Statisitcal Analyses Plan (d) I4V-MC-JAIN

A Phase 3, Multicenter, Double-Blind, Randomized, Placebo-Controlled Study Evaluating the Safety and Efficacy of Baricitinib in Combination with Topical Corticosteroids in Adult Patients with Moderate-to-Severe atopic Dermatitis Who Have Experienced Failure to Cyclosporine or Are Intolerant to, or Have Contradiction to Cyclosporine

NCT03428100

Approval Date: 19-Dec-2019

1. Statistical Analysis Plan: I4V-MC-JAIN(d): A Phase 3, Multicenter, Double-Blind, Randomized, Placebo-Controlled Study Evaluating the Safety and Efficacy of Baricitinib in Combination with Topical Corticosteroids in Adult Patients with Moderate-to-Severe Atopic Dermatitis Who Have Experienced Failure to Cyclosporine or Are Intolerant to, or Have Contraindication to, Cyclosporine (BREEZE-AD4)

Baricitinib (LY3009104) Atopic Dermatitis

Study I4V-MC-JAIN is a Phase 3, multicenter, double-blind, randomized, 113-week placebo-controlled study evaluating the safety and efficacy of baricitinib 1-mg once daily, 2-mg once daily, and 4 mg once daily compared with placebo in patients with moderate-to-severe atopic dermatitis who are receiving background topical corticosteroid treatment and who have experienced failure to cyclosporine or are intolerant to, or have a contraindication to, cyclosporine.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol I4V-MC-JAIN(d)
Phase 3

Statistical Analysis Plan electronically signed and approved by Lilly on date provided below.

Approval Date: 19-Dec-2019 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

3. Revision History

Statistical Analysis Plan (SAP) Version 1 is based on Protocol I4V-MC-JAIN(b) and was approved prior to the first unblinding.

Statistical Analysis Plan Version 2 is based on Protocol I4V-MC-JAIN(d) and Program Statistical Analysis Plan (PSAP) Version 6. The purpose of SAP Version 2 is to describe and document the 24-week interim analysis. The analysis for assessment after Week 24, Treatment Period 3, including the down-titration substudy of JAIN will be described in a subsequent version of this SAP.

The changes incorporated in Version 2 are as follows:

Change	Section	Summary of Change
Study Objective	Section 4	• Updated primary objective to include estimand language
Primary Endpoint	Section 4.1	• Changed the primary endpoint from IGA of 0 or 1 at Week 16 to EASI75 at Week 16
Key Secondary Endpoints	Section 4.2	• Made IGA of 0 or 1 at Week 16 a key secondary endpoint • Made IGA of 0 or 1 and EASI75 at Week 24 key secondary endpoints
Other Secondary Endpoints	Section 4.2.2	• Moved IGA of 0 or 1 and EASI75 at Week 52 from key secondary endpoints to other secondary endpoints
Exploratory Analysis	Section 4.3	• Added details to the analysis and exploratory endpoints descriptions for POEM, DLQI, Itch NRS, ADSS, Skin Pain NRS, and HADS • Added IGA (0.1), EASI75, and 4-point improvement in Itch NRS to time to event analysis
Assumption in Sample Size Calculation	Section 6.1	• Updated the sample size assumption to align with I4V-MC-JAIN(d) protocol
Analysis Period	Section 6.2	• Updated the efficacy analysis period from Week 0 to Week 16 to Week 0 to Week 24 • Updated the safety analysis period

Change	Section	Summary of Change
Duplicated Diary Records	Section 6.2.2	• Updated the method for handling duplicated diary records
Derived Data	Section 6.2.3	• Added details to derived data
Missing Data Imputation and Sensitivity Analysis	Section 6.3	• Added mBOCF method to key secondary continuous endpoints
Placebo Multiple Imputation	Section 6.3.4	• Updated analysis description • Updated Table JAIN.6.4
Tipping Point	Section 6.3.5	• Updated Table JAIN.6.5
Graphical Testing Scheme	Section 6.6	• Updated the graphical testing procedure
Compliance	Section 6.7	• Updated the detail to support 24-week interim analysis
Background Therapy	Section 6.8	• Updated analysis method to align with Study 14V-MC-JAIY
Concomitant Medication	Section 6.9	• Updated analysis method to align with Study 14V-MC-JAIY
Efficacy Analysis	Section 6.10	• Updated Table JAIN.6.6 • Updated Table JAIN.6.7 • Updated primary outcome

Abbreviations: ADSS = Atopic Dermatitis Sleep Scale; DLQI = Dermatology Life Quality Index; EASI75 = 75% improvement from baseline in Eczema Area and Severity Index score; HADS = Hospital Anxiety Depression Scale; IGA = Investigator's Global Assessment for AD; mBOCF = modified baseline observation carried forward; NRS = Numeric Rating Scale; POEM = Patient-Oriented Eczema Measure.

4. Study Objectives

4.1. Primary Objective

The primary objective of this study is to test the hypothesis that baricitinib 4-mg once daily (QD) plus topical corticosteroids (TCS) or baricitinib 2-mg QD plus TCS is superior to placebo plus TCS in the treatment of patients with moderate-to-severe atopic dermatitis (AD), as assessed by the proportion of patients achieving 75% improvement from baseline using the Eczema Area and Severity Index score (EASI75) at Week 16.

In particular, the associated estimand for this objective is to measure the effect of therapy with baricitinib as assessed by the proportion of patients with a response of EASI75 at Week 16 assuming treatment response disappears after patients are rescued or discontinued from the study or treatment. See Sections 6.3 and 6.10 for details on how this estimand handles outcomes after the occurrence of any intercurrent event through nonresponder imputation (NRI).

4.2. Secondary Objectives

4.2.1. Key Secondary Objectives

Key Secondary *These are prespecified objectives that will be adjusted for multiplicity*	
Objectives	**Endpoints**
To test the hypothesis that baricitinib 1-mg + TCS is superior to placebo + TCS in the treatment of patients with moderate-to-severe AD	• Proportion of patients achieving EASI75 at 16 weeks
To compare the efficacy of baricitinib 4-mg + TCS, baricitinib 2-mg + TCS, or baricitinib 1-mg + TCS to placebo + TCS in AD during the double-blind placebo-controlled treatment period as measured by improvement of signs and symptoms of AD	• Proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement at 16 weeks • Proportion of patients achieving EASI90 at 16 weeks • Percent change from baseline in EASI score at 16 weeks • Proportion of patients achieving SCORAD75 at 16 weeks
To compare the efficacy of baricitinib 4-mg + TCS, baricitinib 2-mg + TCS, or baricitinib 1-mg + TCS to placebo + TCS in AD during the double-blind placebo-controlled treatment period as assessed by patient-reported outcome measures	• Proportion of patients achieving a 4-point improvement in Itch NRS at 16 weeks, 4 weeks, 2 weeks, and 1 week • Mean change from baseline in the score of Item 2 of the ADSS at 16 weeks and 1 week • Mean change from baseline in Skin Pain NRS at 16 weeks
To compare the efficacy of baricitinib 4-mg + TCS, baricitinib 2-mg + TCS, or baricitinib 1-mg + TCS to placebo + TCS in AD during the double-blind placebo-controlled treatment period as measured by improvement of signs and symptoms of AD	• Proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at 24 weeks • Proportion of patients achieving EASI75 at 24 weeks

Key Secondary Objectives

Abbreviations: AD = atopic dermatitis; ADSS = Atopic Dermatitis Sleep Scale; EASI75 = 75% improvement from baseline in Eczema Area and Severity Index score; EASI90; 90% improvement from baseline in Eczema Area and Severity Index score; IGA = Investigator's Global Assessment for AD; NRS = Numeric Rating Scale; SCORAD75 = 75% decrease from baseline in Scoring Atopic Dermatitis values; TCS = topical corticosteroids.

4.2.2. Other Secondary Objectives

Other Secondary *These are prespecified objectives that will not be adjusted for multiplicity*	
Objectives	**Endpoints**
To test the hypothesis that baricitinib 4-mg + TCS, baricitinib 2-mg + TCS, or baricitinib 1-mg + TCS is superior to placebo + TCS in the treatment of moderate-to-severe AD	• Proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement at Week 4 and Week 52 • Proportion of patients achieving EASI75 at Week 4 and Week 52
To compare the efficacy of baricitinib 4-mg + TCS, baricitinib 2-mg + TCS, or baricitinib 1-mg + TCS to placebo + TCS in AD during the double-blind placebo-controlled treatment period as measured by improvement in signs and symptoms of AD	• Proportion of patients achieving EASI50 at 16 weeks • Proportion of patients achieving IGA of 0 at 16 weeks • Mean change from baseline in SCORAD at 16 weeks • Proportion of patients achieving SCORAD90 at 16 weeks • Mean change from baseline in BSA affected at 16 weeks • Proportion of patients developing skin infections requiring antibiotic treatment by Week 16 • Mean number of days without use of background TCS over 16 weeks • Mean gram quantity of background TCS used over 16 weeks (tube weights)
To compare the efficacy of baricitinib 4-mg + TCS, baricitinib 2-mg + TCS, or baricitinib 1-mg + TCS to placebo + TCS in AD during the double-blind placebo-controlled treatment period as assessed by patient-reported outcome measures	• Percent change from baseline in Itch NRS at Week 52, Week 24, Week 16, Week 4, and Week 1 • Proportion of patients achieving a 4-point improvement in Itch NRS at 24 weeks • Mean change from baseline in the total score of the POEM at 16 weeks • Mean change from baseline in the PGI-S-AD scores at 16 weeks • Mean change from baseline in HADS total scores at 16 weeks • Mean change from baseline in the DLQI total scores at 16 weeks • Mean change from baseline in the WPAI-AD total scores at 16 weeks • Mean change from baseline in the EQ-5D-5L total scores at 16 weeks

Other Secondary Objectives

Abbreviations: AD = atopic dermatitis; ADSS = Atopic Dermatitis Sleep Scale; BSA = body surface area; DLQI = Dermatology Life Quality Index; EASI75 = 75% improvement from baseline in Eczema Area and Severity Index score; EQ-5D-5L = the European Quality of Life–5 Dimensions 5 Levels; HADS = Hospital Anxiety Depression Scale; IGA = Investigator's Global Assessment for AD; NRS = Numeric Rating Scale; PGI-S-AD = Patient Global Impression of Severity–Atopic Dermatitis; POEM = Patient-Oriented Eczema Measure; SCORAD90 = 90% decrease from baseline in Scoring Atopic Dermatitis values; WPAI-AD = Work Productivity and Activity Impairment-Atopic Dermatitis; TCS = topical corticosteroids.

4.3. Exploratory Objectives

The exploratory objectives of this study are as follows:

Objectives/Endpoints
Exploratory Endpoints may include evaluating the response to baricitinib treatment regimens on clinical measures and patient-reported outcomes. These endpoints may include dichotomous endpoints or change from baseline for the following measures: IGA, EASI, SCORAD, POEM, DLQI, WPAI-AD, EQ-5D-5L, Itch NRS, ADSS Item 1, 2, and 3 scores, Skin Pain NRS, SF-36, and PGI-S-AD. Patients continuing on placebo as responders will be assessed during the long-term extension for relevant efficacy endpoints. EASI at Week 16: • Proportion of patients achieving EASI ≤7 • Change from baseline in EASI total score SCORAD at Week 16: • Proportion of patients achieving SCORAD ≤25 (in the subset of patients with SCORAD >25 at baseline) • Percent change from baseline in SCORAD To evaluate the effect of maintaining baricitinib effect on key clinical measures: • Proportion of IGA (0,1) responders at Week 24 for those patients with an IGA (0,1) response at Week 16 • Proportion of EASI75 responders at Week 24 for those patients with an EASI75 response at Week 24 Time to event analysis: • Time to first EASI75 reduction response • Time to first IGA (0,1) response • Time to first Itch NRS 4-point improvement response ADSS at Week 16: • Mean change from baseline in ADSS Item 1 score at Week 1 and Week 16 • Mean change from baseline in ADSS Item 3 ADSS at Week 1 and Week 16 • Proportion of patients achieving a ≥1-point improvement in ADSS Item 1 score for those with a baseline Item 1 score ≥1 • Proportion of patients achieving a ≥2-point improvement in ADSS Item 2 score for those with a baseline Item 2 score ≥2 • Proportion of patients achieving a 1-point improvement in ADSS Item 3 score for those with a baseline Item 3 score ≥1 • Proportion of patients achieving a ≥1.5-point improvement in ADSS Item 2 score for those with a baseline Item 2 score ≥1.5 Skin Pain at Week 16: • Proportion of patients achieving a ≥4-point improvement for those with a baseline Skin Pain NRS score ≥4 DLQI at Week 16: • Proportion of patients achieving a ≥4-point improvement in DLQI total score for those with a baseline DLQI total score ≥4

- Proportion of patients achieving a DLQI total score of 5 or less for those with a baseline DLQI total score >5
- Proportion of patients achieving DLQI total score of 0 or 1 for those with a baseline DLQI score >1

POEM at Week 16:

- Proportion of patients achieving a ≥4-point improvement in POEM total score for those with a baseline total score ≥4
- Proportion of patients achieving a ≥3.4-point improvement in POEM total score for those with a baseline total score ≥3.4

HADS:

- Proportion of patients achieving a HADS Anxiety Score <8 for those with a baseline HADS Anxiety Score ≥8
- Proportion of patients achieving a HADS Depression Score <8 for those with a baseline HADS Depression Score ≥8
- Proportion of patients achieving improvement in HADS Anxiety Score or HADS Depression Score <8 for those with a baseline HADS Anxiety Score ≥8 or a baseline HADS Depression Score ≥8
- Change from baseline in HADS total score

Abbreviations: ADSS = Atopic Dermatitis Sleep Scale; DLQI = Dermatology Life Quality Index; EASI = Eczema Area and Severity Index; EASI75 = 75% improvement from baseline in Eczema Area and Severity Index score; EQ-5D-5L = the European Quality of Life–5 Dimensions 5 Levels; HADS = Hospital Anxiety Depression Scale; IGA = Investigator's Global Assessment for AD; NRS = Numeric Rating Scale; PGI-S-AD = Patient Global Impression of Severity–Atopic Dermatitis; POEM = Patient-Oriented Eczema Measure; SF-36 = Medical Outcomes Study 36-item short-form health survey; SCORAD = Scoring Atopic Dermatitis; WPAI-AD = Work Productivity and Activity Impairment-Atopic Dermatitis.

5. Study Design

Study I4V-MC-JAIN (JAIN) is a Phase 3, multicenter, double-blind, randomized, placebo-controlled study evaluating the safety and efficacy of baricitinib 1-mg QD, 2-mg QD, and 4-mg QD compared with placebo in patients with moderate-to-severe AD who are receiving background topical TCS treatment and who have experienced failure to cyclosporine or are intolerant to, or have a contraindication to, cyclosporine.

The study consists of 4 periods:

Period 1: **Screening Period (Visit 1)**: Up to 5 weeks prior to randomization

Period 2: **52-week Double-Blind Treatment Period**: From Week 0 (Baseline; Visit 2) up to Week 52 (Visit 14)

Period 3: **52-week Double-Blind, Long-Term Extension Period**: From Week 52 (Visit 14) to Week 104 (Visit 22)

a. Randomized Down-Titration Substudy (Week 52)
Responders (Investigator's Global Assessment for AD [IGA] 0 or 1) and Partial responders (IGA 2) at Week 52 who are eligible to enter the substudy will be rerandomized as follows:
   - baricitinib 4-mg treatment group 1:1 to baricitinib 2-mg, or baricitinib 4-mg.
   - baricitinib 2-mg treatment group 1:1 to baricitinib 1-mg, or baricitinib 2-mg.

b. Responders (IGA 0 or 1) and Partial responders (IGA 2) at Week 52 who are <u>not</u> eligible to enter the randomized down-titration substudy will continue on the treatment regimen assigned at baseline. However, patients in the placebo, baricitinib 2-mg, and baricitinib 1-mg treatment groups will be automatically rerandomized to either 4-mg or 2-mg baricitinib (1:1) in case of a worsening of symptoms during Period 3, such that IGA increases to ≥3. Patients in the baricitinib 4-mg group will remain on 4-mg.

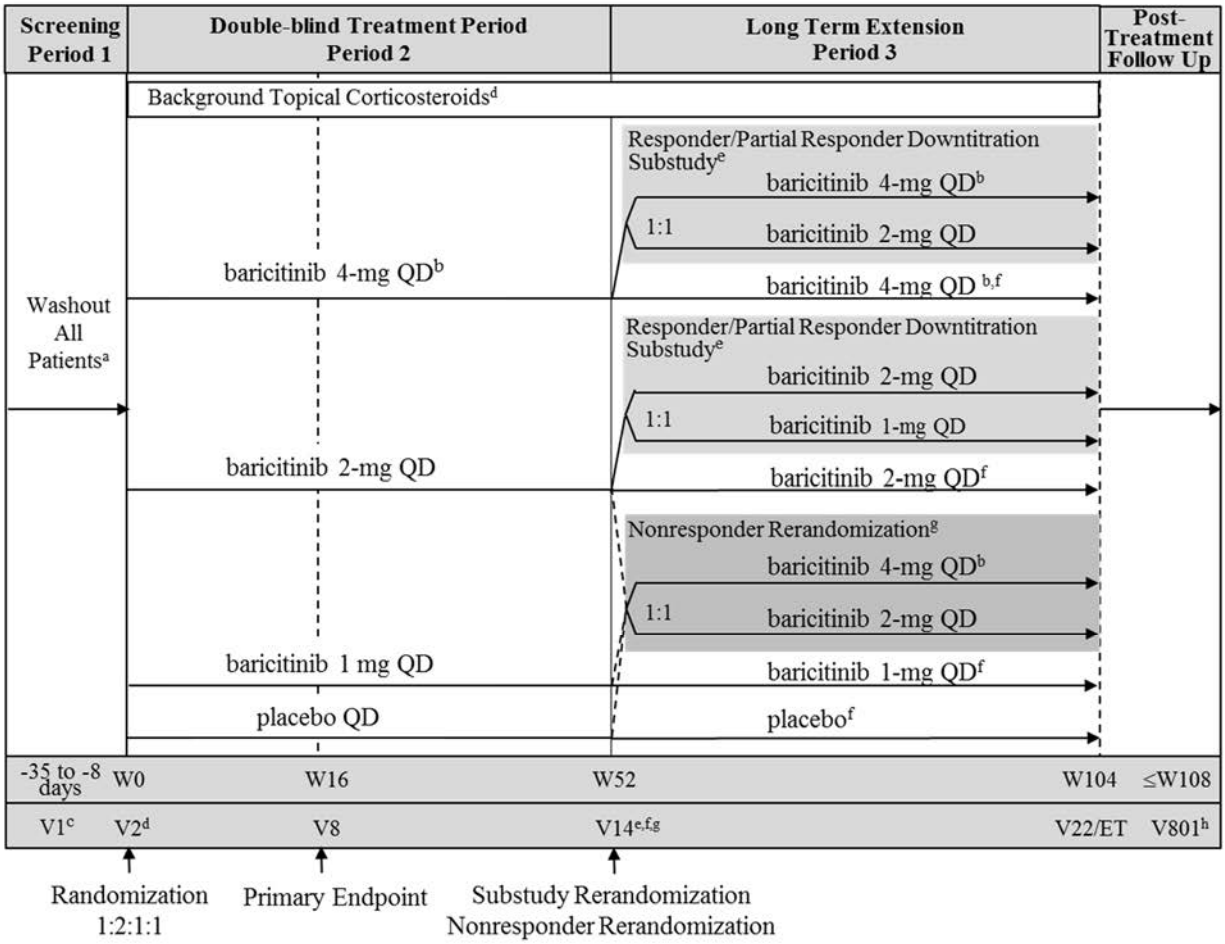
c. Nonresponders (IGA 3 or 4) at Week 52 in the placebo, baricitinib 2-mg, or baricitinib 1-mg treatment groups will be rerandomized at a 1:1 ratio to baricitinib 4-mg or baricitinib 2-mg QD at Week 52. After re-randomization, patients will remain on the same dose of baricitinib for the remainder of the study. Nonresponders who were randomized to baricitinib 4-mg will remain on 4-mg.

Period 4: **Post-Treatment Follow-Up Period**: From last treatment period visit or early termination visit (ETV) to 4 weeks after the last dose of investigational product.

Treatment Arms and Duration: At baseline, patients will be randomized at a 1:1:2:1 ratio to receive placebo QD, baricitinib 1-mg QD, baricitinib 2-mg QD, or baricitinib 4-mg QD for up to 104 weeks.

Number of Patients: Planned enrollment is 500 patients ≥18 years of age.

Figure JAIN.5.1 illustrates the study design.



Abbreviations: AD = atopic dermatitis; eGFR = estimated glomerular filtration rate; ET = early termination; IGA = Investigator's Global Assessment for AD; IP = investigational product; JAIN = I4V-MC-JAIN; PPD = purified protein derivative; QD = once daily; TB = tuberculosis; TCS = topical corticosteroids; V = visit; W = week.

[a] Applicable to patients taking topical treatments (excluding emollients) or systemic treatments for AD at the time of screening.

[b] Maximum dose of baricitinib for patients with renal impairment (defined as eGFR <60 mL/min/1.73 m^2) will be 2-mg QD.

[c] Patients for whom PPD skin test for the evaluation of TB infection was performed at V1 must return and PPD test must be read 48 to 72 hours after Visit 1 (post-PPD).

[d] At Visit 2 (W0), patients will be supplied with mild- and moderate-potency TCS to be applied per the guidelines in JAIN Protocol Section 7.7.2.

[e] At Week 52, responders (IGA 0 or 1) and partial responders (IGA 2) who were assigned to baricitinib 4-mg or 2-mg at randomization, are currently receiving investigational product (does not currently have study drug interrupted), and have not used high- or ultra-high-potency TCS in the previous 14 days will be enrolled into the down-titration substudy. If a patient in the substudy has an IGA ≥3, the patient will be re-treated automatically with the pre-substudy baricitinib dose for the remainder of the study.

[f] At Week 52, responders (IGA 0 or 1) and partial responders (IGA 2) in the baricitinib 4-mg or baricitinib 2-mg treatment groups who are not eligible for the randomized down-titration substudy and those who are in the baricitinib 1-mg or placebo group will remain on their current dose of investigational product. If worsening of AD symptoms occurs any time thereafter such that a patient's IGA is ≥3, with the exception of patients in the baricitinib 4-mg group, they will be re-randomized automatically at a 1:1 ratio to baricitinib 2-mg QD or baricitinib 4-mg QD. Re-randomization will only occur once. Patients in the baricitinib 4-mg group will remain on 4-mg.

[g] Beginning at Visit 14 (Week 52), nonresponders (IGA ≥3) in the placebo, baricitinib 1-mg, or baricitinib 2-mg treatment group will be rerandomized at a 1:1 ratio to baricitinib 4-mg or baricitinib 2-mg QD. Nonresponders randomized to baricitinib 4-mg at baseline will remain on 4-mg. After re-randomization, patients will remain on the same dose of baricitinib for the remainder of the study.

[h] Occurs approximately 28 days after the last dose of IP.

Figure JAIN.5.1. Illustration of study design for Clinical Protocol I4V-MC-JAIN.

5.1. Method of Assignment to Treatment

At Week 0 (Visit 2), patients who meet all criteria for enrollment will be randomized in a 1:1:2:1 ratio to receive double-blind treatment with placebo QD, baricitinib 1-mg QD, 2-mg QD, or 4-mg QD. Randomization will be stratified by geographic region and disease severity at baseline (IGA 3 vs 4). 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 cartons, each containing 4 blister packs of double-blind investigational product tablets to each patient, starting at Visit 2 (Week 0), and at each visit up to and including Visit 22 (Week 104). Site personnel will confirm that they have located the correct carton by entering a confirmation number found on the carton into the IWRS.

This study will be conducted internationally at multiple sites.

describes how regions will be defined for stratification. Regions may be combined for statistical analyses when one of the region strata fails to meet the minimum number of patients. Details are provided in Section 6.2.2.2.

Table JAIN.5.1. Geographic Regions for Stratification

Region	Country
Europe (EU)	Austria, Belgium, Finland, France, Germany, Italy, Netherlands, Poland, Spain, Switzerland, UK
Japan (JP)	Japan
Rest of World (ROW)	Brazil, Russia

In Period 3 (Week 52 [Visit 14] to Week 104 [Visit 22]), eligible patients may be re-randomized depending on their response to treatment. Re-randomization for the Down-Titration Substudy (at Week 52 [Visit 14]) is stratified by IGA (0 or 1 vs 2).

6. A Priori Statistical Methods

6.1. Determination of Sample Size

Study JAIN will aim to enroll 500 patients ≥18 years of age. Ignoring stratification, the proposed sample size will ensure >90% power based on the Chi-square test to detect an absolute difference of approximately 25% between the 4-mg baricitinib and the placebo treatment group and the 2-mg baricitinib and placebo treatment group, each using a 2-sided alpha of 0.025 and assuming approximately 20% placebo response rate for the primary endpoint. These assumptions are based on what was observed in Phase 3 Study 14V-MC-JAIY.

Sample size and power estimates were obtained from nQuery® Advisor 7.0.

6.2. General Considerations

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

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly). The 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. Any display described in this SAP and not included in the CSR are available upon request.

Statistical tests of treatment effects and confidence intervals (CIs) will be performed at a 2-sided significance level of 0.05, unless otherwise stated (for example, graphical multiple testing strategy in Section 6.4).

Data collected at ETVs will be mapped to the closest scheduled visit number for that patient if it falls within the visit window discussed in Section 6.2.2. For by-visit summaries, only visits in which a measure was scheduled to be collected will be summarized.

Any unscheduled visit data will be included in the patient-level listings. However, the data will still 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.

Although Period 3 goes through Week 104 (Visit 22), daily diary collection only continues through Week 68 (16 weeks after Week 52; Visit 18).

6.2.1. Analysis Populations

The following major analysis populations will be used in SAP Version 2 for the 24-week interim analysis (additional analysis populations for specific analysis will be defined in the corresponding analysis section):

Intent-to-treat (ITT) population: The ITT population analysis set is defined as all randomized patients for Period 2.

Per-protocol Set (PPS): The PPS of the ITT population analysis set will include those patients who are not deemed noncompliant with treatment in Period 2 (up to Week 52), who do not have significant protocol violations, and whose investigator site does not have significant good clinical practice (GCP) issues that require a report to the regulatory agencies prior to Week 16 (Visit 8). Qualifications for and identification of significant protocol deviations will be determined while the study remains blinded, prior to database lock. Details can be found in Section 6.15.

Unless otherwise specified, the efficacy and health outcome analyses for Period 2 will be conducted on the ITT population (Gillings and Koch 1991), which seeks to preserve the benefits of randomization and avoids the issue of selection bias. Patients will be analyzed according to the treatment to which they were assigned, even if the patient does not receive the correct treatment, or otherwise did not follow the protocol. In addition, the primary and key secondary analyses will be repeated using the PPS population.

Safety population: The safety population is defined as all randomized patients who receive at least 1 dose of investigational product 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 rates are not underestimated.

Safety analyses will be done using the safety population. Patients will be analyzed according to the treatment 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 exist (for example, unscheduled laboratory assessments) after first dose of study drug, a listing of the data or a patient profile will be provided when requested.

The time period for safety analysis of the 24-week interim analysis will be following criteria whichever occurs first:

- up to last dose plus up to 30 days in the study
- the first dose in Period 3
- up to dose chance or up to the data cutoff date

Table JAIN.6.1. Safety Population Treatment Group Analyses for 24-Week Interim Analysis

Treatment Group	Definition
PBO	Placebo + TCS at entry to Study JAIN followed up to end of study or treatment change (data censored at treatment change)
BARI 1-mg	BARI 1-mg + TCS at entry to Study JAIN followed up to end of study or dose change (data censored at dose change)
BARI 2-mg	BARI 2-mg + TCS at entry to Study JAIN followed up to dose change (data censored at dose change)
BARI 4-mg	BARI 4-mg + TCS at entry to Study JAIN followed up to dose change (data censored at dose change)

Abbreviations: BARI = baricitinib; JAIN = I4V-MC-JAIN; PBO = placebo; TCS = topical corticosteroids.

Table JAIN.6.2 describes the treatment groups and the comparisons for each study period and the analysis population.

Table JAIN.6.2. Treatment Groups and Comparisons for the 24-Week Interim Analysis and Analysis Population

Study Period	Analysis Population	Treatment Group	Abbreviation	Comparison
52-week Double-Blind Treatment Period (Period 2)	Intent-to-Treat Population Per-protocol Set Safety Population	Placebo + TCS	PBO	BARI-4mg vs PBO BARI-2mg vs PBO BARI-1mg vs PBO
		Baricitinib 4-mg + TCS	BARI-4mg	
		Baricitinib 2-mg + TCS	BARI-2mg	
		Baricitinib 1-mg + TCS	BARI-1mg	
		Pooled Baricitinib + TCS (for safety only)	BARI 1-mg/2-mg/4-mg	
		Total	Total	

Abbreviations: BARI = Baricitinib; TCS = topical corticosteroids; vs. = versus.

6.2.2. General Considerations for Analyses for the 24-Week Interim Analysis

Period 2 starts at randomization (Visit 2, Week 0) and ends with the visit at Week 52 (Visit 14) or the early discontinuation visit (between Weeks 0 and 52).

24-week interim analysis will be done when all the patients complete Week 24 assessment (Visit 10) or discontinue study or treatment before the Week 24 assessment.

6.2.2.1. Definition of Baseline and Postbaseline Measures

The baseline value for the efficacy and health outcomes 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 (Itch Numeric Rating Scale [NRS], Atopic Dermatitis Sleep Scale (ADSS), Skin Pain NRS, Patient Global Impression of Severity–Atopic Dermatitis [PGI-S-AD]) is defined as the mean assessment of the 7 days prior to the date of first study drug administration. Criteria for derivation of the baseline score requires that 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 to obtain the most recent 4 non-missing measurements prior to first study drug administration. If there at least 4 non-missing assessments are not collected prior to the date of first study drug administration using the aforementioned method, the baseline will be designated as missing.

Baseline for the safety analyses is defined as the last non-missing scheduled (planned) measurement on or prior to the date of first study drug administration for continuous measures, by-visit analyses, and all non-missing measurements on or prior to the date of first study drug administration for all other analyses.

Postbaseline measurements are collected after study drug administration through Week 52 (Visit 14) or early discontinuation visit. Data collected on scheduled visits will generally be used for efficacy and health outcome analyses. If data for a scheduled visit are missing, data from the most proximal unscheduled visits, if available, will be used if it falls within visit windows as follows: a 2-day interval around the scheduled visit day (4-day window) is applied to Visits 3 to 6 (Week 1, 2, 4, 8), a 4-day interval around the scheduled visit day (8-day window) is applied to Visits 7 to 9 (Week 12, 16, 20), and a 5-day interval around the scheduled visit day (10-day window) is applied to Visits 10 to 14 (Week 24, 32, 40, 48, 52). 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 only 2 equal distance unscheduled visits are available, the later of the 2 visits will be used.

Postbaseline visits for weekly assessments of daily diary data will be each 7-day visit interval beginning on the day of first study drug administration, creating intervals for Week 1 (Days 1 through 7), Week 2 (Days 8 through 14), Week 3 (Days 15 through 21), etc. Each interval will be represented by the mean of that interval. If there are less than 4 non-missing

measurements in any weekly interval, the mean for that weekly visit interval will be considered missing. Furthermore, as some analyses require use of the primary censoring rule (Section 6.3), assessments collected on or after the day of rescue will be excluded from the weekly visit interval calculation when implementing this censoring rule. If, after exclusion of those records, there are less than 4 non-missing assessments, the mean of the weekly interval that implements the primary censoring rule will be considered missing.

Postbaseline measures for the safety analyses are defined as the non-missing scheduled (planned) measurements after the date of first study drug administration for continuous measures, by-visit analyses, and all non-missing measurements after the date of first study drug administration for all other analyses.

Handling of Duplicate Diary Records

If there is more than 1 diary record on a particular date, the first record on that date will be used in the analysis.

As some analyses require use of the primary censoring rule, assessments collected on or after the day of rescue will be excluded from the weekly visit interval calculation when implementing the rule for the daily diary. If, after exclusion of these records, there are less than 4 non-missing assessments, the weekly interval that implements the primary censoring rule will be considered missing. The poststudy follow-up weekly score for daily diaries will be calculated as the mean of the 7 days prior to the follow-up visit that occur after last dose of study treatment.

Postbaseline measures for the safety analyses are defined as the non-missing scheduled (planned) measurements after the date of first study drug administration for continuous measures, by-visit analyses, and all non-missing measurements after the date of first study drug administration for all other analyses.

6.2.2.2. Covariate Adjustment

This study will be conducted by multiple investigators at multiple sites internationally, and countries will be categorized into geographic regions (see Table JAIN.5.1). Randomization to treatment groups at Week 0 (Visit 2) is stratified by disease severity (IGA) and geographic region as described in Section 5.1. Unless otherwise specified, the statistical analysis models will adjust for disease severity and geographical region.

The covariates used in the logistic model for categorical data will include the parameter value at baseline. The covariates used in the analysis of covariance (ANCOVA) model for continuous data will include the parameter value at baseline. Inclusion of baseline in the model ensures that treatment least-squares means (LSMs) are estimated at the same baseline value. When a mixed-model repeated measures (MMRM) analysis is performed, baseline value and baseline-by-visit interactions will be included as covariates.

Unless otherwise specified, all analyses described in this section will compare estimates (for example, odds ratios, LSMs, proportions) of baricitinib 4-mg, 2-mg, and 1-mg to placebo. Thus, odds ratios are for baricitinib treatment groups relative to placebo; similarly, LSM differences and differences in proportions are between baricitinib treatment groups and placebo.

6.2.2.3. Analysis Method

The main analysis of categorical efficacy variables and health outcomes variables will use a logistic regression analysis with region (as appropriate), baseline disease severity (IGA), baseline value, and treatment group in the model. Firth's correction will be used in 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 difference in percentages and 100(1-alpha)% CI of the difference in percentages using the Newcombe-Wilson without continuity correction will be reported. The p-value from the Fisher's exact test will also be produced as a secondary analysis.

In case geographic region is used as factor in the statistical model, for the analysis of the primary endpoint, treatment-by-region interaction will be added to the logistic regression model as a sensitivity 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 (that is, 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.

The main analysis method for all continuous efficacy and health outcomes variables will use MMRM analysis. The MMRM model will use a restricted maximum likelihood (REML) estimation. The model will include treatment, region (as appropriate), baseline disease severity (IGA), visit, and treatment-by-visit-interaction as fixed categorical effects and baseline and baseline-by-visit-interaction as fixed continuous effects. For daily diary assessments and the model for analyses, the following models will be generated:

- The models include weekly assessment up to Week 16: results from this model will be used in a graphical testing scheme. If the algorithm fails to converge, the scheduled visit weeks will be used: 1, 2, 4, 8, 12, and 16.
- The models include weekly assessment up to Week 24: If the algorithm fails to converge, the scheduled visit weeks will be used: 1, 2, 4, 8, 12, 16, 20, and 24.

For non-daily assessments associated with a continuous endpoint such as EASI percent change from baseline, the model will use scheduled visit as covariate.

An unstructured (co)variance 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 LSM will be estimated within the framework of the MMRM using type 3 sums of squares. Differences in LSM between each dose of baricitinib and placebo (and associated p-values,

standard errors, and 95% 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 and health outcomes variables may also be made using ANCOVA for primary and key secondary objectives. When an ANCOVA model is used, the model includes region, baseline disease severity, treatment group, and baseline value. Inclusion of baseline in the ANCOVA models ensures that treatment LSMs are estimated at the same baseline value. Treatment LSMs 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. Treatment-by-region interaction will also be added to the model for sensitivity purposes and is discussed in Section 6.10.

Time to event will be analyzed using cumulative incidence function with observed values, defining first time reaching the event IGA (0,1), EASI 75, or Itch NRS 4-pt improvement before rescue as onset, treating rescue and discontinuation for lack of efficacy as competing event censor up to Week 16.

Fisher's exact test will be used to test the difference between each baricitinib dose and placebo in proportion 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 and baseline value in the model. The significance of within-treatment group changes from baseline will be evaluated by testing whether 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.

6.2.3. Derived Data

- Age (year), derived using first dose date as the reference start date and July 1 of birth year and truncated to a whole-year (integer) age. Patients whose derived age is less than 18 will have the required minimum age of 18 at informed consent; however, reporting for age, age groups, and laboratory ranges, will be based on the derived age.
- Age group (<65 years, ≥65 years)
- Age group (<65 years, ≥65 years to <75 years, ≥75 years to <85 years, ≥85 years)
- Body mass index (BMI) (kg/m^2) = weight (kg)/((height (cm)/100)^2)
- BMI category (<25 kg/m^2, ≥25 kg/m^2 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. 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 (0 to <2 years, 2 years to < 5 years, 5 years to <10 years, 10 years to <20 years, ≥20 years)

- Diagnosis age (years), derived using diagnosis date as the reference start date and July 1 of birth year and truncated to a whole-integer age
- Diagnosis age group (<18, ≥18 to <50, ≥50 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 (lb) * 0.454.
- Weight category (<60 kg, ≥60 kg to <100 kg, ≥100 kg)
- Height (cm) = height (in) * 2.54.
- Cyclosporine inadequate efficacy (yes, no)
    - Set **yes** if the reason for discontinuation is inadequate response.
- Cyclosporine intolerance (yes, no)
    - Set **yes** if the reason for discontinuation is intolerance to medication or contraindication (physician indicated cyclosporine was used and a contraindication was noted).
- Cyclosporine contraindication [ineligible] (yes, no)
    - Set to **yes** if cyclosporine never used because of a contraindication
- Cyclosporine inadvisable (yes, no)
    - Set to **yes** if the following reasons were selected for either not using the medication or discontinuing the medication:
        - Reason for not using medication: physician decision, concern about side effects, unfavorable benefit risk, contraindication
        - Reason for discontinuation: inadequate response, intolerance to medication, or contraindication
- Topical calcineurin inhibitor (TCNI) inadequate efficacy (yes, no)
    - Set **yes** if the reason for discontinuation is inadequate response.
- TCNI intolerance (yes, no)
    - Set **yes** if the reason for discontinuation is intolerance to medication or contraindication (physician indicated TCNI was used and a contraindication was noted).
- TCNI contraindication /[ineligible](yes, no)
    - Set to **yes** if TCNI never used because of a contraindication
- TCNI inadvisable (yes, no)
    - Set to **yes** if the following reasons were selected for either not using the medication or discontinuing the medication:
        - Reason for not using medication: physician decision, concern about side effects, unfavorable benefit risk, contraindication

- Reasons for discontinuation: inadequate response, intolerance to medication, or contraindication.

6.3. Handling of Dropouts or Missing Data

Intercurrent events (ICH E9 R1) are events which occur after the treatment initiation and make it impossible to measure a variable or influence how it would be interpreted.

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

- application of one of the censoring rules (including after permanent study drug discontinuation, after rescue therapy, or re-treatment)
- discontinuation of inadvertently enrolled patients
- discontinuation from the study due to enrollment in other trials, medical, safety or regulatory reasons, investigator decision, and patient decision
- missing an intermediate visit prior to discontinuation, rescue, or re-treatment
- loss to follow-up

Noncensor intercurrent events are events that are not due to the application of any censoring rule, that is, the last 4 items in the list above.

Note that as efficacy and health outcome 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 and health outcome observations subsequent to these events depending on the estimand being addressed. These specific censoring rules are described below.

The *primary censoring rule* will censor efficacy and health outcome results after permanent study drug discontinuation or after rescue therapy. This censoring rule will be applied to all continuous and categorical efficacy and health outcome endpoints.

A *secondary censoring rule* will only censor efficacy and health outcome results after permanent study drug discontinuation. As a consequence, data for patients rescued with high or ultra-high potency TCS or with phototherapy will not be censored at the time of rescue as they can continue or only temporarily interrupt study drug. As patients who are rescued to systemic corticosteroids are required to permanently discontinue study drug, they will also have post-rescue observations censored. The secondary censoring rule will be applied to primary and key secondary efficacy and health outcome endpoints.

Nonresponder imputation (for categorical variables) and MMRM (for continuous variables) will be the primary methods used to handle missing data. Sections 6.3.1 through 6.3.5 summarize the imputation methods for the various efficacy and health outcome endpoints.

Table JAIN.6.3. Imputation Techniques for Various Variables

Efficacy and Health Outcome Endpoints	Imputation Method
IGA (0,1), EASI75, 4-point Itch NRS improvement	NRI[a,b], pMI[a], Tipping point[a]
EASI90, SCORAD75	NRI[a,b], pMI[a]
EASI percent change, ADSS Item 2 change, Skin Pain NRS change	MMRM[a,b], mLOCF[a], pMI[a], mBOCF[a]
All remaining categorical measures	NRI[a]
All remaining continuous efficacy and health outcome measures	MMRM[a], mLOCF[a]

Abbreviations: ADSS = Atopic Dermatitis Sleep Scale; EASI = Eczema Area and Severity Index score; EASI75 = 75% improvement from baseline in Eczema Area and Severity Index score; EASI90 = 90% improvement from baseline in Eczema Area and Severity Index score; IGA = Investigator's Global Assessment for AD; mBOCF = modified baseline observation carried forward; mLOCF = modified last observation carried forward; MMRM = mixed model repeated measures; NRI = nonresponder imputation, NRS = Numeric Rating Scale; pMI = placebo multiple imputation; SCORAD75 = 75% decrease from baseline in Scoring Atopic Dermatitis values.

[a] Analyses utilizing the primary censoring rule.

[b] Analyses utilizing the secondary censoring rule.

6.3.1. Nonresponder Imputation

An NRI method can be justified based on the composite strategy (ICH E9R1) for handling intercurrent events. This imputation procedure assumes that the effects of treatments disappear after the occurrence of the intercurrent event. It will be the primary imputation method for the analysis of categorical efficacy and health outcomes variables such as IGA (0,1) and EASI50/75/90.

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 6.2.2.2). Additionally, all primary and key secondary categorical endpoints will utilize NRI after applying the secondary censoring rule. For analyses which utilize either of the censoring methods, randomized patients without at least 1 postbaseline observation will also be defined as nonresponders for all visits. As well, patients who are missing a value prior to discontinuation or rescue (if censoring on rescue) (that is, the patient is missing an intermediate visit) will be imputed as nonresponders at that visit only.

6.3.2. Mixed Model for Repeated Measures

For the continuous secondary and exploratory efficacy and health outcome variables, such as EASI score and Scoring Atopic Dermatitis (SCORAD) score, data after the occurrence of intercurrent events (including application of any of the censoring rules) will be set to missing. Mixed model repeated measures analyses will be performed to mitigate the impact of missing data. This approach assumes that missing observations are missing-at-random (missingness is related to observed data) during the study and borrows information from patients in the same treatment arm taking into account both the missingness of data and the correlation of the repeated measurements.

Essentially, this method tries to measure the effect of initially randomized treatments had all patients remained on their randomized 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. A placebo multiple imputation (pMI) or modified

baseline observation carried forward (mBOCF) may be done to bridge the gap between the 2 imputation procedures.

All continuous endpoints will utilize MMRM after applying the primary censoring rule to patients who permanently discontinued study drug or who were rescued (described in Section 6.2.2.2). All key secondary continuous endpoints will also utilize MMRM after applying the secondary censoring rule.

6.3.3. Modified Last Observation Carried Forward

A modified last observation analysis 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 non-missing baseline and at least 1 postbaseline observation will be included in the analyses. These mLOCF analyses help ensure that 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 non-missing 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 non-missing 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.

6.3.4. Placebo Multiple Imputation

The pMI method will be used as an additional analysis when an intercurrent event occurs for the analysis of the primary efficacy endpoint, IGA (0,1) with ≥2-point improvement, as well as the key secondary endpoints at Week 16 (Visit 8) and Week 52 (Visit 14), where applicable. The primary censoring rule will be applied to these sensitivity analyses.

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 are 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 reasonable approach than mLOCF because, unlike mLOCF, it accounts for uncertainty of imputation and therefore does not underestimate standard errors, and it limits bias by taking into account study/placebo effects. In the efficacy context, pMI is a specific form of a missing not at random analysis expected to yield a conservative estimate of efficacy.

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 dataset 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 datasets.

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

The number of imputed datasets 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 JAIN.6.2.

Table JAIN.6.4. 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 and Week 24, with data up to rescue as the primary censoring rule.	123450
Percent change from baseline in EASI score at 16 weeks, with data up to rescue. EASI75 (at 16 weeks and at 24 weeks) and EASI90 (at 16 weeks).	123451
Proportion of patients achieving SCORAD75 at 16 weeks, with data up to rescue	123452
Proportions of patients achieving a 4-point improvement from baseline in Itch NRS at Week 16, with data up to rescue	123453
Mean change from baseline in Skin Pain NRS at Week 16, with data up to rescue	123454
Mean change from baseline in the score of Item 2 of the ADSS at Week 16, with data up to rescue	123455

Abbreviations: ADSS = Atopic Dermatitis Sleep Scale; EASI = Eczema Area and Severity Index score; EASI75 = 75% improvement from baseline in Eczema Area and Severity Index score; EASI90 = 90% improvement from baseline in Eczema Area and Severity Index score IGA = Investigator's Global Assessment for AD; NRS = Numeric Rating Scale; SCORAD75 = 75% decrease from baseline in Scoring Atopic Dermatitis values.

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

6.3.5. Tipping Point Analyses

To investigate the missing data mechanism, additional analyses using multiple imputation (MI) under the missing not at random assumption will be provided for the following primary and key secondary objectives:

- IGA (0,1) with ≥2-point improvement at Week 16, baricitinib 4-mg + TCS or baricitinib 2-mg + TCS compared to placebo + TCS
- EASI percent change from baseline to Week 16, baricitinib 4-mg + TCS or baricitinib 2-mg + TCS compared to placebo + TCS
- Itch NRS 4-point improvement from baseline to Week 16, baricitinib 4-mg + TCS or baricitinib 2-mg + TCS compared to placebo + TCS

All patients in the ITT population will be included in the analyses. Data after the occurrence of intercurrent events (after application of the primary censoring rule) will be set to missing.

Within each analysis, a most extreme case will be considered, in which all missing data for patients randomized to baricitinib 1-mg, 2-mg, or 4-mg will be imputed using the worst possible result and all missing data for patients randomized to placebo will be imputed with the best possible result. Treatment differences will be analyzed using logistic regression or ANCOVA (Section 6.2.2.2) as appropriate.

For continuous variables, the following process will be used to determine the tipping point:

1. To handle intermittent missing visit data, a Markov chain Monte Carlo method (MCMC) (SAS® Proc MI with MCMC option) will be used to create a monotone missing pattern.
2. A set of Bayesian regressions (using SAS® Proc MI with MONOTONE option) will be used for the imputation of monotone dropouts. Starting from the first visit with at least 1 missing value, the regression models will be fit sequentially with treatment as a fixed effect and values from the previous visits as covariates.
3. A delta score is added to all imputed scores at the primary time point for patients in the baricitinib treatment groups, thus worsening the imputed value. The delta score is capped for patients based on the range of the outcome measure being analyzed.
4. Treatment differences between baricitinib and placebo are analyzed for each imputed dataset using ANCOVA (Section 6.2.2.2). Results across the imputed datasets are aggregated using SAS® Proc MI ANALYZE to compute a p-value for the treatment comparisons for the given delta value.
5. Steps 3 and 4 are repeated, and the delta value added to the imputed baricitinib scores is gradually increased. The tipping point is identified as the delta value which leads to a loss of statistical significance (aggregated p-value >0.05) when evaluating baricitinib relative to the placebo group.

As a reference, for each delta value used in Steps 3 through 5, a fixed selection of delta values (ranging from slightly negative to slightly positive) will be added to imputed values in

the placebo group, and Step 4 will be performed for the combination. This will result in a 2-dimensional table, with the columns representing the delta values added to the imputed placebo responses, and the rows representing the delta values added to the imputed baricitinib responses. Separate 2-dimensional tables will compare each baricitinib dose group to placebo.

A similar process will be used for the categorical variables:

1. Missing responses in the baricitinib groups will be imputed with a range of low response probabilities, including probabilities of 0, 0.1, and 0.2. These ranges may be changed after unblinding to provide the display of the observed data.
2. For missing responses in the placebo group, a range of response probabilities (for example, probability = 0, 0.05, … 1) will be used to impute the missing values. Multiple imputed datasets will be generated for each response probability. These ranges may be changed after unblinding to provide the display of the observed data.
3. Treatment differences between baricitinib and placebo are analyzed for each imputed dataset using logistic regression (Section 6.2.2.2). Results across the imputed datasets are aggregated using SAS® Proc MIANALYZE to compute a p-value for the treatment comparisons for the given response probability. If the probability values do not allow for any variation between the multiple imputed datasets (for example, all missing responses in the placebo and baricitinib groups are imputed as responders and nonresponders, respectively), then the p-value from the single imputed dataset will be used.

The tipping point is identified as the response probability value within the placebo group that leads to a loss of statistical significance when evaluating baricitinib relative to placebo.

For tipping point analyses, the number of imputed datasets will be m=100. The seed values to start the pseudorandom number generator of SAS Proc MI (same values for MCMC option and for MONOTONE option) are given in Table JAIN.6.3.

Table JAIN.6.5. Seed Values for Imputation

Analysis	Seed Value
Proportion of patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16 and at Week 24, with data up to rescue	123470
Percent change from baseline in EASI score at 16 weeks, with data up to rescue. EASI75 (at 16 weeks and at 24 weeks).	123471
Proportion of patients achieving a 4-point improvement from baseline in Itch NRS at Week 16, with data up to rescue	123472

Abbreviations: EASI = Eczema Area and Severity Index score; IGA = Investigator's Global Assessment for AD; NRS = Numeric Rating Scale.

6.3.6. Modified Baseline Observation Carried Forward

A baseline observation analysis is performed by carrying forward the baseline assessment for the continuous measure. This assumes that the effect of treatments will be lost and patient status

will return to the baseline after the occurrence of the intercurrent event (after application of the primary censoring rule). The mBOCF analyses will be applied to the ITT population where data from patients with non-missing baseline value will be included in the analyses.

6.4. Multiple Comparisons/Multiplicity

In **Period 2**, the primary and key secondary endpoints will be adjusted for multiplicity to control the overall family-wise Type I error rate at a 2-sided alpha level of 0.05. In **Period 3**, no adjustment for multiplicity will be performed.

For countries whose primary endpoint is IGA 0 or 1, a different graphical testing strategy will be described. The primary and key secondary endpoints will be adjusted for multiplicity to control the overall family-wise Type I error rate at a 2-sided alpha level of 0.05.

The following is a list of primary and key secondary endpoints to be tested.

Primary Null Hypotheses:

- Null Hypothesis[EASI75 W16]: Proportion of baricitinib 4-mg patients achieving EASI75 at Week 16 is equal to the proportion of placebo patients achieving EASI75 at Week 16.
- Null Hypothesis[EASI75 W16]: Proportion of baricitinib 2-mg patients achieving EASI75 at Week 16 is equal to the proportion of placebo patients achieving EASI75 at Week 16.

Key Secondary Null Hypotheses:

Baricitinib 4 mg:

- Null Hypothesis[IGA01 W16]: Proportion of baricitinib 4-mg patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline is equal to the proportion of placebo patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16.
- Null Hypothesis[IGA01 W24]: Proportion of baricitinib 4-mg patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline is equal to the proportion of placebo patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 24.
- Null Hypothesis[EASI75 W24]: Proportion of baricitinib 4-mg patients achieving EASI75 is equal to the proportion of placebo patients achieving EASI75 at Week 24.
- Null Hypothesis[ITCH W16]: Proportion of baricitinib 4-mg patients achieving a 4-point improvement in Itch NRS is equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 16 among patients with baseline Itch NRS score ≥4.
- Null Hypothesis[EASI PCFB]: Percent change from baseline in EASI score for baricitinib 4-mg patients is equal to the percent change from baseline in EASI score for placebo patients at Week 16.
- Null Hypothesis[ITCH W4]: Proportion of baricitinib 4-mg patients achieving a 4-point improvement in Itch NRS is equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 4 among patients with baseline Itch NRS score ≥4.

- Null Hypothesis[SCORAD75]: Proportion of baricitinib 4-mg patients achieving SCORAD75 is equal to the proportion of placebo patients achieving SCORAD75 at Week 16.
- Null Hypothesis[EASI 90]: Proportion of baricitinib 4-mg patients achieving EASI90 is equal to the proportion of placebo patients achieving EASI90 at Week 16.
- Null Hypothesis[PAIN NRS]: Mean change from baseline in Skin Pain NRS for baricitinib 4-mg patients is equal to the mean change from baseline in Skin Pain NRS for placebo patients at Week 16.
- Null Hypothesis[ADSS2 W16]: Mean change from baseline in ADSS Item 2 score for baricitinib 4-mg patients is equal to the mean change from baseline in ADSS Item 2 score for placebo patients at Week 16.
- Null Hypothesis[ITCH W2]: Proportion of baricitinib 4-mg patients achieving a 4-point improvement in Itch NRS is equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 2 among patients with baseline Itch NRS score ≥4.
- Null Hypothesis[ADSS2 W1]: Mean change from baseline ADSS Item 2 the score for baricitinib 4-mg patients is equal to the mean change from baseline in ADSS Item 2 score for placebo patients at Week 1.
- Null Hypothesis[ITCH W1]: Proportion of baricitinib 4-mg patients achieving a 4-point improvement in Itch NRS is equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 1 among patients with baseline Itch NRS score ≥4.

Baricitinib 2 mg:

- Null Hypothesis[IGA01 W16]: Proportion of baricitinib 2-mg patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline is equal to the proportion of placebo patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16.
- Null Hypothesis[IGA01 W24]: Proportion of baricitinib 2-mg patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline is equal to the proportion of placebo patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 24.
- Null Hypothesis[EASI75 W24]: Proportion of baricitinib 2-mg patients achieving EASI75 is equal to the proportion of placebo patients achieving EASI75 at Week 24.
- Null Hypothesis[ITCH W16]: Proportion of baricitinib 2-mg patients achieving a 4-point improvement in Itch NRS is equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 16 among patients with baseline Itch NRS score ≥4.
- Null Hypothesis[EASI PCFB]: Percent change from baseline in EASI score for baricitinib 2-mg patients is equal to the percent change from baseline in EASI score for placebo patients at Week 16.
- Null Hypothesis[ITCH W4]: Proportion of baricitinib 2-mg patients achieving a 4-point improvement in Itch NRS is equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 4 among patients with baseline Itch NRS score ≥4.
- Null Hypothesis[SCORAD75]: Proportion of baricitinib 2-mg patients achieving SCORAD75 is equal to the proportion of placebo patients achieving SCORAD75 at Week 16.

- Null Hypothesis[EASI90 W16]: Proportion of baricitinib 2-mg patients achieving EASI90 is equal to the proportion of placebo patients achieving EASI90 at Week 16.
- Null Hypothesis[PAIN NRS W16]: Mean change from baseline in Skin Pain NRS for baricitinib 2-mg patients is equal to the mean change from baseline in Skin Pain NRS for placebo patients at Week 16.
- Null Hypothesis[ADSS2 W16]: Mean change from baseline in ADSS Item 2 score for baricitinib 2-mg patients is equal to the mean change from baseline in ADSS Item 2 score for placebo patients at Week 16.
- Null Hypothesis[ITCH W2]: Proportion of baricitinib 2-mg patients achieving a 4-point improvement in Itch NRS is equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 2 among patients with baseline Itch NRS score ≥4.
- Null Hypothesis[ADSS2 W1]: Mean change from baseline in ADSS Item 2 score for baricitinib 2-mg patients is equal to the mean change from baseline in ADSS Item 2 score for placebo patients at Week 1.
- Null Hypothesis[ITCH W1]: Proportion of baricitinib 2-mg patients achieving a 4-point improvement in Itch NRS is equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 1 among patients with baseline Itch NRS score ≥4.

Baricitinib 1 mg:

- Null Hypothesis[IGA01 W16]: Proportion of baricitinib 1-mg patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline is equal to the proportion of placebo patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 16.
- Null Hypothesis[IGA01 W24]: Proportion of baricitinib 1-mg patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline is equal to the proportion of placebo patients achieving IGA of 0 or 1 with a ≥2-point improvement from baseline at Week 24.
- Null Hypothesis[EASI75 W16]: Proportion of baricitinib 1-mg patients achieving EASI75 is equal to the proportion of placebo patients achieving EASI75 at Week 16.
- Null Hypothesis[EASI75 W24]: Proportion of baricitinib 1-mg patients achieving EASI75 is equal to the proportion of placebo patients achieving EASI75 at Week 24.
- Null Hypothesis[ITCH W16]: Proportion of baricitinib 1-mg patients achieving a 4-point improvement in Itch NRS is equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 16 among patients with baseline Itch NRS score ≥4.
- Null Hypothesis[EASI PCFB]: Percent change from baseline in EASI score for baricitinib 1-mg patients is equal to the percent change from baseline in EASI score for placebo patients at Week 16.
- Null Hypothesis[ITCH W4]: Proportion of baricitinib 1-mg patients achieving a 4-point improvement in Itch NRS is equal to proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 4 among patients with baseline Itch NRS score ≥4.
- Null Hypothesis[SCORAD75 W16]: Proportion of baricitinib 1-mg patients achieving SCORAD75 is equal to proportion of placebo patients achieving SCORAD75 at Week 16.
- Null Hypothesis[EASI 90 W16]: Proportion of baricitinib 1-mg patients achieving EASI90 is equal to the proportion of placebo patients achieving EASI90 at Week 16.

- Null Hypothesis[PAIN NRS W16]: Mean change from baseline in Skin Pain NRS for baricitinib 1-mg patients is equal to the mean change from baseline in Skin Pain NRS for placebo patients at Week 16.
- Null Hypothesis[ADSS2 W16]: Mean change from baseline in ADSS Item 2 score for baricitinib 1-mg patients is equal to the mean change from baseline in ADSS Item 2 score for placebo patients at Week 16.
- Null Hypothesis[ITCH W2]: Proportion of baricitinib 1-mg patients achieving a 4-point improvement in Itch NRS is equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 2 among patients with baseline Itch NRS score ≥4.
- Null Hypothesis[ADSS2 W1]: Mean change from baseline in ADSS Item 2 score for baricitinib 1-mg patients is equal to the mean change from baseline in ADSS Item 2 score for placebo patients at Week 1.
- Null Hypothesis[ITCH W1]: Proportion of baricitinib 1-mg patients achieving a 4-point improvement in Itch NRS is equal to the proportion of placebo patients achieving a 4-point improvement in Itch NRS at Week 1 among patients with baseline Itch NRS score ≥4.

A multiple testing strategy for the primary and key secondary endpoints is implemented through a graphical testing scheme depicted in Figure JAIN.6.1.

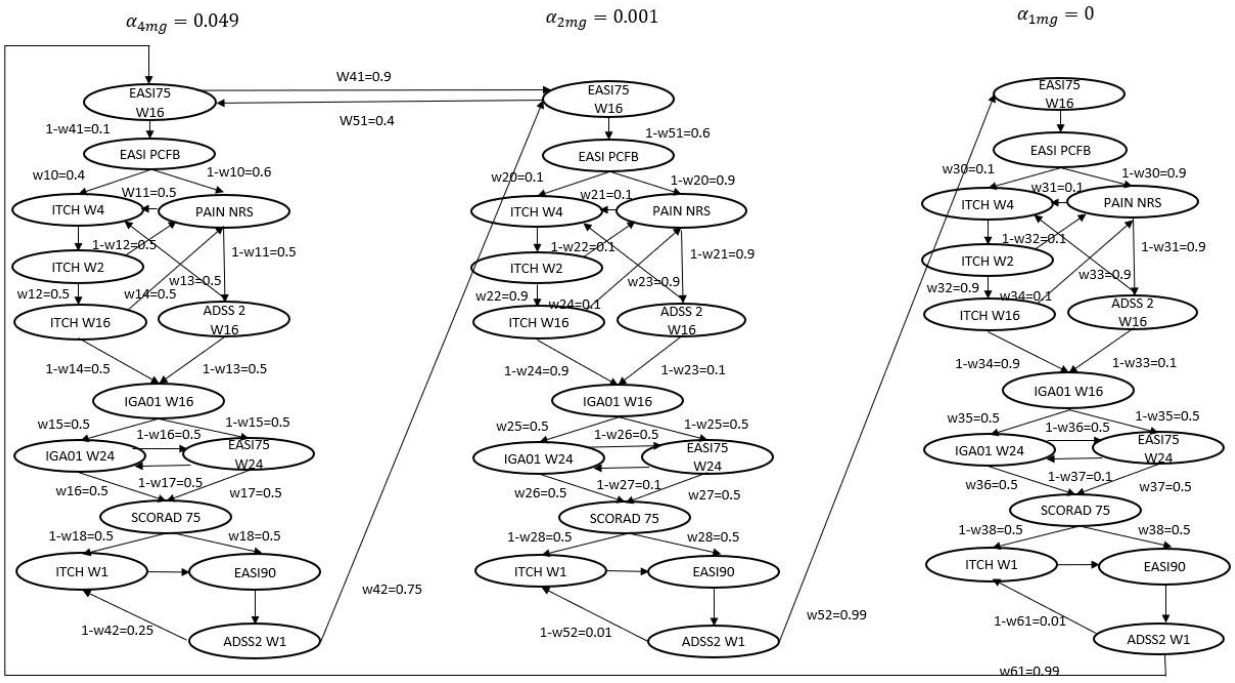


Figure JAIN.6.1. Illustration of graphical multiple testing procedure with initial α allocation and weights.

The primary null hypothesis includes testing whether the baricitinib 4-mg + TCS and the baricitinib 2-mg dose + TCS are superior to placebo + TCS at the primary endpoint defined as the proportion of patients achieving EASI75 at Week 16.

Figure JAIN.6.1 depicts the graphical testing scheme (including testing order, interrelationships, Type I error allocation, and the associated propagation).

There will be no adjustment for multiple comparisons for any other analyses.

6.5. 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.

For 24-week interim analysis, patient disposition through Period 2 will be summarized using the ITT population. Frequency counts and percentages of patients who complete the study treatment or discontinue treatment early will also be summarized separately by treatment group for patients who are not rescued and for patients who are rescued, along with the reason for study treatment discontinuation.

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, will be summarized overall and separately by treatment group for patients who are not rescued and for patients who are rescued, along with the reason for study discontinuation.
Reasons for discontinuation from the study will be summarized by treatment group and compared between groups using Fisher's exact test.

6.6. Patient Characteristics

Patient characteristics including demographics and baseline characteristics will be summarized descriptively for the **24-week interim analysis** (ITT population).

Historical illnesses and pre-existing conditions will be summarized descriptively for the **24-week interim analysis** (ITT population) by treatment group for the ITT population.

No formal statistical comparisons will be made among treatment groups unless otherwise stated.

6.6.1. Demographics

The following demographic information will be categorized and presented as above:

- Age (years)
- Age group (<65 years vs ≥65 years)
- Age group (<65 years, ≥65 years to <75 years, ≥75 years to <85 years, ≥85 years)
- 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)
- Country
- Weight (kg)
- Weight group (<60 kg, ≥60 kg to <100 kg, ≥100 kg)
- Height (cm)
- BMI (kg/m^2)
- BMI group (<25 kg/m^2, ≥25 to <30 kg/m^2, ≥30 kg/m^2)
- Ethnicity (Hispanic, non-Hispanic)

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

6.6.2. 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)
- Duration since AD diagnosis category (0 to <2 years, 2 years to <5 years, 5 years to <10 years, 10 years to <20 years, ≥20 years)
- Age at Diagnosis (years)
- Age Group at Diagnosis (<18 years, ≥18 years to <50 years, ≥50 years)
- Habits (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)
- Validated IGA) score
- Eczema Area and Severity Index (EASI) score
- Scoring Atopic Dermatitis (SCORAD)
- Body Surface Area (BSA) affected by AD
- Hospital Anxiety Depression Scale (HADS) subscales
- Patient Oriented Eczema Measure (POEM)
- Itch Numeric Rating Scale (NRS)
- Atopic Dermatitis Sleep Scale (ADSS) Item 2
- Dermatology Life Quality Index (DLQI)
- Skin Pain NRS
- Patient Global Impression of Severity AD (PGI-S-AD)
- Baseline renal function status: impaired (estimated glomerular filtration rate [eGFR] <60 mL/min/1.73 m^2) or not impaired (eGFR ≥60 mL/min/1.73 m^2)
- Immunoglobulin E (IgE): intrinsic (<200 kU/I) or extrinsic (≥200 kU/I)
- Prior therapy
    - Topical therapy only
    - Systemic therapy only
- Prior phototherapy use
    - Yes, No
    - Yes (type of phototherapy), No
    - If yes, reason for discontinuation of latest phototherapy
- Prior topical calcineurin inhibitor (TCNI) therapy use
    - Yes, No (reason for not using TCNI)

- Prior cyclosporine use (all prior uses, independent of the outcome)
    - Yes, No
- Prior cyclosporine use (cyclosporine ineligible, eligibility criteria)
    - Yes, No

Note: Cyclosporine ineligible (yes, no)
Set to "yes" if the following reasons were selected for either not using the medication or discontinuing the medication:

- Reason for not using the medication: contraindication
- Reason for discontinuation (at least once in case patient took cyclosporine on several occasions): inadequate response, intolerance to medication, or contraindication

All patients in JAIN should be cyclosporine ineligible.

6.6.3. 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 selected historical diagnoses will be summarized by treatment group using the ITT population. Historical diagnoses will be categorized using the Medical Dictionary for Regulatory Activities (MedDRA®, most current available version) algorithmic standardized MedDRA queries (SMQs) or similar predefined lists of preferred terms (PTs) of interest.

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 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. Pre-existing conditions will be categorized using the MedDRA SMQs or similar predefined lists of PTs of interest. Frequency counts and percentages of patients with selected pre-existing conditions will be summarized by treatment group using the ITT population.

6.7. Treatment Compliance

For 24-week interim analysis, patient compliance with study medication will be assessed for the following:

- Week 0 (Visit 2) through Week 16 (Visit 8) or Early Termination using the ITT population.
- Week 0 (Visit 2) through Week 24 (Visit 10) or Early Termination using the ITT population.

All patients are expected to take 3 tablets daily from a blister pack as described in the protocol. Each blister pack contains 27 tablets. 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. 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 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 – date of first dose + 1) – number of days of temporary drug interruption]*number of tablets taken per day

A patient will be considered significantly noncompliant if he or she misses more than 20% of the prescribed doses of investigational product during the study, unless the patient's investigational product is withheld by the investigator for safety reasons. Similarly, a patient will be considered significantly noncompliant if he or she is judged by the investigator to have intentionally or repeatedly taken 20% more than the prescribed amount of investigational product during the study. Patients who are significantly noncompliant through Week 16 will be excluded from the PPS population.

Descriptive statistics for percent compliance and noncompliance rate will be summarized for the ITT population by treatment group for Week 0 through Week 16 and Week 0 through Week 24. Subintervals of interest, such as compliance between visits, may also be presented. The number of expected doses, tablets dispensed, tablets returned, and percent compliance will be listed by patient for Week 0 through Week 16 and Week 0 through Week 24.

6.8. Background Therapy

Throughout the study, background TCS therapy (for example, triamcinolone 0.1% cream and/or hydrocortisone 2.5% ointment) are to be used on active lesions (see JAIN protocol Section 7.7.2).

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 the patients did not use background TCS will be summarized by both potencies throughout the entire 16-week treatment period and 24-week treatment period. The main analysis applies censoring rule #1. After patients who are rescued or discontinue investigational products, it is assumed that

background TCS would be applied each day. In case of missing values in the daily diary, it will be assumed that background TCS has been used. Analysis will be done for Week 16 via ANOVA, with geographic region, baseline disease severity, and treatment as factors in the model. A secondary analysis will apply censoring rule #2, with the same assumptions for missing values as described above.

- 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 16-week and 24-week treatment period. The total amount in grams for low and moderate potency, as well as the sum of both potencies will be summarized between visits. If a returned tube is not weighed in grams, 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 main analysis on the total amount of background TCS throughout the entire 16-week treatment period will apply censoring rule #1. After patients who get rescued or discontinue investigational product, whichever is earlier, it is assumed that they would use the same amount of TCS as they did before. Analysis will be done via analysis of variance (ANOVA), with geographic region, baseline disease severity, and treatment as factors in the model. The secondary analysis will apply censoring rule #2 with the same assumptions 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 number of days without background TCS cannot be performed. Whether any background TCS is used for each patient is also collected in the diary device each day starting from the first dose date throughout the study.

6.9. Previous and Concomitant Therapy

For 24-week interim analysis, summaries of previous and concomitant medications will be based on the ITT population.

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 (for example, 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 concomitant for the treatment period.

Summaries of previous medications will be as follows:

- Previous AD therapies

Summaries of concomitant medications, with sponsor and non-sponsor provided background TCS included, will be as follows:

- General 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 6.2.2.

Efficacy analyses will generally be analyzed according to the following formats and patients will be analyzed according to the investigational product to which they were randomized at Week 0 (Visit 2):

- Week 0 to Week 24 with data up to rescue (primary censoring rule)
- Week 0 to Week 24, including data after rescue to nonsystemic therapy for primary and key secondary objectives (secondary censoring rule)

Table JAIN.6.4 includes the descriptions and derivations of the primary, secondary, and exploratory efficacy outcomes (IGA, EASI, BSA, SCORAD).

Table JAIN.6.5 provides the detailed analyses including analysis type, method and imputation, population, time point, and comparisons for efficacy analyses.

Descriptions, derivations and analyses of Health Outcomes/Quality-of-Life Measures (for example, NRS, ADSS, POEM) are detailed in Section 6.11.

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

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with 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. 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: $EASI_{region}$ = (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. Then total EASI score is as follows: $EASI = 0.1*EASI_{head\ and\ neck} + 0.3*EASI_{trunk} + 0.2*EASI_{upper\ limbs} + 0.4*EASI_{lower\ limbs}$	N/A – partial assessments cannot be saved.
		• Change from baseline in EASI score	Change from baseline: observed EASI score – baseline EASI score	Missing if baseline or observed value is missing.
		• Percent change from baseline EASI score	% change from baseline: $100 \times \frac{Observed\ score - Baseline}{Baseline}$	
		• EASI50	% Improvement in EASI score from baseline ≥50%: % change from baseline ≤-50	Missing if baseline or observed value is missing.
		• EASI75	% Improvement in EASI score from baseline ≥75%: % change from baseline ≤-75	Missing if baseline or observed value is missing.
		• EASI90	% Improvement in EASI score from baseline ≥90%: % change from baseline ≤-90	Missing if baseline or observed value is missing.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
		• Time to reaching EASI75	First time reaching EASI75 as event 1, rescue and discontinue for lack of efficacy as event 2, censor up to Week 16	Use observed value, rescue and discontinue for lack of efficacy as competing event, censor up to Week 16.
Validated Investigator's Global Assessment for AD (vIGA AD)	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 0 represents "clear" 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 • IGA (0)	• Observed score of 0 or 1 and change from baseline ≤-2 • Observed score of 0	• Missing if baseline or observed value is missing. • Single item, missing if missing.
		• Time to reaching first IGA (0,1)	• First time reaching IGA (0,1) as event 1, rescue and discontinue for lack of efficacy as event 2, censor up to Week 16	• Use observed value, rescue and discontinue for lack of efficacy as competing event, censor up to Week 16.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with 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 on 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: $BSA = BSA_{head\ and\ neck}/100/0.10 + BSA_{trunk}/100/0.0333 + BSA_{upper\ limbs}/100/0.05 + BSA_{lower\ limbs}/100/0.025$	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.
Scoring Atopic Dermatitis (SCORAD)	The Scoring Atopic Dermatitis 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	• SCORAD score	SCORAD = A/5 + 7B/2 + C, where A is extent of disease, range 0-100 B is disease severity, range 0-18 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 • Percent change from baseline in SCORAD score	Change from baseline: observed SCORAD score – baseline SCORAD score % change from baseline: $100 \times \frac{Observed\ score - Baseline}{Baseline}$	Missing if baseline or observed value is missing.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
	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).	• SCORAD75	% Improvement in SCORAD from baseline ≥75%: % change from baseline ≤-75	Missing if baseline or observed value is missing.
		• SCORAD90	% Improvement in SCORAD from baseline ≥90%: % change from baseline ≤-90	Missing if baseline or observed value is missing.

Abbreviations: AD = atopic dermatitis; EASI50 = 50% improvement from baseline in Eczema Area and Severity Index score; EASI75 = 75% improvement from baseline in Eczema Area and Severity Index score; EASI90 = 90% improvement from baseline in Eczema Area and Severity Index score; N/A = not applicable; SCORAD75 = 75% decrease from baseline in Scoring Atopic Dermatitis values; SCORAD90 = 90% decrease from baseline in Scoring Atopic Dermatitis values.

Table JAIN.6.7. Description of Primary, Secondary, and Exploratory Efficacy Analyses (Period 2)

Measure	Variable	Analysis Method (Section 6.2.2)	Population (Section 6.2.1)	Comparison[a]/Time Point	Analysis Type
Eczema Area and Severity Index (EASI) [*categorical*]	• Proportion of patients achieving EASI75	Logistic regression using NRI (both censoring rules, respectively)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Primary analysis (censoring rule #1)**
		Logistic regression using NRI (both censoring rules, respectively)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 24**	**Key secondary analysis**
		Logistic regression using NRI; with treatment-by-region interaction (censoring rule #1)[b]	ITT	Bari 4-mg or Bari 2-mg vs Bari 1-mg vs PBO; **Week 16, 24**	Sensitivity analysis
		Logistic regression using NRI (censoring rule #1)	PPS	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 24**	Sensitivity analysis
		Logistic regression using pMI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 24**	Sensitivity analysis
		Tipping point analysis (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg vs Bari 1-mg PBO; **Week 16, 24**	Sensitivity analysis
	• Proportion of patients achieving EASI90	Logistic regression using NRI (both censoring rules, respectively)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Key secondary** analysis
		Logistic regression using NRI (censoring rule #1)	PPS	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
		Logistic regression using pMI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis

Measure	Variable	Analysis Method (Section 6.2.2)	Population (Section 6.2.1)	Comparison[a]/Time Point	Analysis Type
		Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory analysis
	• Proportion of patients achieving EASI50	Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Secondary** analysis for Week 16
	• Proportion of patients achieving EASI≤7	Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory analysis
Eczema Area and Severity Index (EASI) [*continuous*]	• EASI score percent change from baseline	MMRM (both censoring rules, respectively)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Key secondary** analysis
		MMRM (censoring rule #1)	PPS	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
		ANCOVA; pMI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
		ANCOVA; mBOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
		MMRM & ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week16**	Exploratory analysis
	• EASI score change from baseline	MMRM & ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory analysis
Validated Investigator's Global Assessment for AD (vIGA AD)	• Proportion of patients achieving IGA (0,1) with a ≥2-point improvement	Logistic regression using NRI (both censoring rules, respectively)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 24**	**Key secondary** analysis **(censoring rule #1)**
		Logistic regression using NRI; with treatment-by-region interaction (censoring rule #1)[b]	ITT	Bari 4-mg or Bari 2-mg vs PBO; **Week 16, 24**	Sensitivity analysis

Measure	Variable	Analysis Method (Section 6.2.2)	Population (Section 6.2.1)	Comparison[a]/Time Point	Analysis Type
		Logistic regression using NRI (censoring rule #1)	PPS	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 24**	Sensitivity analysis
		Logistic regression using pMI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg **vs PBO; Week 16, 24**	Sensitivity analysis
		Tipping point analysis (censoring rule #1)	ITT	Bari 4-mg or BARI 2-mg vs Bari 1-mg vs PBO; **Week 16, 24**	Sensitivity analysis
	• Proportion of patients achieving IGA (0,1) with a ≥2-point improvement	Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 4**	**Secondary** analysis
	• Proportion of patients achieving IGA (0)	Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Secondary** analysis
Body Surface Area (BSA) Affected by AD	• BSA change from baseline	MMRM (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Secondary** analysis for Week 16
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis for Week 16
Scoring Atopic Dermatitis (SCORAD) [*categorical*]	• Proportion of patients achieving SCORAD75	Logistic regression using NRI (both censoring rules, respectively)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Key secondary** analysis
		Logistic regression using NRI (censoring rule #1)	PPS	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
		Logistic regression using pMI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
	• Proportion of patients achieving SCORAD90	Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Secondary** analysis for Week 16

Measure	Variable	Analysis Method (Section 6.2.2)	Population (Section 6.2.1)	Comparison[a]/Time Point	Analysis Type
	• Proportion of patients achieving SCORAD ≤25 (in the subset of patients with SCORAD>25 at baseline)	Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory analysis
Scoring Atopic Dermatitis (SCORAD) [*continuous*]	• SCORAD score change from baseline	MMRM (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Secondary** analysis for Week 16
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis for Week 16
	• SCORAD score percent change from baseline	MMRM & ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory analysis

Abbreviations: AD = atopic dermatitis; ANCOVA = analysis of covariance; Bari = baricitinib; EASI50 = 50% improvement from baseline in Eczema Area and Severity Index score; EASI75 = 75% improvement from baseline in Eczema Area and Severity Index score; EASI90 = 90% improvement from baseline in Eczema Area and Severity Index score; ITT = intent-to-treat; mLOCF = modified last observation carried forward; MMRM = mixed model repeated measures; NRI = nonresponder imputation; NRS = Numeric Rating Scale; PBO = placebo; pMI = placebo multiple imputation; PPS = per protocol set; TCS = TCS = topical corticosteroids.

[a] The addition of the background therapy to the treatment arms ("+ TCS") has been omitted to save space and ease the reading.

[b] This analysis will only be done in case there are sufficient patient numbers per regions to allow the factor "region" in the main analysis model. See Section 6.2.2.2 for details.

6.10.1. Rescue Treatment

A summary of the initial rescue therapy and the reason for rescue will be produced, as well as a summary of the proportion of patients rescued, by study visit and overall, and by type of rescue therapy (high/ultra-high potency TCS, phototherapy, systemic therapy).

A summary of all rescue therapies will be provided.

6.10.2. Primary Outcome and Methodology

Both EASI score and IGA are commonly used in clinical trials, both for qualifying patients for enrollment and for evaluating treatment efficacy (Langley et al. 2015; Futamura et al. 2016; Bożek and Reich 2017). There is no single 'gold standard' disease severity scale for AD; however, IGA scales provide clinically meaningful measures to patients and investigators that are easily described and that correspond to disease severity categories (for example, moderate to severe), and a 75% improvement from Baseline (EASI75) is a commonly used measure of treatment effect in AD clinical trials.

The primary objective of this study is to test the hypothesis that baricitinib 4-mg + TCS or baricitinib 2-mg + TCS is superior to placebo + TCS in the treatment of moderate-to-severe AD, as measured by the proportion of patients achieving EASI75 at Week 16 using the ITT population, and assuming the treatment response disappears after patients are rescued or discontinue from study or 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 noncensor intercurrent events will be imputed using the NRI method described in Section 6.3.1.

A supplemental estimand is to test the hypothesis that baricitinib 4-mg + TCS or baricitinib 2-mg + TCS is superior to placebo + TCS 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 discontinue from study or treatment. In this supplemental estimand, missing data due to the application of the secondary censoring rule and the occurrence of other noncensor intercurrent events will be imputed using the NRI method described in Section 6.3.1.

A logistic regression analysis as described in Section 6.2.2.2 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. Missing data will be imputed using the NRI method described in Section 6.3.1.

Multiplicity controlled analyses will be performed on the primary and key secondary (see Section 6.4) 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 6.4.

6.10.3. Secondary and Exploratory Efficacy Analyses

Multiplicity controlled analyses will be performed on the primary and key secondary (see Section 6.10) objectives 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 6.4.

There will be no adjustment for multiple comparisons for any other analyses. The secondary and exploratory efficacy analyses are detailed in Table JAIN.6.4. Health Outcomes/Quality-of-Life analyses are described in Section 6.11.

For planed exploratory analyses in Section 4.3, if details for exploratory analyses are not included in Table JAIN.6.4, the detail of analysis method will be documented in a supplementary SAP or a supplementary list of analyses.

6.10.4. Sensitivity Analyses

Sensitivity analyses for select outcomes have been previously described and include the following:

- Analyses of key endpoints using the Per-protocol Analysis Set (Section 6.10.2)
- Analyses of key endpoints using the secondary censoring rule (Section 6.10.2)
- Placebo multiple imputation (Section 6.3.4)
- Tipping point analysis (Section 6.3.5)
- The addition of a treatment-by-region interaction to the logistic regression model for the primary outcome (Section 6.2.2.2)
- Analysis of continuous outcomes with ANCOVA (Section 6.2.2.2), with missing data imputed using mLOCF (Section 6.3.3).
- Analysis of continuous outcomes with ANCOVA (Section 6.2.2), with missing data imputed using mBOCF (Section 6.3.6).

6.11. 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 6.2.2.

Health outcomes and quality-of-life measures will generally be analyzed according to the formats discussed in Section 6.2.2.

Table JAIN.6.6 includes the descriptions and derivations of the health outcomes and quality-of-life measures.

Table JAIN.6.7 provides the detailed analyses including analysis type, method and imputation, population, time point, and comparisons for health outcomes and quality-of-life measures.

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

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

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with 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 for how to derive the visit score.
		• Change from baseline in Itch NRS • Percent change from baseline in Itch NRS	Change from baseline: observed Itch score – baseline Itch score % change from baseline: $100 \times \frac{Observed\ score - Baseline}{Baseline}$	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.
		• Cumulative Incidence Function of Time to reaching Itch NRS 4-pt improvement (primary censoring rule)	First time reaching Itch NRS 4-pt improvement as event 1, rescue and discontinue for lack of efficacy as event 2, censor up to Week 16	Use observed value, rescue and discontinue for lack of efficacy as competing event, censor up to Week 16.
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	• Skin Pain NRS score	Single item; range 0 to 10. Refer to Section 6.2.2 for how to derive the visit score.	Refer to Section 6.2.2 for 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.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
	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.	• 4-point Skin Pain improvement in subgroup of patients with baseline Skin Pain NRS ≥4	Change from baseline ≤-4 and baseline ≥4	Missing if baseline is missing or <4 or observed value is missing.
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. Patients 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. Refer to Section 6.2.2 for details on how to calculate the weekly score, which will be used in the continuous analysis.	• Item 1 score of ADSS • Item 2 score of ADSS • 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 visit score.
		• Change from baseline in score of Item 1 of ADSS • Change from baseline in score of Item 2 of ADSS • 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.
		• 1.5-point improvement on Item 2 of ADSS	Change from baseline <= -1.5 and baseline >=1.5 in score of Item 2 of ADSS	Missing if baseline is missing or <1.5 or observed value is missing.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with 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 (Charman 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 one question is unanswered, then the tool is not scored. If more than one 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.
		• 4-point POEM improvement in subgroup of patients with baseline POEM score ≥4	Change from baseline ≤ -4 and baseline ≥4	Missing if baseline is missing or <4 or observed value is missing.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
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." Refer to Section 6.2.2 for details on how to calculate the weekly score which will be used in the continuous analysis.	• 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.
Hospital Anxiety Depression Scale (HADS)	The HADS is a 14-item self-assessment scale that determines the levels of anxiety and depression that a patient is experiencing over the past week. The HADS utilizes a 4-point Likert scale (eg, 0 to 3) for each question and is intended for ages 12 to 65 years (Zigmond and Snaith 1983; White et al. 1999). Scores for each domain (anxiety and depression) can range from 0 to 21, with higher scores indicating greater anxiety or depression (Zigmond and Snaith 1983; Snaith 2003).	• HADS score for anxiety and depression domains	Anxiety domain score is sum of the seven anxiety questions, range 0 to 21; Depression domain score is sum of the seven depression questions, range 0 to 21.	N/A – partial assessments cannot be saved.
		• Change from baseline in HADS domain	Change from baseline: observed HADS domain score – baseline HADS domain score	Missing if baseline or observed value is missing.
		• Change from baseline in HADS total	Change from baseline: observed HADS domain score – baseline HADS total score	Missing if baseline or observed value is missing.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
		• HADS Anxiety <8 in subgroup of patients with baseline HADS Anxiety score ≥8 • HADS Depression <8 in subgroup of patients with baseline HADS Depression score ≥8 • HADS Anxiety or Depression score <8 in subgroup of patients with baseline HADS Anxiety or Depression score ≥8	observed HADS post-baseline <8 and baseline score >=8 for each HADS domain score	Missing if baseline is missing or <8 or observed value is missing.
Dermatology Life Quality Index (DLQI)	The DLQI is a simple, patient-administered, 10-item, validated, quality-of-life questionnaire that covers 6 domains including symptoms and feelings, daily activities, leisure, work and school, personal relationships, and treatment. The recall period of this scale is over the "last week." Response categories include "a little," "a lot," and "very much," with corresponding scores of 1, 2, and 3, respectively, and "not at all," or unanswered ("not relevant") responses scored as 0. Scores range from 0-30 with higher scores indicating greater impairment of quality of life. A	• Symptoms and feelings domain	Sum of questions 1 and 2, range 0 to 6.	N/A – partial assessments cannot be saved.
		• Daily activities domain	Sum of questions 3 and 4, range 0 to 6.	N/A – partial assessments cannot be saved.
		• Leisure domain	Sum of questions 5 and 6, range 0 to 6.	N/A – partial assessments cannot be saved.
		• Work and school domain	Sum of questions 7 and 7B (if it is answered), range 0 to 3. Responses of "yes" and "no" on Question 7 are given scores of 3 and 0 respectively. If Question 7 is answered "no" then Question 7b is answered with "a lot", "a little", "not at all" getting scores of 2, 1, 0 respectively.	N/A – partial assessments cannot be saved.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
	DLQI total score of 0 to 1 is considered as having no effect on a patient's health-related QoL (Hongbo et al. 2005), and a 4-point change from baseline is considered as the minimal clinically important difference threshold (Khilji et al. 2002; Basra et al. 2015).	• Personal relationships domain	Sum of questions 8 and 9, range 0 to 6.	N/A – partial assessments cannot be saved.
		• Treatment domain	Question 10, range 0 to 3.	N/A – partial assessments cannot be saved.
		• DLQI total score	DLQI total score: sum of all 6 DLQI domain scores, range 0 to 30.	N/A – partial assessments cannot be saved.
		• Change from baseline in DLQI	Change from baseline: observed DLQI score – baseline DLQI score	Missing if baseline or observed value is missing.
		• DLQI total score ≤ 5 in subgroup of patients who had baseline DLQI >5	Post-baseline DLQI total score ≤ 5 with baseline total score >5	Missing if baseline is missing or <=5 or observed value is missing
		• DLQI total score in (0,1)	Post-baseline DLQI total score in (0,1)	N/A – partial assessments cannot be saved.
		• 4-point DLQI improvement in subgroup of patients with baseline DLQI total score ≥ 4	Change from baseline ≤ -4 and baseline ≥ 4	Missing if baseline is missing or <4 or observed value is missing.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
Work Productivity and Activity Impairment: Atopic Dermatitis (WPAI-AD)	The WPAI-AD records impairment due to AD during the past 7 days. The WPAI-AD 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), with higher scores indicating greater impairment and less productivity.	• Employment status	Q1	Single item, missing if missing.
		• Change in employment status	Employed at baseline and remained employed: Q1 = 1 at post-baseline visit and at baseline visit. Not employed at baseline and remain unemployed: Q1 = 0 at post-baseline 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 with 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–5 Levels (EQ-5D-5L)	The EQ-5D-5L is a standardized measure of health status that provides a simple, generic measure of health for clinical and economic appraisal. The EQ-5D-5L consists of 2 components: a descriptive system of the respondent's health and a rating of his or her current health state using a 0 to 100 mm VAS. The descriptive system comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The respondent is asked to indicate his or her health state by ticking (or placing a cross) in the box associated with the most appropriate statement in each of the 5 dimensions. It should be noted that the numerals 1 to 5 have no arithmetic properties and should not be used as an ordinal score. The VAS records the respondent's self-rated health on a	• EQ-5D mobility • EQ-5D self-care • EQ-5D usual activities • EQ-5D pain/ discomfort • EQ-5D anxiety/ depression	Five health profile dimensions, each dimension has 5 levels: 1 = no problems 2 = slight problems 3 = moderate problems 4 = severe problems 5 = extreme problems It should be noted that the numerals 1 to 5 have no arithmetic properties and should not be used as a primary score.	Each dimension is a single item, missing if missing.
		• EQ-5D VAS	Single item. Range 0 to 100. 0 represents "worst health you can imagine" 100 represents "best health you can imagine"	Single item, missing if missing.
		• Change from baseline in EQ-5D VAS	Change from baseline: observed EQ-5D VAS score – baseline EQ-5D VAS score	Missing if baseline or observed value is missing.
		• EQ-5D-5L UK Population-based index score (health state index)	Derive EQ-5D-5L UK Population-based index score according to the link by using the UK algorithm to produce a patient-level index score between -0.59 and 1.0 (continuous variable).	N/A – partial assessments cannot be saved on the eCOA tablet.
		• Change from baseline in EQ-5D-5L UK Population-based index score	Change from baseline: observed EQ-5D-5L UK score – baseline EQ-5D-5L UK score	Missing if baseline or observed value is missing.
		• EQ-5D-5L US Population-based index	Derive EQ-5D-5L US Population-based index score according to the link by using	N/A – partial assessments

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
	vertical VAS where the endpoints are labeled "best imaginable health state" and "worst imaginable health state." This information can be used as a quantitative measure of health outcome. The EQ-5D-5L health states, defined by the EQ-5D-5L descriptive system, may be converted into a single summary index by applying a formula that essentially attaches values (also called weights) to each of the levels in each dimension (Herdman et al. 2011; EuroQol Group 2015 [WWW]).	score (health state index)	the US algorithm to produce a patient-level index score between -0.11 and 1.0 (continuous variable).	cannot be saved on the eCOA tablet.
		• Change from baseline in EQ-5D-5L US Population-based index score	Change from baseline: observed EQ-5D-5L US score – baseline EQ-5D-5L US score	Missing if baseline or observed value is missing.
Medical Outcomes Study 36-Item Short Form Health Survey Version 2 Acute (SF-36)	The SF-36 is a 36-item patient administered measure designed to be a short, multipurpose assessment of health in the areas of physical functioning, role – physical, role – emotional, bodily pain, vitality, social functioning, mental health, and general health. The 2 overarching domains of mental well-being and physical well-being are captured by the Mental Component Summary and Physical Component Summary scores. The summary scores range from 0 to 100; higher scores indicate better levels of function and/or	8 associated domain scores: • Physical Functioning, • Role Physical, • Bodily Pain, • General Health, • Vitality, • Social Functioning, • Role Emotional, • Mental Health 2 component Scores: • MCS Score • PCS Score	Per copyright owner, the QualityMetric Health Outcomes™ Scoring Software 4.5 will be used to derive SF-36 domain and component scores. After data quality-controls, the SF-36 software will re-calibrate the item-level responses for calculation of the domain and component scores. These raw scores will be transformed into the domain scores (t-scores) using the 1-week recall period. The procedure to derive the SF-36 scores is described in the JAIN ADaM specs. It entails exporting the patient data in a CSV or tab delimited file for import, generation of the SF-36 scores and reports, and export of the calculated scores in a CSV or tab delimited file for integration into SDTM/ADaM datasets. The	N/A – partial assessments cannot be saved.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
	better health. Items are answered on Likert scales of varying lengths. The SF-36 version 2 (acute version) will be used, which utilizes a 1-week recall period (Ware 2000). (Brazier et al. 1992; Ware and Sherbourne 1992). The following responder definition values, in terms of T-score points, are proposed for SF-36 v2 component and scale individual respondent scores: PCS, 3.8; MCS, 4.6 (Maruish 2011).		summary scores range from 0 to 100.	
		• Change from Baseline in health domain scores	Change from Baseline: observed SF-36 domain score – Baseline SF-36 domain score	Missing if Baseline or observed value is missing.
		• Change from Baseline in PCS	Change from Baseline: observed SF-36 PCS – Baseline SF-36 PCS	Missing if Baseline or observed value is missing.
		• Change from Baseline in MCS	Change from Baseline: observed SF-36 MCS – Baseline SF-36 MCS	Missing if Baseline or observed value is missing.

Abbreviations: AD = atopic dermatitis; eCOA = Electronic version of Clinical Outcome Assessment; EQ-5D = European Quality of Life–5 Dimensions; N/A = not applicable; QoL = quality of life; VAS = visual analog scale.

Table JAIN.6.9. Description of Health Outcomes and Quality-of-Life Measures Analyses (Period 2)

Measure	Variable	Analysis Method (Section 6.2.2.2)	Population (Section 6.2.1)	Comparison[a]/Time Point	Analysis Type
Itch Numeric Rating Scale (NRS) [*categorical*]	• Proportion of patients achieving a 4-point improvement in Itch NRS (in the subset of patients who had baseline Itch NRS ≥4)	Logistic regression using NRI (both censuring rules, respectively)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 4, 2, 1**	**Key Secondary** Analysis
		Logistic regression using NRI (censoring rule #1)	PPS	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 4, 2, 1**	Sensitivity analysis
		Logistic regression using pMI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 4, 2, 1**	Sensitivity analysis
		Tipping point analysis (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg vs PBO; **Week 16**	Sensitivity analysis
		Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week24**	**Secondary** Analysis
	• Number of Itch-free (Itch NRS = 0) Days	Descriptive statistics (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Baseline to Week 16**	Exploratory Analysis
	• Proportion of patients achieving an Itch NRS of 0	Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 4, 16**	Exploratory Analysis
Itch Numeric Rating Scale (NRS) [*continuous*]	• Itch NRS score percent change from baseline	MMRM (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 24, 16, 4, 1**	**Secondary** Analysis
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 24, 16, 4, 1**	Sensitivity analysis
	• Itch NRS score change from baseline	MMRM & ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 24, 16, 4, 1**	Exploratory Analysis

Measure	Variable	Analysis Method (Section 6.2.2.2)	Population (Section 6.2.1)	Comparison[a]/Time Point	Analysis Type
	• Time to 4-point reduction in Itch NRS (in the subset of patients who had baseline Itch NRS ≥4)	CPH model (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO;	Exploratory Analysis
Skin Pain Numeric Rating Scale (NRS)	• Skin Pain NRS score change from baseline	MMRM (both censuring rules, respectively)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Key Secondary** Analysis
		MMRM (censoring rule #1)	PPS	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
		ANCOVA; pMI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
		ANCOVA; mBOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
	• Number of Skin Pain-free (Skin pain NRS = 0) Days	Descriptive statistics	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Baseline to Week 16**	Exploratory Analysis
Atopic Dermatitis Sleep Scale (ADSS)	• ADSS item 2 score change from baseline	MMRM (both censuring rules, respectively)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 1**	**Key Secondary** Analysis
		MMRM (censoring rule #1)	PPS	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 1**	Sensitivity analysis
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 1**	Sensitivity analysis
		ANCOVA; mBOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 1**	Sensitivity analysis
		ANCOVA; pMI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 1**	Sensitivity analysis
	• ADSS items 1 and 3 scores change from baseline	MMRM & ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16, 1**	Exploratory analysis

Measure	Variable	Analysis Method (Section 6.2.2.2)	Population (Section 6.2.1)	Comparison[a]/Time Point	Analysis Type
	• 1.5-point improvement on Item 2 of ADSS	Logistic regression using NRI (primary censoring rule)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory Analysis
Patient-Oriented Eczema Measure (POEM) *[continuous]*	• POEM score change from baseline	MMRM (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Secondary** analysis
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
Patient-Oriented Eczema Measure (POEM) *[categorical]*	• Proportion of Patients achieving an MCID of POEM improvement ≥3.4 • Proportion of Patients achieving an MCID of POEM improvement ≥4	Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory analysis
Patient Global Impression of Severity–Atopic Dermatitis (PGI-S-AD)	• PGI-S-AD score change from baseline	MMRM (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Secondary** analysis
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
Hospital Anxiety Depression Scale (HADS)	• HADS change from baseline in domain scores: - anxiety - depression	MMRM (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Secondary** analysis
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis

Measure	Variable	Analysis Method (Section 6.2.2.2)	Population (Section 6.2.1)	Comparison[a]/Time Point	Analysis Type
	• Proportion of patient achieving HADS Anxiety < 8 in subgroup of patients who had baseline HADS Anxiety >= 8 - Anxiety - Depression - total	Logistic regression using NRI (primary censoring rule)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory analysis
Dermatology Life Quality Index (DLQI) [*categorical*]	• Proportion of patients achieving a DLQI score of 0 or 1 in patients with DLQI>1 at baseline • DLQI 4-pt improvement in subgroup of patients who had baseline DLQI >= 4 • DLQI total score <=5 in subgroup of patients who had baseline DLQI > 5	Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory Analysis
Dermatology Life Quality Index (DLQI) [*continuous*]	• DLQI total score change from baseline	MMRM (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Secondary** analysis
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
	• DLQI Observed and change from baseline in domain scores - Symptoms and feelings - Daily activities - Leisure - Work and school - Personal relationships - Treatment	MMRM & ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory analysis
Work Productivity and Activity Impairment: Atopic	• Observed and Change from baseline in employment status	Descriptive statistics (observed, no censoring)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO: **Week 16**	**Secondary** analysis (Week 16)

Measure	Variable	Analysis Method (Section 6.2.2.2)	Population (Section 6.2.1)	Comparison[a]/Time Point	Analysis Type
Dermatitis (WPAI-AD)	• Observed and Change from baseline in: - absenteeism - presenteeism - overall work impairment - activity impairment	MMRM (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Secondary** analysis
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
European Quality of Life–5 Dimensions–5 Levels (EQ-5D-5L)	• Observed values in - EQ-5D mobility - EQ-5D self-care - EQ-5D usual activities - EQ-5D pain/ discomfort - EQ-5D anxiety/ depression	Logistic Regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO: **Week 16**	Exploratory Analysis
	• Observed and Change from baseline in - EQ-5D VAS - EQ-5D-5L UK Population-based index score - EQ-5D-5L US Population-based index score	MMRM (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	**Secondary** analysis
		ANCOVA; mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Sensitivity analysis
Medical Outcomes Study 36-Item Short Form Health Survey Version 2 Acute (SF-36)	• Observed and Change from baseline in summary scores: - MCS (mental component score) - PCS (physical component score)	MMRM & ANCOVA;mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory analysis
	• Proportion of Patients achieving a response of: - MCS improvement ≥4.6 - PCS improvement ≥3.8	Logistic regression using NRI (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory analysis

Measure	Variable	Analysis Method (Section 6.2.2.2)	Population (Section 6.2.1)	Comparison[a]/Time Point	Analysis Type
Medical Outcomes Study 36-Item Short Form Health Survey Version 2 Acute (SF-36)	• Observed and Change from baseline in domain scores: - physical functioning - bodily pain - role limitations due to physical problems - role limitations due to emotional problems - general health perceptions - mental health - social function - vitality	MMRM & ANCOVA;mLOCF (censoring rule #1)	ITT	Bari 4-mg or Bari 2-mg or Bari 1-mg vs PBO; **Week 16**	Exploratory analysis

Abbreviations: ANCOVA = analysis of covariance; Bari = baricitinib; CPH = Cox proportional hazard; ITT = intent-to-treat; MCID = minimal clinically important change; mLOCF = modified last observation carried forward; MMRM = mixed model repeated measures; NRI = nonresponder imputation; PBO = placebo; pMI = placebo multiple imputation; PPS = per protocol set; VAS = visual analogue scale.

[a] The addition of the background therapy to the treatment arms ("+ TCS") has been omitted to save space and ease the reading.

6.12. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

Pharmacokinetic, Pharmacodynamic and Biomarker analyses to address secondary and exploratory objectives of this study will be described by Lilly in separate Pharmacokinetic/Pharmacodynamic and Biomarker analysis plans.

6.13. Safety Analyses

The general methods used to summarize safety data, including the definition of baseline value are described in Section 6.2.2.

Safety analyses will include all data including rescue to any of the 3 rescue therapy options, unless otherwise stated, and patients will be analyzed according to the investigational product to which they were randomized at Week 0 (Visit 2). Additional analyses may be conducted using data after rescue to systemic therapy for some safety topics such as systemic treatment-emergent adverse events (TEAEs), and serious adverse events (SAEs). Safety analyses will use the safety population defined in Section 6.2.1.

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 6.13.5), and investigational product interruptions.

Unless otherwise specified, by-visit summaries will include planned on-treatment visits. For tables that summarize events (such as AEs, categorical laboratory abnormalities, shift to maximum severity), 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 including rescue, regardless of study period.

For the interim lock(s), all safety data from continuing patients at time of the interim lock will be included in the safety analysis. Safety data from patients who permanently discontinued study drug prior to an interim lock will be included in the interim lock safety analysis up to 30 days post last dose.

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

The Compound level safety standards provide further details and information.

6.13.1. 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*

Duration of exposure (in days) for the "PBO/BARI 1-mg to BARI 2-mg" and the "PBO/BARI 1-mg/BARI 2-mg to BARI 4-mg" groups 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 after the switch + 1*

Last dose of study drug including rescue is calculated as last date on study drug. See the Compound level safety standards for more details.

Total patient-years (PY) of exposure will be reported for each treatment group 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 will be reported in weeks using the following:

- ≥4 weeks, ≥8 weeks, ≥12 weeks, ≥16 weeks, ≥24 weeks, ≥52 weeks, ≥76 weeks, and ≥104 weeks
- >0 to <4 weeks, ≥4 weeks to <8 weeks, ≥8 weeks to <12 weeks, ≥12 to <16 weeks, ≥16 to 24 weeks, ≥24 to 52 weeks, ≥52 to 76 weeks, ≥76 to 104 weeks, and ≥104 weeks

The exposure ranges for the interim lock(s) will be adjusted accordingly.

Overall exposure will be summarized in total PY, which is calculated according to the following formula:

Exposure in PY (PYE) = sum of duration of exposure in days (for all patients in treatment group) / 365.25

6.13.2. Adverse Events

Adverse events are recorded in the eCRFs. Each AE will be coded to System Organ Class (SOC) and Preferred Term (PT) using the Medical Dictionary for Regulatory Activities (MedDRA) version that is current at the time of database lock. Severity of AEs is recorded as mild, moderate, or severe.

A treatment-emergent adverse event (TEAE) is defined as an event that either first occurred or worsened in severity after the first dose of study treatment and on or prior to the last visit date during the analysis period. The analysis period is defined as the treatment period plus up to 30 days off-drug follow-up time.

Adverse events are classified based upon the MedDRA PT. 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 pre-existing during the baseline period, and persists during the treatment period, then the baseline severity will be considered mild for determining treatment-emergence (that is, the event is treatment-emergent if the severity is coded moderate or

severe postbaseline and not treatment-emergent if the severity is coded mild postbaseline). 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. The day and time for events where onset is on the day of the first dose of study treatment will both be used to distinguish between pretreatment and posttreatment to derive treatment-emergence. Should there be insufficient data for AE start date to make this comparison (for example, the AE start year is the same as the treatment start year, but the AE start month and day are missing), the AE will be considered treatment-emergent.

In general, summaries will include the number of patients in the safety population (N), frequency of patients experiencing the event (n), and relative frequency (that is, percentage; n/N*100). For any events that are gender-specific based on the displayed PT, the denominator used to compute the percentage will only include patients from the given gender.

For events of interest, incidence rate (IR) per 100 patient-years of observation (PYR) may be provided for the final analysis. See the Compound level safety standards for more details.

In an overview table, the number and percentage of patients in the safety population who experienced death, an SAE, any TEAE, discontinuation from the study due to an AE, permanent discontinuation from study drug due to an AE, temporary interruption of study drug due to an AE, or a severe TEAE will be summarized by treatment group.

The number and percentage of patients with TEAEs will be summarized by treatment group in 2 formats:

- by MedDRA PT nested within SOC with decreasing frequency in SOC, and events ordered within each SOC by decreasing frequency in the baricitinib 4-mg group
- by MedDRA PT with events ordered by decreasing frequency in the baricitinib 4-mg group.

6.13.2.1. Common Adverse Events

Common TEAEs are defined as TEAEs that occurred in ≥2% (before rounding) of patients in any treatment group including placebo. The number and percentage of patients with common TEAEs will be summarized by treatment using MedDRA PT ordered by decreasing frequency in the baricitinib 4-mg group.

The number and percentage of patients with TEAEs will be summarized by maximum severity by treatment using MedDRA PT ordered by decreasing frequency in the baricitinib 4-mg group for the common TEAEs. For each patient and TEAE, the maximum severity for the MedDRA level being displayed is the maximum postbaseline severity observed from all associated LLTs mapping to that MedDRA PT.

6.13.2.2. Serious Adverse Event Analyses

Consistent with the International Conference on Harmonisation (ICH) E2A guideline (1994) and 21 Code of Federal Regulations (CFR) 312.32 (a) (2010), an SAE is any AE that results in any one of the following outcomes:

- Death
- Initial or prolonged inpatient hospitalization
- A life-threating experience (that is, immediate risk of dying)
- Persistent or significant disability/incapacity
- Congenital anomaly/birth defect
- Important medical events that may not be immediately life-threatening or result in death or hospitalization but may jeopardize the patient or may require intervention to prevent one of the other outcomes listed in the definition above.

The number and percentage of patients who experienced any SAE will be summarized by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency in the baricitinib 4-mg group within decreasing frequency in SOC. The SAEs will also be summarized by treatment using MedDRA PT without SOC.

An individual listing of all SAEs will be provided. A listing of deaths, regardless of when they occurred during the study, will also be provided.

6.13.2.3. Other Significant Adverse Events

Other significant AEs to be summarized will provide the number and percentage of patients who

- permanently discontinued study drug because of an AE or death, and
- temporarily interrupted study drug because of AE

by treatment using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency in the baricitinib 4-mg group within decreasing frequency in SOC.

A summary of temporary interruptions of study drug will also be provided, showing the number of patients who experienced at least 1 temporary interruption and the number of temporary interruptions per patient with an interruption. Further, the duration of each temporary interruption (in days), the cumulative duration of dose interruption (in days) using basic descriptive statistics, and the reason for dose interruption will be provided.

A listing of all AEs leading to permanent discontinuation from the study drug or from the study will be provided. A listing of all temporary study drug interruptions, including interruptions for reasons other than AEs, will be provided.

6.13.2.4. Criteria for Notable Patients

Patient narratives will be provided for all patients who experience certain “notable” events prior to data cutoff date for the submission. See the Compound level safety standards for a list of criteria.

6.13.3. Clinical Laboratory Evaluation

For the categorical laboratory analyses (shift and treatment-emergent), the analysis period is defined as the treatment period plus up to 30 days off-drug follow-up time. The analysis period for the continuous laboratory analyses (for example, change from baseline by time point) is defined as the treatment period excluding off-drug follow-up time.

All laboratory tests will be presented using the International Système (SI) and US conventional (CN) units. The performing central laboratory reference ranges will be used to define the low and high limits. Key results pertaining to the 4 key hepatic laboratory assessments (alanine aminotransferase [ALT], aspartate aminotransferase [AST], total bilirubin, and alkaline phosphatase [ALP]) will be included as a separate analysis to address the risk of liver injury as a special safety topic (see Section 6.13.5.1).

There is 1 special circumstance for laboratory values to be derived based on regularly scheduled, protocol-specified analytes. The low-density lipoprotein/high-density lipoprotein (LDL/HDL) ratio will be derived as the ratio of LDL cholesterol to HDL cholesterol. There are no central lab reference ranges for the LDL/HDL ratio.

The following will be conducted for the laboratory analytes collected quantitatively:

- Box plots: Values at each visit (starting from randomization) and change from last baseline to each visit and to last postbaseline measure will be displayed in box plots for patients who have both a baseline and at least 1 postbaseline visit. The last non-missing observation in the treatment period will be used as the last observation. Individual measurements outside of reference limits will also be displayed using distinct symbols overlaying the box plot. Original-scale data will be used for the display but for some analytes (for example, immunoglobulins) a logarithmic scale may be used to aid in viewing the measures of central tendency and dispersion. Unplanned measurements will be excluded. Descriptive summary statistics will be included below the box plot along with p-values resulting from between treatment comparison in change from last baseline to last observation. An ANCOVA model with explanatory term for treatment and the baseline value as a covariate will be used. These box plots will be used to evaluate trends over time and to assess a potential impact of outliers on central tendency summaries.
- Treatment-emergent high/low analyses: The number and percentage of patients with treatment-emergent high and low laboratory results at any time will be summarized by treatment group. Planned and unplanned measurements will be included. A treatment-emergent **high** result is defined as a change from a value less than or equal to the high limit at all baseline visits to a value greater than the high limit at any time during the treatment period. A treatment-emergent **low** result is defined as a change from a value greater than or equal to the low limit at all baseline visits to a value less than the low limit at any time during the treatment period. The Fisher's exact test will be used for the treatment comparisons.

For laboratory analyte measurements collected qualitatively, a listing of abnormal findings will be provided. The listing will include but not be limited to patient identifier (ID), treatment

group, laboratory collection date, analyte name, and analyte finding. If needed by the safety physician/scientist, for analytes measured qualitatively, the number and percentage of patients with treatment-emergent abnormal laboratory results at any time will be summarized by treatment. Planned and unplanned measurements will be included. A treatment-emergent abnormal result is defined as a change from normal at all baseline visits to abnormal at any time postbaseline.

The listing of specific reference ranges used in analysis of laboratory data will be provided.

Note that additional analyses of certain laboratory analytes will be discussed within subsections of Section 6.13.5 pertaining to Special Safety topics (Section 6.13.5.1 for hepatic analytes, Section 6.13.5.2 for analytes related to hematologic changes, Section 6.13.5.3 for analytes related to lipids, Section 6.13.5.4 for analytes related to renal function, and Section 6.13.5.5 for creatine phosphokinase [CPK]).

6.13.4. Vital Signs and Other Physical Findings

For the treatment-emergent categorical analyses (shift and treatment-emergent), the analysis period is defined as the treatment period plus up to 30 days off-drug follow-up time. The analysis period for the continuous analyses (for example, change from baseline by time point) is defined as the treatment period excluding off-drug follow-up time.

Vital signs and physical characteristics include systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse, weight, and BMI. Original-scale data will be analyzed. When these parameters are analyzed as continuous numerical variables, unplanned measurements will be excluded. When these parameters are analyzed as categorical outcomes and/or treatment-emergent abnormalities, planned and unplanned measurements will be included.

The planned analyses described for the laboratory analytes in Section 6.13.3 will be used to analyze the vital signs and physical characteristics.

Table JAIN.6.8 defines the low and high baseline values as well as the criteria used to define treatment-emergence based on postbaseline values. The blood pressure and pulse rate criteria are consistent with the document *Selected Reference Limits for Pulse/Heart Rate, Arterial Blood Pressure (Including Orthostasis), and Electrocardiogram Numerical Parameters for Use in Analyses of Phase 2-4 Clinical Trials Version 1.3* approved on April 29, 2015 as recommended by the Lilly Cardiovascular Safety Advisory Committee (CVSAC).

Table JAIN.6.10. Categorical Criteria for Abnormal Treatment-Emergent Blood Pressure and Pulse Measurement and Categorical Criteria for Weight Changes for Adults

Parameter (Units of Measure)	Low	High
Systolic Blood Pressure (mm Hg)	≤90 (low limit) and decrease from lowest value during baseline ≥20 if >90 at each baseline visit	≥140 (high limit) and increase from highest value during baseline ≥20 if <140 at each baseline visit
Diastolic Blood Pressure (mm Hg)	≤50 (low limit) and decrease from lowest value during baseline ≥10 if >50 at each baseline visit	≥90 (high limit) and increase from highest value during baseline ≥10 if <90 at each baseline visit
Pulse (beats per minute)	<50 (low limit) and decrease from lowest value during baseline ≥15 if ≥50 at each baseline visit	>100 (high limit) and increase from highest value during baseline ≥15 if ≤100 at each baseline visit
Weight (kilograms)	(Loss) decrease ≥7% from lowest value during baseline	(Gain) increase ≥7% from highest value during baseline

6.13.5. 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.

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.

6.13.5.1. Abnormal Hepatic Tests

Analyses for abnormal hepatic tests will involve 4 laboratory analytes: ALT, AST, total bilirubin, and ALP. In addition to the analyses described in Section 6.2.2.2, this section describes specific analyses for this topic.

First, the number and percentage of patients with the following abnormal elevations in hepatic laboratory tests at any time will be summarized between treatment groups:

- The percentages of patients with an ALT measurement ≥3×, 5×, and 10× the central laboratory upper limit of normal (ULN) during the treatment period will be summarized for all patients with a postbaseline value and for subsets based on various baseline values.

  - The analysis of 3× ULN will contain 4 subsets: patients whose non-missing maximum baseline value is ≤1× ULN, patients whose maximum baseline is >1× ULN but <3× ULN, patients whose maximum baseline value is ≥3× ULN, and patients whose baseline values are missing.
  - The analysis of 5× ULN will contain 5 subsets: patients whose non-missing maximum baseline value is ≤1× ULN, patients whose maximum baseline is >1× ULN but <3× ULN, patients whose maximum baseline is ≥3× ULN but <5× ULN, patients whose maximum baseline value is ≥5× ULN, and patients whose baseline values are missing.
  - The analysis of 10× ULN will contain 6 subsets: patients whose non-missing maximum baseline value is ≤1× ULN, patients whose maximum baseline is >1× ULN but <3× ULN, patients whose maximum baseline is ≥3× ULN but <5× ULN, patients whose maximum baseline is ≥5× ULN but <10× ULN, patients whose maximum baseline value is ≥10× ULN, and patients whose baseline values are missing.
- The percentages of patients with an AST measurement ≥3×, 5×, and 10× the central laboratory ULN during the treatment period will be summarized for all patients with a postbaseline value and for subsets based on various levels of baseline. Analyses will be constructed as described above for ALT.
- The percentages of patients with a total bilirubin measurement ≥2× the central laboratory ULN during the treatment period will be summarized for all patients with a postbaseline value and subset into 4 subsets: patients whose non-missing maximum baseline value is ≤1× ULN, patients whose maximum baseline is >1× ULN but <2× ULN, patients whose maximum baseline value is ≥2× ULN, and patients whose baseline values are missing.
- The percentages of patients with an ALP measurement ≥1.5× the central laboratory ULN during the treatment period will be summarized for all patients with a postbaseline value and subset into 4 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 value is ≥1.5× ULN, and patients whose baseline values are missing.

Information collected from additional hepatic safety data collection forms will be provided in patient profiles.

Second, to further evaluate potential hepatotoxicity, an Evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot using maximum postbaseline ALT divided by ULN versus maximum postbaseline total bilirubin divided by ULN will be created that includes all patients from the safety populations for the studies included in the submission (any phase, any medication). Each patient with at least 1 postbaseline ALT and total bilirubin contributes 1 point to the plot. The measurements do not need to be taken at the same blood draw. Symbols will be used to indicate randomized treatment.

When criteria are met for hepatic evaluation and completion of the hepatic safety case report form (CRF), investigators are required to answer a list of questions (see the Compound level

safety standards). A listing of this collected information will be generated together with a graphical patient profile. This includes demographics, disposition, and a display of study drug exposure, AEs, medications, and the liver-related measurements over time will be provided for these patients and any additional patients meeting ALT or AST measurement greater than or equal to 5× ULN (on a single measurement) or ALP measurement greater than or equal to 2× ULN (on a single measurement).

6.13.5.2. Hematologic Changes

Hematologic changes will be defined based on clinical laboratory assessments. Common Terminology Criteria for Adverse Events (CTCAEs) will be applied for selected laboratory tests and are described in the Compound level safety standards. These CTCAE grading schemes are consistent with both Version 3.0 and Version 4.03 of the CTCAE guidelines (CTCAE 2003, 2010).

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. Planned and unplanned measurements will be included. Treatment-emergence will be characterized using the following 5 criteria (as appropriate to the grading scheme):

- any increase in postbaseline CTCAE grade from worst baseline grade
- increase to Grade 1 or above at worst postbaseline
- increase to Grade 2 or above at worst postbaseline
- increase to Grade 3 or above at worst postbaseline
- increase to Grade 4 at worst postbaseline

Shift tables will show the number and percentage of patients based on baseline to maximum during the treatment period, with baseline depicted by the most extreme 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 within the following categories:

- Decreased: postbaseline category < baseline category
- Increased: postbaseline category > baseline category
- Same: postbaseline category = baseline category

A laboratory-based treatment-emergent outcome related to increased platelet count will be summarized in similar fashion. Treatment-emergent thrombocytosis as a laboratory-based abnormality will be defined as an increase in platelet count from a maximum baseline value ≤600 billion/L to any postbaseline value >600 billion/L (Lengfelder et al. 1998). Planned and unplanned measurements will be included.

A listing of patients with treatment-emergent thrombocytosis may be provided for safety review.

6.13.5.3. Lipid Effects

Lipid effects will be assessed through analysis of elevated total cholesterol, elevated LDL cholesterol, decreased HDL cholesterol, and elevated triglycerides as described in Section 6.13.3 and with TEAEs potentially related to hyperlipidemia.

Categorical analyses will be performed using National Cholesterol Education Program (NCEP) Adult Treatment Panel (ATP) III guidelines (2002) as shown in the Compound level safety standards. The grade-like categories shown in this table are ordered from traditionally most desirable to least desirable for the purposes of these analyses.

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. With each shift table, a shift table summary displaying the number and percentage of patients with the least desirable postbaseline results will be presented by treatment group for each treatment period within the following categories:

- Decreased: postbaseline category more desirable than baseline category
- Increased: postbaseline category less desirable than baseline category
- Same: postbaseline category = baseline category

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 analysis period will be tabulated using the NCEP categories shown in the Compound level safety standards.

Treatment-emergent elevated total cholesterol will be characterized as follows:

- increase to categories ‘Borderline high’ or ‘High’
- increase to category ‘High’

Treatment-emergent elevated triglycerides will be characterized as

- increase to categories ‘Borderline high,’ ‘High,’ or ‘Very high’
- increase to categories ‘High’ or ‘Very high’
- increase to category ‘Very high’

Treatment-emergent elevated LDL cholesterol will be characterized as

- increase to categories ‘Borderline high,’ ‘High,’ or ‘Very high’
- increase to categories ‘High’ or ‘Very high’
- increase to ‘Very high’

Treatment-emergent abnormal HDL cholesterol will be characterized as

- decreased
    - decrease to categories ‘Normal’ or ‘Low’
    - decrease to category ‘Low’

- increased
    - increase to categories 'Normal' or 'High'
    - increase to category 'High'.

The percentages of patients with treatment-emergent potential hyperlipidemia will be summarized by treatment group, ordered by decreasing frequency in the baricitinib 4-mg group using a predefined MedDRA list of PTs that is a subset of the narrow scope PTs in the MedDRA SMQ 'Dyslipidemia' (code 200000026) [see Compound level safety standards].

6.13.5.4. 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 as shown in the Compound level safety standards. This CTCAE grading scheme is consistent with both Version 3.0 and Version 4.03 of the CTCAE guidelines.

Shift tables will show the number and percentage of patients based on baseline to maximum during the analysis period, with baseline depicted by highest grade during the baseline period. Treatment-emergent laboratory abnormalities related to elevated creatinine occurring at any time during the treatment period will be tabulated. Refer to the Compound level safety standards for details.

6.13.5.5. Elevations in Creatine Phosphokinase

Elevations in CPK will be addressed using CTCAE criteria as shown in the Compound level safety standards. This CTCAE grading scheme is consistent with both Version 3.0 and Version 4.03 of the CTCAE guidelines. Refer to the Compound level safety standards for details. A listing of elevated CPK (CTCAE Grade of 3 or above) will be provided for medical safety review.

Treatment-emergent adverse events potentially related to muscle symptoms may be analyzed based on reported AEs. The Muscle Symptoms special search category is a predefined MedDRA search criteria list that contains the narrow scope terms from the Rhabdomyolysis/myopathy SMQ (code 20000002) plus selected terms from the Musculoskeletal SOC. These terms are shown in the Compound level safety standards.

6.13.5.6. Infections

Infections will be defined using all the PTs from the Infections and Infestations SOC as defined in MedDRA. Serious infection will be defined as all the infections that meet the SAE criteria.

The number and percentage of patients with TEAEs of infections, serious infections, and infections resulting in permanent study drug discontinuation will be summarized by treatment group using MedDRA PTs. The proportion of patients developing skin infections requiring antibiotic treatment by Week 16 and end of study will also be summarized on the overview of infections table.

The number and percentage of patients with TEAEs of infections by maximum severity will be summarized by treatment group using MedDRA PTs.

The IR and 95% CI will be calculated for the overall observation time for infections of special interest (serious infections, treatment-emergent herpes zoster, treatment-emergent tuberculosis, treatment-emergent opportunistic infections) for the final analysis.

Treatment-emergent infections will be reviewed in context of other clinical and laboratory parameters via a listing (details see Compound level safety standards).

The TEAE infections will be further analyzed in terms of potential opportunistic infection, herpes zoster, and herpes simplex. Summary of hepatitis B virus (HBV) deoxyribonucleic acid (DNA) monitoring results and association between infection and neutropenia/lymphopenia will also be provided in the context of infections.

Opportunistic Infection

To identify potential opportunistic infections (POIs), the following approach will be used:

- identifying the POIs using a list of MedDRA PTs (refer to Compound level safety standards)

Potential opportunistic infections identified through this approach may be combined in one list for medical review and final classification of whether the case met the modified Winthrop (2015) definitions for opportunistic infection (OI).

A final listing for OIs will be provided for the CSR and to assist in composition of patient narratives.

Herpes Zoster

Cases of herpes zoster will be further classified as follows:

- localized or non-multidermatomal involvement of the primary and/or adjacent dermatomes only
    - complicated – documented ocular (cornea or deeper structure; eg, iritis, keratitis, retinitis, etc.) or motor nerve involvement (eg, palsy; postherpetic neuralgia does not meet criteria for motor nerve involvement)
    - uncomplicated-localized or non-multidermatomal cases that are not complicated
- multidermatomal – involvement beyond primary and adjacent dermatomes (that is, >3 contiguous dermatomes) or involvement of 2 or more non-contiguous dermatomes
    - complicated-documented ocular (cornea or deeper structure; for example, iritis, keratitis, retinitis) or motor nerve involvement
    - uncomplicated-multidermatomal cases
- disseminated-systemic infection, visceral or widespread cutaneous (for example, ≥5 dermatomes or 3 to 4 dermatomes including at least 1 non-contiguous dermatome)
- Recurrent – >1 infection occurring in an individual patient during the course of participation in the baricitinib clinical program

All cases of herpes zoster will undergo medical review to determine the classification as described above.

A summary of herpes zoster table will be provided. The summary table will also include event maximum severity, seriousness, whether resulting in temporary study drug interruption, whether resulting in study drug discontinuation, whether treated with antiviral medication, and event outcome. Of note, in the context of herpes zoster, antiviral treatment is defined as that the follows: the medication was initiated at the event start date or within 30 days before or after the event start date. The antiviral medication for herpes zoster includes but is not limited to aciclovir, brivudine, cidofovir, famciclovir, foscarnet, ganciclovir, penciclovir, valaciclovir, valganciclovir, vidarabine (best presented by J05AB, J05AC, J05AE, and J05AH ATC codes). Medical representatives will review the concomitant medication list prior to the database lock and make adjustment of the above list if necessary.

If a patient has more than 1 event of herpes zoster, the event with the maximum severity will be used in these summary tables. If more than 1 event of herpes zoster occurs with the same severity, the event with the longest duration will be used in the summary table.

Herpes Simplex

A summary analysis of herpes simplex will be provided. Herpes simplex will be defined based on MedDRA PT as listed in Compound level safety standards (both narrow and broad terms in the herpes simplex section). The list needs to be reviewed by GPS/medical prior to data locks (final and interim). The summary table will include event maximum severity, seriousness, whether resulting in temporary study drug interruption, whether resulting in study drug discontinuation, and whether treated with antiviral medication.

If a patient has more than 1 event of herpes simplex, the event with the maximum severity will be used in these summary tables. If more than 1 event of herpes simplex occurs with the same severity, the event with the longest duration will be used in the summary table.

Skin Infections

A summary analysis of skin infections may be provided. Skin infections may be defined based on MedDRA preferred term as listed in Compound level safety standards.

HBV DNA

A listing of patients with detectable HBV DNA postbaseline will be provided.

HBV DNA status post baseline (not detectable, detectable but not quantifiable [that is, < lower limit of detection (LLOD)], quantifiable [that is, ≥LLOD]) will be summarized by treatment group stratified by baseline HBV serology status, specifically:

1. HBsAb+ / HBcAb+
2. HBsAb- / HBcAb+

Association Between Infection and Neutropenia/Lymphopenia

Depending on the number of cases with CTCAE Grade 2 or greater, a summary table may be provided for treatment-emergent infections that were preceded or accompanied by neutropenia. For this analysis, neutropenia is defined as (1) CTCAE Grade 2 or greater or (2) CTCAE Grade 3

or greater. Infection events with onset date ≤14 days before or after the Grade 2 neutrophil count collection date will be considered as infections preceded or accompanied by neutropenia.

Similar analyses as above will be conducted to evaluate the association between infection and lymphopenia.

6.13.5.7. Major Adverse Cardiovascular Events and Other Cardiovascular Events

Potential major adverse cardiovascular events (MACE) and other cardiovascular events requiring adjudication will be analyzed.

Categories and subcategories analyzed will include, but are not limited to the following:

- MACE
  - Cardiovascular death
  - Myocardial infarction (MI)
  - Stroke
- Other cardiovascular events
  - Transient ischemic attack
  - Hospitalization for unstable angina
  - Hospitalization for heart failure
  - Serious arrhythmia
  - Resuscitated sudden death
  - Cardiogenic shock
  - Coronary interventions (such as coronary artery bypass surgery or percutaneous coronary intervention)
- Non-cardiovascular death
- All-cause death

In general, events requiring adjudication are documented by investigative sites using an endpoint reporting CRF. This CRF is then sent to the adjudication center which uses an adjudication reporting CRF to document the final assessment of the event as a MACE, as some other cardiovascular event, or as no event (according to the Clinical Endpoint Committee Charter). In some cases, however, the investigator may not have deemed that an event had met the endpoint criteria, but the event was still sent for adjudication as a potential MACE, other cardiovascular event, or no event. These events are included in the adjudication process to ensure adequate sensitivity. In these instances, the adjudication reporting CRF will not have a matching endpoint reporting CRF from the investigator. Events generated from these circumstances will be considered as events sent for adjudication in the absence of an investigator's endpoint reporting form.

The number and percentage of patients with MACE, other cardiovascular events, non-cardiovascular death, and all-cause death, as positively adjudicated, will be summarized by treatment group based on the categories and subcategories above.

A listing of the 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.

6.13.5.8. Venous and Pulmonary Artery Thromboembolic Events

Events identified as representative of venous thromboembolic event (VTE) disease will be further classified as deep vein thrombosis (DVT), pulmonary embolism (PE), or other peripheral venous thrombosis and will be analyzed. The following definitions apply:

- DVT: Clinical diagnosis of a thrombosis in a deep vein above the knee that must be confirmed by objective evidence of either: a filling defect of deep veins of the leg on venography or a non-compressible venous segment on ultrasound or confirmation by other imaging modality (for example, computed tomography [CT], magnetic resonance imaging [MRI])
- PE: Clinical diagnosis of pulmonary embolus that must be confirmed by objective evidence of either: a filling defect of pulmonary arteries by either pulmonary angiography or CT angiography or by a high probability ventilation perfusion (VQ) scan
- Other Peripheral Venous Thrombosis: Clinical diagnosis of a venous thrombosis not specified by either DVT or PE above. Other peripheral venous thrombosis disease must be confirmed by objective evidence by imaging including venography, ultrasound, CT scan, or MRI. Examples of these would include non-superficial below knee thrombosis, portal vein, subclavian vein, or mesenteric vein. Superficial thrombophlebitis alone is not considered a VTE event.

In general, events requiring adjudication are documented by investigative sites using an endpoint reporting CRF. Refer to Section 6.13.5.7 for more details, as the process is the same as that for MACE.

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. Note that the below knee thrombosis captured in the 'other peripheral venous thrombosis' category will be summarized within DVT.

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.

6.13.5.9. Arterial Thromboembolic Events

Refer to the Compound level safety standards.

6.13.5.10. Malignancies

Malignancies will be identified using terms from the malignant tumors SMQ (SMQ 20000194). Malignancies excluding non-melanoma skin cancers (NMSC) and NMSC will be reported separately.

All the cases identified by malignant tumors SMQ will be assessed through medical review to determine confirmed NMSC cases.

First, a listing including all the malignancy cases will be prepared before database lock, along with the *planned* NMSC flag according to the current MedDRA version PTs (the list will be updated depending on the MedDRA version used for analysis):

- Squamous cell carcinoma of skin (10041834)
- Bowen's disease (10006059)
- Basal cell carcinoma (10004146)
- Basosquamous carcinoma (10004178)
- Basosquamous carcinoma of skin (10004179)
- Squamous cell carcinoma (10041823)
- Skin squamous cell carcinoma metastatic (10077314)
- Skin cancer (10040808)
- Carcinoma in situ of skin (10007390)
- Keratoacanthoma (10023347)
- Vulvar squamous cell hyperplasia (10079905)
- Skin squamous cell carcinoma recurrent (10081136)

This internal review is to occur prior to database lock. The case review and subsequent summary analyses will include all the cases reported in the study database or by Lilly Safety System report, disregarding the length of gap between the last treatment dose date and the event date. The NMSC flag will be confirmed by the medical team during the internal review process.

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

6.13.5.11. Allergic Reactions/Hypersensitivities

A search will be performed using the current MedDRA version SMQs to search for relevant events, using the following queries:

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

Assessment of the Anaphylactic reaction SMQ includes an algorithmic query. An algorithmic case comprises one or more events associated with an individual administration of study drug, where the events include:

- A narrow term from the SMQ (Category A of the SMQ);
- Multiple terms from the SMQ, comprising terms from at least 2 of the following categories from the SMQ:
    - Category B - (Upper Airway/Respiratory)
    - Category C - (Angioedema/Urticaria/Pruritus/Flush)
    - Category D - (Cardiovascular/Hypotension)

In the present studies, where study drug is administered daily, events will be considered as associated with an individual administration of study drug when the events are reported on the same study day.

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, algorithmic, or broad), reported AE term, AE onset and end dates, severity, seriousness, outcome, etc.

In addition, summary tables will be provided. Refer to the Compound safety level standards for details.

6.13.5.12. Gastrointestinal Perforations

Treatment-emergent adverse events related to potential gastrointestinal (GI) perforations will be analyzed using reported AEs. Identification of these events will be based on review of the PTs of the MedDRA SMQ 20000107, GI perforations (note that this SMQ holds only narrow terms and has no broad terms). Potential GI perforations identified by the above 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 GI perforation. A summary table based on medical review may be provided and treatment comparisons will be made using Fisher's exact test.

6.13.5.13. Columbia-Suicide Severity Rating Scale

Suicidal ideation, suicidal behavior, and self-injurious behavior without suicidal intent, based on the C-SSRS, will be listed by patient and visit. Only patients that show suicidal ideation/behavior or self-injurious behavior without suicidal intent during treatment will be displayed along with all their ideation and behavior, even if not positive (that is, if a patient's answers are all 'no' for the C-SSRS, then that patient will not be displayed). A summary of the C-SSRS categories during treatment and a shift summary in the C-SSRS categories from baseline during treatment may be provided.

6.13.5.13.1. 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.

6.14. Subgroup Analyses

In Period 2, subgroup analyses comparing each dose of baricitinib to placebo will be performed on the ITT population **at Week 16**, with data up to rescue (primary censoring rule) for the following:

- Proportion of patients achieving IGA 0 or 1 with a 2-point improvement
- Proportion of patients achieving EASI75 Response Rate
- Proportion of patients achieving Itch NRS 4-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 (<65 years, ≥65 years)
  - Age group (<65 years, ≥65 years to <75 years, ≥75 years to <85 years, ≥85 years)
  - Baseline weight: (<60 kg, ≥60 kg to <100 kg, ≥100 kg)
  - Baseline BMI (<25 kg/m^2, ≥25 kg/m^2 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)
  - Ethnicity (Hispanic, non-Hispanic)
  - Baseline renal function status: impaired (eGFR <60 mL/min/1.73 m^2) or not impaired (eGFR ≥60 mL/min/1.73 m^2)

- Geographic Region Subgroups:
  - Region (as defined in)
  - Specific regions (Europe, other)
  - Specific country (Japan, other)

- Previous and Concomitant Therapy Subgroups:
  - Prior use of TCNI (yes, no)
  - Prior systemic therapy use (yes, no)

- Baseline Disease-Related Characteristics Subgroup:
  - Baseline disease severity (IGA score: 3, 4)

Descriptive statistics will be provided for each treatment and stratum of a subgroup as outlined, regardless of sample size.

Subgroup analyses for categorical outcomes will be performed via logistic regression, using Firth's correction to accommodate (potential) sparse response rates. The model will include the categorical outcome as the dependent variable and baseline value (for EASI and itch), treatment, subgroup, and treatment-by-subgroup interaction as explanatory variables. Missing data will be imputed using NRI (Section 6.3.1). The treatment-by-subgroup interaction will be tested at the 0.1 significance level. 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% confidence interval (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.

In case any level of a subgroup comprises <10% of the overall sample size, only descriptive summary statistics will be provided for treatment arms, and no treatment group comparisons will be performed within these subgroup levels.

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

6.15. 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 the important protocol deviations (IPDs) identified, a subset occurring during Treatment Period 2 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.

The Trial Issue Management Plan includes the categories and subcategories of IPDs and whether these deviations will result in the exclusion of patients from the PPS.

The number and percentage of patients having IPD(s) will be summarized within category and subcategory of deviation by treatment group for Treatment Period 2 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.

6.16. Interim Analyses and Data Monitoring

An interim analysis will be conducted at when the last patient completes Visit 10 (Week 24) or ETV.

The baricitinib AD, alopecia areata (AA), and systemic lupus erythematosus (SLE) Phase 3 programs Data Monitoring Committee (DMC) is an independent expert advisory group commissioned and charged with the responsibility of evaluating cumulative safety at regular intervals. As such, the primary objective of the DMC is to monitor the safety of the subjects enrolled in the baricitinib AD, AA, and SLE Phase 3 programs by reviewing the available clinical data at scheduled time points, as described in the DMC Charter, as well as on an ad hoc basis, as needed. 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 dermatology, statistics, cardiology, 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, and vital sign data. The DMC may recommend continuation of the study, as designed;

temporary suspension of enrollment; or the discontinuation of a particular 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 Baricitinib Atopic Dermatitis DMC charter, and further details are given in the Interim Analysis Plan in Section 6.16.1.

Besides DMC members, a limited number of preidentified individuals may gain access to the limited unblinded data, as specified in the unblinding plan, prior to the interim or final database lock, for preparation of regulatory documents. 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.

6.16.1. Interim Analysis Plan

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

- patient disposition, demographics, and baseline characteristics
- exposure (if request)
- AEs, to include the following:
  - TEAEs
  - SAEs, including deaths
  - selected special safety topics
- clinical laboratory results
- vital signs
- C-SSRS

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

The DMC may request efficacy data if they feel there is value and to confirm a reasonable benefit/risk profile for ongoing patients in the studies. If efficacy data is requested, it will be mean change from baseline of EASI score. Further details are given in the DMC charter.

6.17. Planned Exploratory Analyses

The planned exploratory objectives of this study are documented in Section 4.3 and Section 6.10.3. Details of the analysis methods are documented in Table JAIN.6.4. Any information not provided in Table JAIN.6.4 will be documented in a supplementary SAP or a supplementary list of analyses.

6.18. Annual Report Analyses

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

6.19. 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 and 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 (eg, CSR, manuscripts).

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

7. Unblinding Plan

Refer to the 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.

Basra MK, Salek MS, Camilleri L, Sturkey R, Finlay AY. Determining the minimal clinically important difference and responsiveness of the Dermatology Life Quality Index (DLQI): further data. *Dermatology*. 2015;230(1):27-33.

Bożek A, Reich, A. Assessment of intra-and inter-rater reliability of three methods for measuring atopic dermatitis severity: EASI, objective SCORAD, and IGA. *Dermatology*. 2017;233(1):16-22.

Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ*. 1992;305(6846):160-164.

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;140:1513-1519.

[CTCAE] Common Terminology Criteria for Adverse Events, Cancer Therapy Evaluation Program, Version 3.0, DCTD, NCI, NIH, DHHS, March 31, 2003. Available at: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcaev3.pdf. Accessed 25 July 2017.

[CTCAE] Common Terminology Criteria for Adverse Events, Version 4.03, DHHS NIH NCI, NIH Publication No. 09-5410, June 14, 2010. Available at: https://www.eortc.be/services/doc/ctc/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf. Accessed 25 July 2017.

EuroQol Group. EQ-5D-5L User Guide. Version 2.1. Available at: http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-5L_UserGuide_2015.pdf. Accessed March 2018.

Futamura M, Leshem YA, Thomas KS, Nankervis H, Williams HC, Simpson EL. A systematic review of Investigator Global Assessment (IGA) in atopic dermatitis (AD) trials: many options, no standards. *J Am Acad Dermatol*. 2016;74(2):288-294.

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.

Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-1736.

Hongbo Y, Thomas CL, Harrison MA, Salek MS, Finlay AY. Translating the science of quality of life into practice: What do dermatology life quality index scores mean? *J Invest Dermatol*. 2005;125(4):659-664.

[ICH] International Conference on Harmonisation. Harmonised Tripartite Guideline: Clinical safety data management: definitions and standards for expedited reporting. E2A. 1994; Step 4. Available at: http://www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html. Accessed 21 July 2017.

[ICH] International Conference on Harmonisation. Statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials. E9(R1). 2017. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analysis-clinical. Accessed 12 December 2019.

Khilji FA, Gonzalez M, Finlay AY. Clinical meaning of change in Dermatology Life Quality Index scores. *Br J Dermatol*. 2002;147(suppl 62):25-54.

Kimball AB, Naegeli AN, Edson-Heredia E, Lin CY, Gaich C, Nikai E, 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 A, Labrèze, Stadler 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.

Langley RG, Feldman SR, Nyirady J, van de Kerkhof P, Papavassilis C. The 5-point Investigator's Global Assessment (IGA) scale: a modified tool for evaluating plaque psoriasis severity in clinical trials. *J Dermatolog Treat*. 2015;26(1):23-31.

Lengfelder E, Hochhaus A, Kronawitter U, Hoche D, Queisser W, Jahn-Eder M, Burkhardt R, Reiter A, Ansari H, Hehlmann R. Should a platelet limit of 600 x 10(9)/l be used as a diagnostic criterion in essential thrombocythaemia? An analysis of the natural course including early stages. *Br J Haematol*. 1998;100(1):15-23.

Maruish ME. User's manual for the SF-36v2 Health Survey (3rd ed.). Lincoln, RI: QualityMetric Incorporated. 2011.

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.

[NCEP] National Cholesterol Education Program. Third Report of the NCEP Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel [ATP] III) Final Report, National Cholesterol Education Program, National Heart, Lung, and Blood Institute, National Institutes of Health, NIH Publication No. 02-5215, September 2002. *Circulation*. 2002;106(25):3143-421.

Program Safety Analysis Plan: Baricitinib (LY3009104) Version 6. Report on file, Eli Lilly and Company.

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 MMG, 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.

Snaith RP. The Hospital Anxiety and Depression Scale. *Health and Quality of Life Outcomes*. 2003;1:29.

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.

US National Archives and Records Administration. Code of Federal Regulations (CFR). Investigational New Drug Application (IND) safety reporting. 2010;Title 21:Section 312.32. Available at: https://www.ecfr.gov/cgi-bin/text-idx?SID=fd87d5f6fb00b95672dbe924a6d3b2f0&mc=true&node=se21.5.312_132&rgn=div8. Accessed 21 July 2017.

Ware JE Jr. SF-36 health survey update. *Spine (Phila Pa 1976)*. 2000;25(24):3130-3139.

Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36), I: conceptual framework and item selection. *Med Care*. 1992;30(6):473-483.

White D, Leach C, Sims R, Atkinson M, Cottrell D. Validation of the Hospital Anxiety and Depression Scale for use with adolescents. *Br J Psychiatry*. 1999;175(5):452-454.

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.

Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand*. 1983;67(6):361-370.

Leo Document ID = c9e609fc-3ef2-4f28-92b2-572181818e90

Approver: PPD

Approval Date & Time: 19-Dec-2019 21:47:51 GMT

Signature meaning: Approved