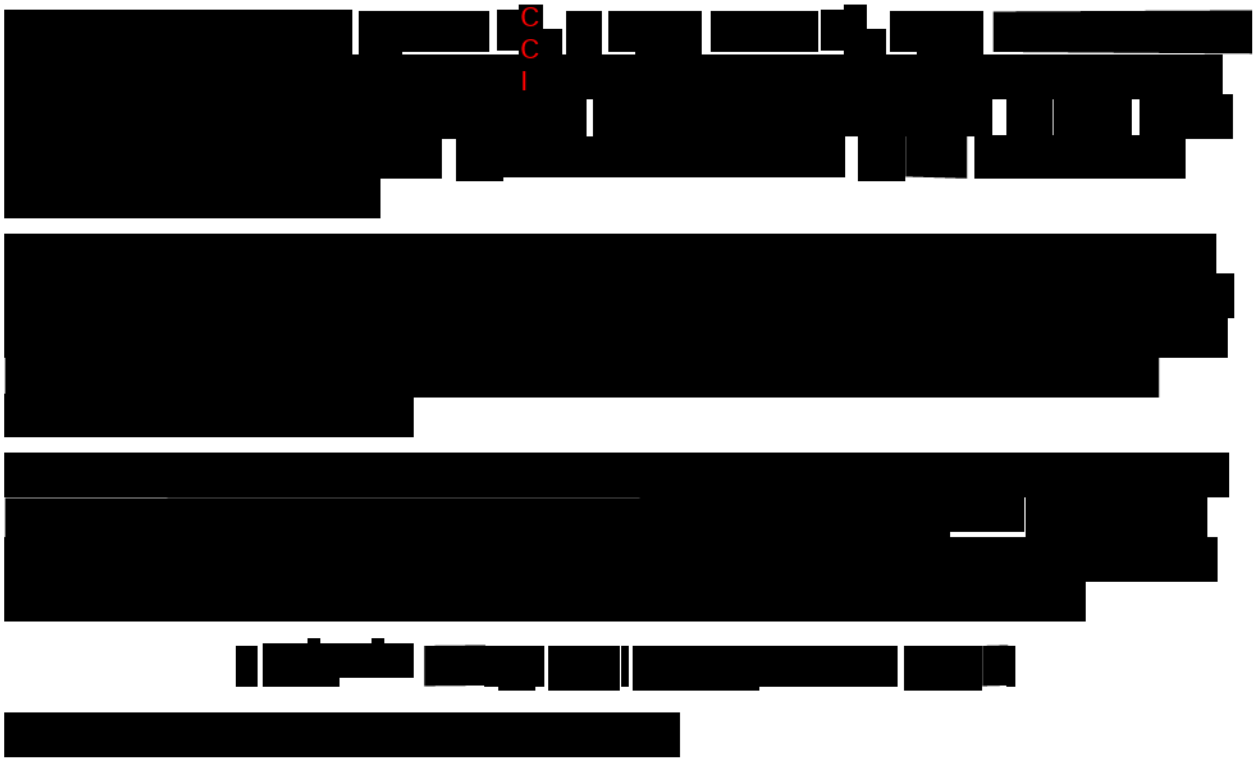
Winthrop KL, Novosad SA, Baddley JW, Calabrese L, Chiller T, Polgreen P, Bartalesi F, Lipman M, Mariette X, Lortholary O, Weinblatt ME, Saag M, Smolen J. Opportunistic infections and biologic therapies in immune-mediated inflammatory diseases: consensus recommendations for infection reporting during clinical trials and postmarketing surveillance. *Ann Rheum Dis*. 2015;74(12):2107-2116.

# 9. Appendices

# Appendix 1. Bayesian Analysis Methodology

CCI





# Appendix 2. Definition of Medications for Tertiary Censoring Rule

| Medication type | ATC Code or Standardized Medication Name | Routes |
|---|---|---|
| Systemic | ATC level 3 code in "G01B" "H02A" "H02B" "L04A"<br>OR<br>ATC level 4 code in "A01AC" "A07EA" "C05AA" "M01BA" "R03BA" "L03AB" "L03AC"<br>OR<br>Standardized Medication Name in "TRALOKINUMAB" "OMALIZUMAB" "USTEKINUMAB" "DUPILUMAB" | INTRALMUSCULAR<br>INTRAVENOUS<br>INTRAVENOUS CENTRAL VEIN<br>INTRAVENOUS PERIPHERAL VEIN<br>ORAL<br>SUBCUTANEOUS<br>INTRA-ARTERIAL<br>INTRACORONARY<br>EPIDURAL<br>INTRA-ARTICULAR (allowed on limited basis)<br>INTRAPERICARDIAL<br>INTRAPERITONEAL<br>INTRAPLEURAL<br>INRTATHECAL<br>INTRATRACHEAL<br>INTRAVESICAL<br>RECTAL<br>SUBLINGUAL<br>URETHRAL<br>VAGINAL |
| Topical | ATC level 4 code in "D07AC" "D07BC" "D07CC" "D07XC" "D07AD" "D07BD" "D07CD" "D07XD" | INTRADERMAL<br>TOPICAL<br>TRANSDERMAL |

Leo Document ID = 90cee1ec-9473-4454-8d19-5f0e36685993

Approver: PPD

Approval Date & Time: 01-Jun-2022 13:48:55 GMT

Signature meaning: Approved