Protocol I4V-MC-JAHN(c)
A Phase 3 Multicenter, Double-Blind Study to Evaluate the Long-Term Safety and Efficacy of Baricitinib in Adult Patients with Atopic Dermatitis

BREEZE-AD3

EudraCT Number: 2017-000873-35

Baricitinib (LY3009104)

A 4-year Phase 3 multicenter, double-blind long-term extension study to evaluate the safety and efficacy of placebo, baricitinib 1-mg, baricitinib 2-mg, and baricitinib 4-mg in adult patients with atopic dermatitis, including a blinded randomized treatment withdrawal and randomized downtitration.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol Electronically Signed and Approved by Lilly: 25 Jul 2017
Amendment (a) Electronically Signed and Approved by Lilly: 15 Dec 2017
Amendment (b) Electronically Signed and Approved by Lilly: 29 Mar 2019
Amendment (c) Electronically Signed and Approved by Lilly on approval date provided below

Approval Date: 27-Nov-2019 GMT

Table of Contents

Section **Page**

List of Tables

Table **Page**

List of Figures

Figure **Page**

List of Appendices

1. Synopsis

Title of Study:
A Phase 3 Multicenter, Double-Blind Study to Evaluate the Long-Term Safety and Efficacy of Baricitinib in Adult Patients with Atopic Dermatitis

Rationale:
Atopic dermatitis (AD) is a pruritic, chronic or chronically relapsing, highly symptomatic inflammatory skin disease characterized by excessive T cell activation leading to significant skin infiltration by T cells and dendritic cells (Bieber 2010). Presentation is varied, but includes skin manifestations and pruritus, with associated sleep disturbances. The course of disease includes relapses of varying duration and severity.

Baricitinib is an oral Janus kinase (JAK) inhibitor with potency and selectivity for JAK1 and JAK2 and less potency for JAK3 or tyrosine kinase 2 (TYK2) (Fridman et al. 2010). The pathogenesis of AD is thought to be modulated through thymic stromal lymphopoietin, interleukin (IL)-13, IL-5, IL-4, IL-22, and IL-31, many of which activate receptors with downstream signaling through intracellular JAK1/JAK2/TYK2 (Nomura and Kabashima 2015). This activity profile suggests that baricitinib may inhibit cytokines involved in AD pathogenesis.

Clinical studies have established that baricitinib is effective in autoimmune and autoinflammatory diseases involving the joints, kidneys, and skin. Baricitinib was effective in reducing swollen and tender joints in patients with rheumatoid arthritis (Genovese et al. 2016; Dougados et al. 2017; Fleischmann et al. 2017; Taylor et al. 2017), was effective in reducing disease severity in patients with moderate to severe plaque psoriasis (Papp et al. 2016), and was effective in reducing the urinary albumin-to-creatinine ratio in patients with diabetic kidney disease (Tuttle et al. 2015). In addition, in a recently completed Phase 2 study (I4V-MC-JAHG), baricitinib was effective in reducing disease severity in patients with moderate to severe AD. The mechanism of action, combined with demonstration of clinical benefit in inflammatory diseases involving joints, kidneys, and skin, provides the rationale for evaluating baricitinib in moderate to severe AD.

Objectives/Endpoints:

Objectives	Endpoints
Primary	
To estimate the effect of long-term therapy with baricitinib on responders and partial responders at entry of JAHN.	• Proportion of patients with a response of IGA 0 or 1 assessed at Weeks 16, 36, and 52
Secondary (Weeks 0–52)	
Baricitinib Patients at Entry to Study JAHN To evaluate the effect of increasing or maintaining baricitinib dose on clinical measures and patient-reported outcomes.	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at Weeks 16, 36, and 52 • Proportion of patients with a response of IGA 0 or 1 assessed at Weeks 16, 36, and 52 (nonresponders) • Proportion of patients achieving response of EASI75 from baseline of originating study assessed at Weeks 16, 36, 52 • Proportion of patients with a 4-point improvement from baseline of originating study in Itch NRS at 16 weeks

Objectives	Endpoints
Placebo Nonresponders at Entry to Study JAHN To evaluate the effect of starting baricitinib 2-mg versus 4-mg on clinical measures and patient-reported outcomes.	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at Weeks 4, 16, 24, 52 • Proportion of patients with a response of IGA 0 or 1 assessed at Weeks 4, 16, 24, 52 • Proportion of patients achieving response of EASI75 from baseline of originating study assessed at Weeks 4, 16, 24, 52 • Proportion of patients with a 4-point improvement from baseline of originating study in Itch NRS at 16 weeks
Secondary (Weeks 52–104)	
All Patients Entering the Substudy To evaluate the change in clinical response after treatment withdrawal or downtitration. **Patients Entering the Substudy with IGA 0 or 1** To evaluate the change in clinical response after treatment withdrawal or downtitration.	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at 16 weeks after rerandomization (Week 68) and Week 104 • Proportion of patients with a response of IGA 0 or 1 assessed at 16 weeks after rerandomization (Week 68) and Week 104 • Proportion of patients with a response of EASI75 from baseline of originating study assessed at 16 weeks after rerandomization (Week 68) and Week 104 • Time to retreatment (time to IGA ≥3)
Patients Retreated During Substudy To evaluate the ability to recapture efficacy based on clinical measures after experiencing a loss of treatment benefit:	• Proportion of patients with a response of IGA 0, 1, or 2 within 16 weeks of retreatment • Proportion of patients with a response of IGA 0 or 1 within 16 weeks of retreatment • Proportion of patients with a response of EASI75 from baseline of originating study within 16 weeks of retreatment
Patients Not Entered into the Substudy To evaluate the effect of maintaining baricitinib dose on clinical measures.	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at Week 104 • Proportion of patients with a response of IGA 0 or 1 assessed at Week 104 • Proportion of patients with a response of EASI75 from baseline of originating study assessed at Week 104

Abbreviations: EASI = Eczema Area and Severity Index; IGA = Investigator's Global Assessment; NRS = numeric rating scale.

Summary of Study Design:

Study I4V-MC-JAHN (JAHN) is a Phase 3, multicenter, double-blind study to evaluate the long-term safety and efficacy of baricitinib (1-mg once daily, 2-mg once daily, and 4-mg once daily) in adult patients with AD for up to 4 years. The study population will include patients aged 18 years or older who completed an originating study (such as I4V-MC-JAHL, I4V-MC-JAHM, or 14V-MC-JAIY) and were eligible for enrollment into JAHN. There is a single substudy included that will evaluate treatment withdrawal and dose downtitration.

Treatment Arms and Duration:

The treatment study duration will be up to 4 years. The study consists of 3 study periods and 1 substudy: randomized treatment withdrawal and downtitration.

Screening and Baseline Period: For most patients, the last visit of the originating study will be the first visit and screening/baseline period for Study JAHN. Patients still completing a washout from systemic therapies from originating study will have this period extended for a maximum of 8 weeks and have their baseline established after washout is completed. At baseline, patients will be classified into the following groups based on their response to treatment in the originating study: "Responders (Investigator's Global Assessment [IGA] 0 or 1 and never rescued) and Partial Responders (IGA 2 and never rescued)" or "Nonresponders (IGA 3 or 4 or rescued)."

Treatment Period: The full treatment period will last from Week 0/Visit 1 through Week 200/Visit 22. Patients will continue using emollients daily and topical corticosteroid (TCS) use will be permitted at the investigator's discretion and provided automatically at the time of rescue or retreatment.

Treatment Period 1: Week 0 (Visit 1) up to Week 52 (Visit 8):

At Week 0/Visit 1:

- Nonresponders receiving placebo, baricitinib 1-mg, or baricitinib 2-mg in the originating study will be rerandomized in a blinded fashion 1:1 to either baricitinib 2-mg or baricitinib 4-mg. Nonresponders receiving baricitinib 4-mg will remain on baricitinib 4-mg.
- Responders and Partial Responders receiving placebo, baricitinib 1-mg, 2-mg, or 4-mg at the completion of the originating study will remain on their assigned treatment. If a patient has an IGA ≥3 during treatment period 1, then interactive web-response system will assign the following treatments: placebo and baricitinib 1-mg will be rerandomized in a blinded fashion 1:1 to either baricitinib 2-mg or baricitinib 4-mg; baricitinib 2-mg and 4-mg will remain on the same dose.

Treatment Period 2: Week 52 (Visit 8) through Week 200 (Visit 22):

At Week 52/Visit 8: All patients will be assessed for eligibility for the randomized withdrawal and downtitration substudy.

- Patients with an IGA 0, 1, or 2 at Week 52 who were assigned to baricitinib 4-mg or 2-mg upon enrollment to JAHN, who are currently receiving investigational product (IP), and who have not used a high-potency TCS for the last 14 days will be enrolled into the substudy. Patients receiving baricitinib 4-mg will be rerandomized 1:1:1 to either placebo, baricitinib 2-mg, or baricitinib 4-mg. Patients receiving baricitinib 2-mg will be rerandomized 1:1:1 to either placebo, baricitinib 1-mg, or baricitinib 2-mg. If a patient has an IGA ≥3 during this treatment period, they will be retreated with their presubstudy baricitinib dose.
- All other patients are not eligible for the randomized withdrawal and downtitration substudy, and will continue on their assigned treatment.

Post-Treatment Follow-Up Period: This period spans from the last treatment visit at Week 200/Visit 22, or early termination visit to approximately 28 days following the last dose of IP.

Number of patients:

Based on the above-mentioned originating studies, Study JAHN will enroll approximately 1425 patients.

Statistical Analysis:

Unless otherwise specified, the efficacy and health outcome analyses will be conducted on the modified intent-to-treat population and safety analyses will be conducted on the safety population.

Primary and secondary discrete efficacy variables will be descriptively summarized by treatment group in terms of frequencies and percentages. Treatment comparisons of discrete efficacy variables between treatment groups may be made using a logistic regression analysis with disease severity and treatment group in the model. The percentages, difference in percentages, and 95% confidence interval (CI) of the difference in percentages would then be reported. All patients who discontinue the study or the study treatment at any time for any reason or are rescued will be defined as nonresponders for the nonresponder imputation analysis for categorical variables after discontinuation or rescue and onward.

Continuous efficacy variables will be descriptively summarized by treatment group in terms of number of patients, mean, standard deviation, median, minimum, and maximum. Treatment comparisons of continuous efficacy and health outcome variables may be made using a restricted maximum likelihood-based mixed model for repeated measures. The model will include treatment, baseline severity, visit, and treatment-by-visit-interaction as fixed categorical effects and baseline score and baseline score-by-visit-interaction as fixed continuous effects. An unstructured (co)variance structure will be used to model the between- and within-patient errors. Type III sums of squares for the least squares means (LSM) will be used for the statistical comparison; 95% CI will also be reported.

Treatment comparisons of continuous efficacy and health outcome variables may also be made using analysis of covariance (ANCOVA) with disease severity, treatment group, and baseline value in the model. Type III tests for LSM may be used for statistical comparison between treatment groups. The LSM difference, standard error, p-value, and 95% CI may also be reported.

All safety data will be descriptively summarized by treatment group and treatment period. For categorical events, Fisher exact test may be used to perform comparisons between each baricitinib dose and the placebo group. Fisher exact test may also be used for all discontinuation, and other categorical safety data for between-treatment group comparisons. Continuous vital signs, body weight, and other continuous safety variables including laboratory variables will be analyzed by ANCOVA with treatment and baseline value in the model.

2. Schedule of Activities

Table JAHN.1. I4V-MC-JAHN Schedule of Activities

	Treatment Period 1							Treatment Period 2															PTFU Period
	Screening and Baseline Period																						
Visit number	**1[a]**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22/ ET[s]**	**801[b]**
Weeks from entry into JAHN	**0**	**4**	**8**	**16**	**24**	**36**	**48**	**52**	**56**	**60**	**64**	**68**	**76**	**84**	**92**	**104**	**120**	**136**	**152**	**168**	**184**	**200**	**204**
Visit tolerance interval (days) from entry into JAHN	0 to 56 from last visit of originating study[c]	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±7	±7	±7	±7	±7	±7	28 ± 4 after last dose
Procedures																							
Inclusion and exclusion criteria review	X[d]																						
Informed consent	X[e]																						
Abbreviated demographics	X																						
Clinical Assessments																							
Weight	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Vital signs (BP and pulse)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Symptom-directed physical examination[f]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Adverse events	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Concomitant medication	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ePRO (patient diary) dispensed	X	X	X				X	X[g]	X[g]	X[g]	X[g]												
ePRO (patient diary) returned	X	X	X	X				X	X[g]	X[g]	X[g]	X[g]										X[h]	
Rerandomization[i, j]	X							X															
IWRS	X[j]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
IP dispensed[j]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
IP returned and compliance assessed		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Dispense TCS (as needed)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X								

	Treatment Period 1							Treatment Period 2															PTFU Period
	Screening and Baseline Period																						
Visit number	**1[a]**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22/ ET[s]**	**801[b]**
Weeks from entry into JAHN	**0**	**4**	**8**	**16**	**24**	**36**	**48**	**52**	**56**	**60**	**64**	**68**	**76**	**84**	**92**	**104**	**120**	**136**	**152**	**168**	**184**	**200**	**204**
Visit tolerance interval (days) from entry into JAHN	0 to 56 from last visit of originating study[c]	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±7	±7	±7	±7	±7	±7	28 ± 4 after last dose
Weigh (tube with cap) and record returned TCS (as needed)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X						X[s]	X[t]
Physician-Assessed Efficacy Measures																							
IGA	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
EASI	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
SCORAD	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Health Outcomes Measures and Other Questionnaires[k]																							
POEM	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
DLQI	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
WPAI-AD	X	X		X	X	X	X	X	X			X	X	X	X	X						X[s]	X[t]
EQ-5D-5L	X	X		X	X	X	X	X	X			X	X	X	X	X						X[s]	X[t]
Itch NRS	X	X	X	X			X	X	X	X	X	X										X[h]	
Skin Pain NRS	X	X	X	X			X	X	X	X	X	X										X[h]	
ADSS	X	X	X	X			X	X	X	X	X	X										X[h]	
PGI-S-AD	X	X	X	X			X	X	X	X	X	X										X[h]	
HADS	X	X	X	X	X	X	X	X	X		X		X	X	X	X						X[s]	X[t]
C-SSRS[l] and Self-Harm Supplement Form	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Self-Harm Follow-up Form[m]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Laboratory Assessments																							
Clinical chemistry[n]	X[o]	X	X	X	X	X		X			X		X		X	X	X	X	X	X	X	X	X
Hematology	X[o]	X	X	X	X	X		X			X		X		X	X	X	X	X	X	X	X	X
Lipids (fasting)[p]	X[o]			X		X		X			X		X			X	X	X	X	X	X	X	
Urinalysis	X[o]	X		X		X		X			X		X		X	X						X[s]	X[t]
HBV DNA[q]	X[o]			X		X		X			X		X		X	X	X	X	X	X	X	X	X
Urine pregnancy[r]	X[o]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

	Treatment Period 1							Treatment Period 2															PTFU Period
	Screening and Baseline Period																						
Visit number	**1[a]**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22/ ET[s]**	**801[b]**
Weeks from entry into JAHN	**0**	**4**	**8**	**16**	**24**	**36**	**48**	**52**	**56**	**60**	**64**	**68**	**76**	**84**	**92**	**104**	**120**	**136**	**152**	**168**	**184**	**200**	**204**
Visit tolerance interval (days) from entry into JAHN	0 to 56 from last visit of originating study[c]	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±7	±7	±7	±7	±7	±7	28 ± 4 after last dose
Serum immunoglobulin (IgE)				X		X		X								X						X[s]	
Stored serum and plasma samples for exploratory analysis				X		X		X								X						X[s]	
Stored blood for RNA analysis				X		X		X								X						X[s]	

Abbreviations: ADSS = Atopic Dermatitis Sleep Scale; BP = blood pressure; C-SSRS = Columbia–Suicide Severity Rating Scale; DLQI = Dermatology Life Quality Index; EASI = Eczema Area and Severity Index; eGFR = estimated glomerular filtration rate; ePRO = electronic patient-reported outcome (device); EQ-5D-5L = the European Quality of Life–5 Dimensions–5 Levels; ET = early termination; HADS = Hospital Anxiety Depression Scale; HBcAb = anti-hepatitis B core antibody; HBsAb = hepatitis B surface antibody; HBV = hepatitis B virus; IGA = Investigator's Global Assessment; IP = investigational product; IWRS = interactive web-response system; NRS = numeric rating scale; PGI-S-AD = Patient Global Impression of Severity – Atopic Dermatitis; POEM = Patient-Oriented Eczema Measure; PTFU = post-treatment follow-up; RNA = ribonucleic acid; SCORAD = SCORing Atopic Dermatitis; TCS = topical corticosteroids; WPAI-AD = Work Productivity and Activity Impairment Questionnaire – Atopic Dermatitis.

[a] For the majority of patients, Visit 1 will also be the last visit of the originating study; thus any assessments/procedures conducted during the final visit of the patients' originating study should not be repeated during their first visit for Study JAHN.

[b] Patients who have discontinued IP but remain in the study for more than 28 days without IP can combine their Visit 22/ET with Visit 801 (follow-up visit).

[c] For patients not requiring washout, the visit tolerance interval should be 0 days for Visit 1; however, those patients returning after completion of a 4-week washout from systemic therapy have a tolerance interval of ≤56 days (8 weeks) from last visit of originating study. When a patient returns after completion of a systemic washout, all Visit 1 procedures should be repeated unless indicated otherwise.

[d] For patients requiring a washout from systemic therapy, the review of inclusion and exclusion criteria should be performed at both the last visit of the originating study and at the time when the patient returns to complete Visit 1 procedures.

[e] For patients requiring a washout from systemic therapy, informed consent should be signed at the last visit of the originating study.

[f] If necessary, a symptom-directed physical examination should be performed prior to enrollment to evaluate any condition that would meet the exclusion criteria. The symptom-directed physical examination may be repeated at the investigator's discretion any time a patient presents with physical complaints.

[g] Patient diaries are not required for patients known by the investigator to not meet substudy eligibility criteria (IGA score ≥3, patients receiving high-potency TCS and/or on a study drug interruption at Week 52 [Section 5.1.3.1]). Thus, patient diaries will not need to be dispensed/returned for visits indicated.

[h] Applies only if ET visit occurs while patient has diary.

[i] Rerandomization occurs only for nonresponders entering JAHN on placebo, baricitinib 1-mg, or baricitinib 2-mg.

[j] For patients requiring a washout from systemic therapy, IWRS registration of Visit 1, dispensing of IP, and rerandomization will not occur until washout is complete.

[k] The following measures (POEM, DLQI, EQ-5D-5L, WPAI-AD) should be completed prior to any clinical assessments being performed on days when study visits occur.

[l] Suicidal ideation and behavior subscales excerpt – adapted for the assessment of 11 preferred ideation and behavior categories.

[m] The Self-Harm Follow-up Form is only required if triggered by the Self-Harm Supplement Form.

[n] Clinical chemistry will include the following value calculated from serum creatinine: eGFR (calculated using the CKD-EPI Creatinine 2009 equation).

[o] If laboratory examinations have been performed ≤4 weeks from previous study visit, then they do not need to be repeated.

[p] Fasting lipid profile; patients should not eat or drink anything except water for 12 hours prior to sample collection. If a patient attends these visits in a nonfasting state, this will not be considered a protocol violation. Unscheduled lipid testing can be performed at the discretion of the investigator.

[q] HBV DNA will be performed per the schedule (see Section 9.4.7) in patients who tested positive for HBcAb at screening in the originating study, regardless of patient's HBsAb status.

[r] For all women of childbearing potential; if required per local regulations and/or institutional guidelines, pregnancy testing can occur at other times during the study treatment period.

[s] An ET visit should be conducted if a patient discontinues from the study before Week 200. Early termination visit activities do not need to be duplicated if occurring at the time of a scheduled visit. Weighing of TCS, collection of WPAI-AD, EQ-5D-5L, HADS, urinalysis, serum IgE exploratory storage samples, RNA and biomarker samples should only be performed at the ET visit if it occurs at or before Week 104.

[t] For V801, the weighing of TCS, and collection of WPAI-AD, EQ-5D-5L, HADS, and urinalysis should only be performed if it occurs at or before Week 104.

NOTE: Patients completing V16 and planning to sign ICF for amendment (c) can participate as long as they have not completed a V801.

3. Introduction

3.1. Background

Atopic dermatitis (AD), also known as eczema or atopic eczema, is a common, chronic, relapsing, highly symptomatic inflammatory skin disease (Bieber 2010). Patients with AD may present with skin lesions that can be acute with oozing, crusted, eroded vesicles or papules on erythematous plaques. Patients may also present with lesions that have a subacute appearance, with thick and excoriated plaques, or chronic appearance, with lichenified, slightly pigmented, excoriated plaques (Bieber 2010). Atopic dermatitis causes pruritus attacks throughout the day, which is the primary source of morbidity in this disorder (Simpson 2012). Pruritus often leads to an "itch–scratch" cycle, further compromising the epidermal barrier and resulting in dry skin, microbial colonization, and secondary infections (Krakowski et al. 2008), with 36% of patients reporting that they often or always scratch until their skin bleeds (Langenbruch et al. 2014). Pruritus from AD can worsen at night, resulting in sleep disturbances, with approximately 27% of adult patients with AD experiencing sleep disturbance as a result of itching (Langenbruch et al. 2014). In adult patients with moderate to severe AD, sleep quality and latency were significantly associated with poor quality of life (QoL) (Yano et al. 2013).

In clinical practice, AD is classified as mild, moderate, or severe based on a variety of clinical features, including severity of skin lesions and pruritus, and extent of disease (body surface area involved).

Until recently, there were no Food and Drug Administration (FDA)-approved systemic treatments for patients with moderate to severe AD, with the exception of systemic corticosteroids. In March 2017, Dupixent (dupilumab) injection, an IgG4 monoclonal antibody that inhibits interleukin (IL)-4 and IL-13, was approved by the FDA for this patient population. In the European Union, only cyclosporine has been approved for the treatment of patients with severe AD (Bieber and Straeter 2015). A recently completed Phase 2 study (I4V-MC-JAHG [JAHG]) evaluated the safety and efficacy of baricitinib (Janus kinase [JAK] inhibitor) in AD and results showed significant improvement in disease severity compared to placebo and no new safety concerns were identified.

In addition to AD, baricitinib has also been studied in Phase 3 in patients with rheumatoid arthritis (RA) and in Phase 2 in patients with diabetic nephropathy, moderate to severe psoriasis, and systemic lupus erythematosus.

Through 13 February 2019, baricitinib has been studied in more than 548 healthy volunteers and 6555 patients have received baricitinib in clinical studies. As of 13 February 2018, more than 2700 patients have been treated with baricitinib for more than a year and more than 1800 patients have been treated with baricitinib for more than 2 years at doses of 2-mg once daily or 4-mg once daily across the RA clinical program. Baricitinib has been administered as single doses ranging from 1- to 40-mg and as repeat oral doses ranging from 2- to 20-mg to healthy subjects. Baricitinib has also been administered to patients with RA at doses up to 15-mg daily for 4 weeks, 10-mg daily for 24 weeks, 8-mg daily for 76 weeks, and lower doses up to 4-mg daily for up to approximately 7 years.

3.2. Study Rationale

The underlying cause of AD is not completely understood. Loss-of-function mutations in the gene for *filaggrin* (filament aggregating protein), a key protein in terminal differentiation of the epidermis contributing to barrier function, has been identified as the strongest genetic risk factor for AD in European populations (Palmer et al. 2006). At a cellular level, AD is characterized by excessive T cell activation caused by genetic and environmental factors, leading to significant skin infiltration by T cells and dendritic cells. The cytokine thymic stromal lymphopoietin (TSLP) is thought to act as a master switch that triggers the initiation and maintenance of AD (Moniaga et al. 2013; Ziegler et al. 2013). Overexpression of TSLP in keratinocytes, the most prevalent cell type in the skin, triggers robust itch-evoked scratching and the development of an AD-like skin phenotype in mice (Li et al. 2005). In addition to directly inducing itch by activating sensory neurons in the skin, TSLP also enhances maturation and differentiation of dendritic cells and naive CD4+ T cells and induces production of Th2-related cytokines involved in AD pathogenesis (Wilson et al. 2013; Divekar and Kita 2015). Thymic stromal lymphopoietin and other key cytokines involved in AD pathogenesis, such as IL-13, IL-5, IL-22, and IL-31, signal through receptors associated with intracellular JAK1/JAK2/tyrosine kinase 2(TYK2) signaling (Ziegler et al. 2013; Nomura and Kabashima 2015).

Janus kinases are a family of tyrosine kinases that mediate cytokine receptor signaling through phosphorylation and activation of signal transducers and activators of transcription (STAT) proteins. There are 4 known JAK family members: JAK1, JAK2, JAK3, and TYK2 (Clark et al. 2014). The relative affinity of JAK inhibitors for different members of the JAK kinase family allows for differentiation of JAK inhibitors in relation to their enzymatic inhibitory profile. In vitro assays indicate that baricitinib is a selective inhibitor of JAKs with potency and selectivity for JAK1/JAK2 and less potency for JAK3 or TYK2 (Fridman et al. 2010). The balanced JAK1/JAK2 inhibitory profile of baricitinib suggests that baricitinib will have greatest modulatory effect in cytokines signaling through a JAK1/JAK2 heterodimer intracellularly (or a JAK1/JAK2/TYK2, such as IL-6, TSLP, IL-13, or IL-31) (Vaddi and Luchi 2012).

The recently completed Phase 2 study of baricitinib in AD, JAHG, met its primary objective of proportion of patients achieving a 50% improvement from baseline in Eczema Area and Severity Index (EASI) scores compared to placebo. Baricitinib also showed statistically significant improvements for other disease severity analyses as well as multiple different patient-reported outcomes (PROs) compared to placebo, further supporting the hypothesis that JAK1/JAK2 signaling plays a key role in AD pathogenesis.

3.3. Benefit/Risk Assessment

More information about the known and expected benefits, risks, serious adverse events (SAEs), and reasonably anticipated adverse events (AEs) of baricitinib are to be found in the Investigator's Brochure (IB).

4. Objectives and Endpoints

Table JAHN.2. Objectives and Endpoints

Objectives	Endpoints
Primary	
To estimate the effect of long-term therapy with baricitinib on responders and partial responders at entry of JAHN.	• Proportion of patients with a response of IGA 0 or 1 assessed at Weeks 16, 36, and 52
Secondary (Weeks 0–52)	
Baricitinib Patients at Entry to Study JAHN To evaluate the effect of increasing or maintaining baricitinib dose on clinical measures and patient-reported outcomes.	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at Weeks 16, 36, and 52 • Proportion of patients with a response of IGA 0 or 1 assessed at Weeks 16, 36, and 52 (nonresponders) • Proportion of patients achieving response of EASI75 from baseline of originating study assessed at Weeks 16, 36, and 52 • Proportion of patients with a 4-point improvement from baseline of originating study in Itch NRS at 16 weeks
Placebo Nonresponders at Entry to Study JAHN To evaluate the effect of starting baricitinib 2-mg versus 4-mg on clinical measures and patient-reported outcomes.	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at Weeks 4, 16, 24, 52 • Proportion of patients with a response of IGA 0 or 1 assessed at Weeks 4, 16, 24, 52 • Proportion of patients achieving response of EASI75 from baseline of originating study assessed at Weeks 4, 16, 24, 52 • Proportion of patients with a 4-point improvement from baseline of originating study in Itch NRS at 16 weeks
Secondary (Weeks 52–104)	
All Patients Entering the Substudy To evaluate the change in clinical response after treatment withdrawal or downtitration. **Patients Entering the Substudy with IGA 0 or 1** To evaluate the change in clinical response after treatment withdrawal or downtitration.	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at 16 weeks after rerandomization (Week 68) and Week 104 • Proportion of patients with a response of IGA 0 or 1 assessed at 16 weeks after rerandomization (Week 68) and Week 104 • Proportion of patients with a response of EASI75 from baseline of originating study assessed at 16 weeks after rerandomization (Week 68) and Week 104 • Time to retreatment (time to IGA ≥3)
Patients Retreated during Substudy To evaluate the ability to recapture efficacy based on clinical measures after experiencing a loss of treatment benefit:	• Proportion of patients with a response of IGA 0, 1, or 2 within 16 weeks of retreatment • Proportion of patients with a response of IGA 0 or 1 within 16 weeks of retreatment • Proportion of patients with a response of EASI75 from baseline of originating study within 16 weeks of retreatment

Objectives and Endpoints

Objectives	Endpoints
Patients Not Entered into the Substudy To evaluate the effect of maintaining baricitinib dose on clinical measures.	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at Week 104 • Proportion of patients with a response of IGA 0 or 1 assessed at Week 104 • Proportion of patients with a response of EASI75 from baseline of originating study assessed at Week 104
Exploratory objectives 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 2 score, Skin Pain NRS, PGI-S-AD. Patients continuing on placebo as responders will be assessed during the long-term extension for relevant efficacy endpoints. Assessments of efficacy may be performed beyond Week 104 up to Week 200. The timing of the data lock(s) for the analysis of the efficacy data from the sub-study will be determined by the retreatment rates (see Section 10.3.7).	

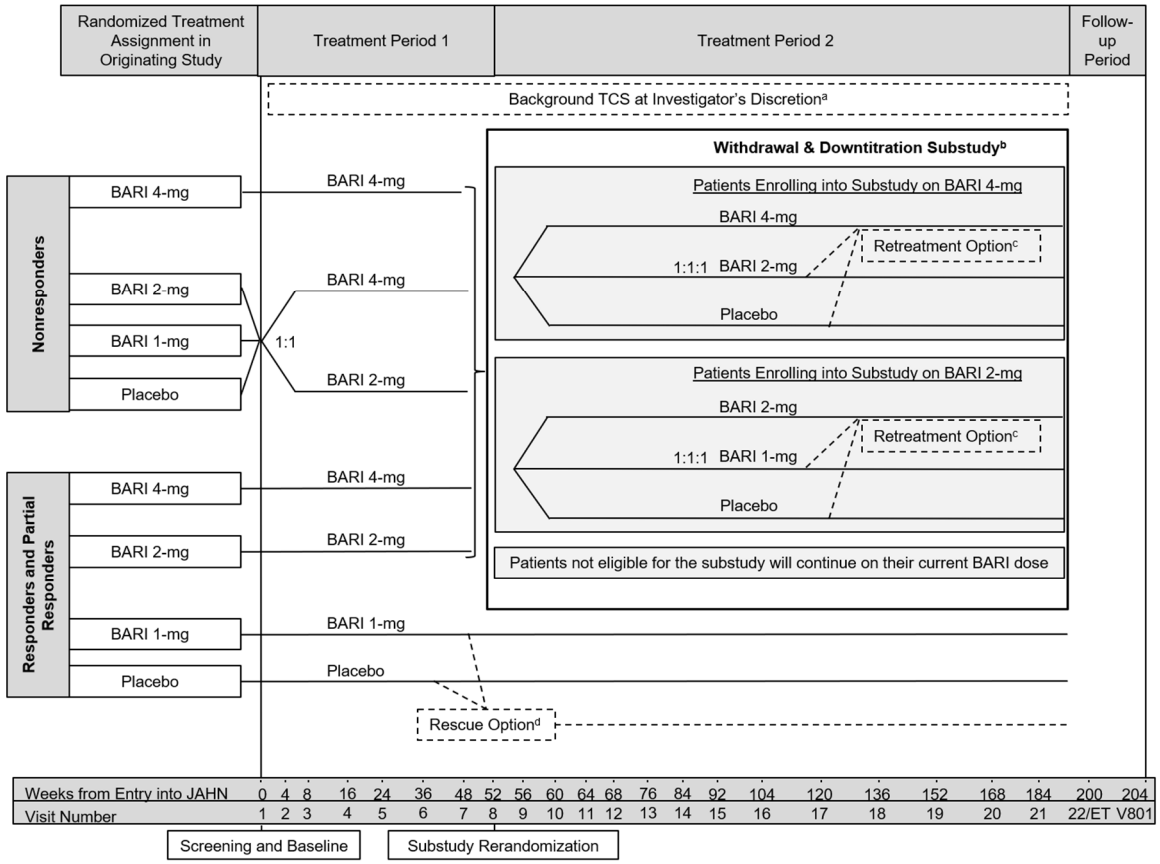
Abbreviations: ADSS = Atopic Dermatitis Sleep Scale; DLQI = Dermatology Life Quality Index; EASI = Eczema Area and Severity Index; EQ-5D-5L = the European Quality of Life–5 Dimensions–5 Levels; IGA = Investigator's Global Assessment; NRS = numeric rating scale; PGI-S-AD = Patient Global Impression of Severity – Atopic Dermatitis; POEM = Patient-Oriented Eczema Measure; SCORAD = SCORing Atopic Dermatitis; TCS = topical corticosteroids; WPAI-AD = Work Productivity and Activity Impairment Questionnaire – Atopic Dermatitis.

5. Study Design

5.1. Overall Design

Study I4V-MC-JAHN (JAHN) is a Phase 3, multicenter, double-blind study to evaluate the long-term safety and efficacy of daily baricitinib 1-mg, 2-mg, and 4-mg in patients with AD for approximately 4 years. The study will consist of 3 study periods and 1 substudy: randomized treatment withdrawal and downtitration. Patients entering Study JAHN will be classified as "Responders and Partial Responders" or "Nonresponders." Figure JAHN.1 illustrates the study design. The full visit schedule is outlined in the Study Schedule of Activities (Section 2). Patients who completed originating Studies I4V-MC-JAHL (JAHL) and I4V-MC-JAHM (JAHM) may be eligible for enrollment into Study JAHN; there may also be additional studies developed that will be eligible to enroll patients directly into Study JAHN.

Patients completing Studies JAHL and JAHM will have had approximately 16 weeks of treatment with investigational product (IP) (baricitinib or placebo) and will enter Study JAHN at Visit 1; the final visit of Study JAHL or JAHM will be the first visit of Study JAHN. Patients will have been using emollients daily in their originating study and this will continue during Study JAHN participation. If oral systemic AD treatments were administered for rescue in the originating study, then a minimum washout of 4 weeks is required prior to study drug initiation in Study JAHN (Visit 1). Daily diary collection will continue through the first 4 months of treatment in JAHN and then collection will resume at Week 48 for an additional 5 months to capture PROs prior to and during the withdrawal and downtitration substudy (except for patients known not to be in the substudy [Section 5.1.3.1]).



Abbreviations: BARI = baricitinib; ET = early termination; IGA = Investigator's Global Assessment; TCS = topical corticosteroids; V = visit.[a] Background TCS may be initiated or reinitiated at any time during the study, following the guidelines in Section 7.7.1 and will be provided as part of rescue or retreatment any time a patient's IGA score becomes ≥3 as described in Sections 5.1.2.1 and 5.1.3.1.

[b] Eligible patients will be rerandomized in the withdrawal and downtitration substudy as described in Section 5.1.3.1. Patients who do not enroll in the substudy will remain on their treatment as described in Section 5.1.3.2.

[c] Patients enrolled in the substudy will automatically be retreated if their IGA score becomes ≥3 as described in Section 5.1.3.1.

[d] Rescue is available as described in Section 5.1.2.1.

Figure JAHN.1. Illustration of study design for clinical Protocol I4V-MC-JAHN.

5.1.1. Screening and Baseline Period

Screening should occur during the last visit of the originating study. However, in particular circumstances, the screening period may be extended:

- if the patient received oral systemic AD treatment as rescue therapy during the originating study and needs additional time to complete the required 4-week washout period. In this situation, patients will need to return in order to complete additional Visit 1 procedures required at the end of the washout period. (See the Study Schedule of Activities [Section 2].) For these patients, the screening period should not exceed 8 weeks following the completion of the Week 16 visit in the originating study unless approval is obtained from the sponsor.
- it may be permissible for screening to occur after the last visit of the originating study if needed, upon approval from the sponsor.

Study eligibility for each patient will be reviewed based on all enrollment criteria (Section 6). Patients who meet all criteria will have treatment allocated by interactive web-response system (IWRS) and begin their treatment period at Visit 1. Doses include placebo, baricitinib 1-mg, baricitinib 2-mg, and baricitinib 4-mg. Patients with renal impairment at screening (Visit 1) of the originating study will not receive doses higher than 2-mg baricitinib. Available treatment assignments will depend on patients' responder status at enrollment. An IWRS will determine patients' responder status based on the following criteria:

- **Responders (Investigator's Global Assessment [IGA] of 0 or 1) and Partial Responders (IGA of 2):** Patients entering Study JAHN who have an IGA score of 0, 1, or 2 AND were not rescued in the originating study.
- **Nonresponders:** any patient who does not meet the "Responder and Partial Responder" definition.

5.1.2. Treatment Period 1: 0 to 52 Weeks

5.1.2.1. Responders and Partial Responders

Treatment for responders and partial responders is diagramed in Figure JAHN.1. Patients entering Study JAHN as a responder or partial responder will remain on their original, randomized treatment assignment: placebo, baricitinib 1-mg, baricitinib 2-mg, or baricitinib 4-mg. These patients did not require rescue during their originating study; thus, they will continue with baricitinib monotherapy unless topical corticosteroids (TCS) are deemed necessary. Topical corticosteroids can be initiated any time during treatment to control worsening and unacceptable symptoms of AD. For management of TCS, follow the guidelines in Section 7.7.1.

If worsening of symptoms results in an IGA ≥3, patients not currently receiving TCS will be provided TCS and automatically rescued as follows:

- patients receiving baricitinib 1-mg or placebo will be rerandomized 1:1 to baricitinib 2-mg or baricitinib 4-mg.
- patients receiving baricitinib 2-mg or 4-mg will continue their same baricitinib dose.

This option is only available once for each patient and will be assigned by IWRS; investigators will remain blinded to treatment assignments. Once this has occurred, the investigator will know that a patient is on the highest dose of baricitinib that they will receive in the study; therefore, if AD symptoms remain unacceptable then discontinuation should be considered. Although patients receiving baricitinib 2-mg and 4-mg will remain on the same dose during rescue, this is done to maintain the blind and assess the effect of long-term therapy with these doses.

At Week 52, patients will be assessed for eligibility for the substudy.

5.1.2.2. Nonresponders

Treatment for nonresponders is diagramed in Figure JAHN.1. Patients entering as a nonresponder on placebo, baricitinib 1-mg, or baricitinib 2-mg will be rerandomized 1:1 to either baricitinib 2-mg or baricitinib 4-mg. Patients entering Study JAHN as a nonresponder on baricitinib 4-mg will remain on baricitinib 4-mg. Patients classified as nonresponders include all patients who were rescued at any time during the originating study or have an IGA ≥3 at baseline of JAHN. For patients with unacceptable symptoms, background TCS is permitted upon enrollment. Higher potency TCS can be used as needed and patients who had discontinued TCS can restart as necessary according to the guidelines in Section 7.7.1. There are no additional treatment modification options for nonresponders after enrollment into JAHN. Patients are on the highest dose of baricitinib that they will receive in the study; therefore, if AD symptoms are unacceptable then discontinuation should be considered.

At Week 52, patients will be assessed for eligibility for the substudy.

5.1.3. Treatment Period 2: 52 to 200 Weeks

5.1.3.1. Substudy at Week 52: Randomized Withdrawal and Downtitration Eligibility

At Week 52, all patients will be evaluated for substudy eligibility. To be eligible, a patient must meet all of the following criteria:

- IGA 0, 1, or 2 at Week 52
- has not used high-potency TCS in the last 14 days (potency classification in Appendix 6)
- does not currently have study drug interrupted
- at entry to JAHN was assigned to baricitinib 2-mg or 4-mg (assessed by IWRS)

Investigators will be aware of the patients not meeting the first 3 criteria, and thus not included in the substudy; however, there will be some patients for which substudy participation is not known based on the originally assigned dose. As such, IGA ≥3 is the criteria for both retreatment (in substudy) and rescue (not in substudy) for these patients to ensure all patients are treated similarly and the blind to treatment group is preserved. All patients who meet the first 3 criteria will continue to follow substudy procedures to maintain the blind.

Treatment

Treatment in the substudy is diagramed in Figure JAHN.1. Patients entering the substudy on baricitinib 4-mg will be rerandomized 1:1:1 to either placebo, baricitinib 2-mg, or baricitinib 4-mg. Patients entering the substudy on baricitinib 2-mg will be rerandomized to either placebo, baricitinib 1-mg, or baricitinib 2-mg. Topical corticosteroids can be continued or initiated to control worsening and unacceptable symptoms of AD any time during treatment. For management of TCS, follow the guidelines in Section 7.7.1.

Retreatment

During the substudy, if worsening of AD symptoms occurs with an IGA ≥3, the patient will be retreated with their original baricitinib dose and will be provided TCS. An unscheduled visit may be needed in order to assess worsening and perform clinical safety and efficacy assessments immediately before retreatment.

5.1.3.2. Patients Not Entered into the Substudy

All patients not entered into the substudy will continue on their current treatment. Topical corticosteroids can be initiated any time during treatment to control worsening and unacceptable symptoms of AD. For management of TCS, follow the guidelines in Section 7.7.1. If worsening of symptoms results in an IGA ≥3, patients not currently receiving TCS will be provided TCS and automatically rescued as follows:

- Patients entering Study JAHN as responders or partial responders who have not previously been rescued during Treatment Period 1 (0 to 52 weeks) are eligible for rescue:
    - patients receiving baricitinib 1-mg or placebo will be rerandomized 1:1 to baricitinib 2-mg or baricitinib 4-mg.
    - patients receiving baricitinib 2-mg or 4-mg will continue their same baricitinib dose.
- Patients entering Study JAHN as nonresponders and patients already rescued during Treatment Period 1 (0 to 52 weeks) are not eligible for a second rescue and will continue on the same dose.

5.1.4. Post-treatment Follow-up Period

Patients who complete the study through Visit 22 (Week 200) will have a post-treatment follow-up visit (Visit 801) approximately 28 days after the last dose of IP.

Patients who discontinue early from the study must have an early termination visit (ETV) and return for the post-treatment safety follow-up visit (Visit 801) approximately 28 days after the last dose of IP.

Patients who have discontinued IP but remain in the study for more than 28 days without IP will have an ETV; however, a separate follow-up visit (Visit 801) is not required.

5.2. Number of Participants

Approximately 1760 patients may be enrolled in Study JAHN, either from originating studies (such as Study JAHL, Study JAHM, or Study JAIY) or future studies or addenda. It is estimated that there will be approximately 600 patients enrolled into the randomized treatment withdrawal and downtitration substudy.

5.3. End of Study Definition

End of the trial is the date of the last visit or last scheduled procedure as shown in the Schedule of Activities (Section 2) for the last patient.

5.4. Scientific Rationale for Study Design

Atopic dermatitis is a chronic, relapsing inflammatory skin disorder (Bieber 2010). Although several therapies are available to patients for the treatment of flares, there are very limited options for long-term management. Most patients rely on chronic use of emollients and intermittent use of TCS or topical calcineurin inhibitors (TCNIs) to regulate skin inflammation in response to flares. In patients with moderate to severe disease, long-term disease control is not always achieved. In order to address this important medical need for a safe and effective long-term therapy, this study will allow for a longer treatment period (up to 4 years).

To adequately assess the long-term effect of specific baricitinib doses, there are instances where patients will remain on their same dose during episodes of worsening disease. However, use of topical treatments (as outlined in Section 7.7.1) is available during Study JAHN.

Treatment withdrawal or dose downtitration for patients with sustained response to therapy are strategies that may be used in the clinical setting. Study JAHN will evaluate the effect of stopping or decreasing the dose of baricitinib in the context of a randomized treatment withdrawal and downtitration substudy starting at Week 52. This timing will allow adequate duration on a stable dose of baricitinib to assess the benefit/risk profile of the dose regimens. It is anticipated that the effects of downtitration and withdrawal will be more reliably assessed in those continuously treated with the 2-mg or 4-mg doses; thus, those being treated with 1-mg are not eligible for the substudy.

5.5. Justification for Dose

The doses proposed for AD Phase 3 studies are baricitinib 1-mg, 2-mg, and 4-mg. These doses were chosen primarily based on the recently completed Phase 2 AD study, JAHG, and are additionally supported by pharmacokinetic (PK), safety, and efficacy data for baricitinib in Phase 2 and Phase 3 RA studies and a Phase 2 psoriasis study.

In the Phase 2 Study JAHG, both the 2-mg and 4-mg doses showed benefit on the primary and major secondary endpoints (EASI, IGA, SCORing Atopic Dermatitis [SCORAD], Patient-Oriented Eczema Measure [POEM], and the Dermatology Life Quality Index [DLQI]) as compared to placebo, and both doses had an acceptable safety profile at Week 16. However, the 4-mg dose appeared to demonstrate a more rapid benefit (at 4 weeks) on the more stringent endpoints (EASI75, EASI90, and IGA 0 or 1) compared to 2-mg dose particularly in the

subgroup of patients with baseline EASI scores ≥16. The 4-mg dose resulted in statistically significant improvement in these endpoints at Week 4 and this level of response was maintained through Week 16. A similar trend between the baricitinib 4-mg and 2-mg doses was observed in patients with RA. Although in Study JAHG the 4-mg dose seemed to perform better than the 2-mg dose on more stringent endpoints, on other endpoints, including EASI-50, and EASI change from baseline, 2-mg and 4-mg doses showed similar efficacy compared to placebo. Thus, based on available data, 3 doses will be included in Phase 3, including a 1-mg dose, to cover the range of exposures where clinical responses could be anticipated.

5.5.1. Dose Adjustment for Renal Impairment

Patients who received a dose adjustment due to renal impairment in their originating study will continue to receive a renal dose adjustment in Study JAHN.

Baricitinib exposure increases with decreased renal function (Study I4V-MC-JADL). Based on PK simulations of baricitinib exposures for the mild and moderate categories of renal function (stratified as estimated glomerular filtration rate [eGFR] 60 to <90 mL/min/1.73 m^2 and eGFR 30 to <60 mL/min/1.73 m^2, respectively), dose adjustment is not required for patients with eGFR ≥60 mL/min/1.73 m^2.

During Study JAHN, patients with eGFR <60 mL/min/1.73 m^2 who are randomized to the 4-mg dose will receive a dose of 2-mg, ensuring that their exposures would not exceed those of the 4-mg dose in patients with eGFR ≥60 mL/min/1.73 m^2. For patients randomized to the 2-mg or 1-mg dose, there is no dose correction based on renal function. See Section 8.1.1 for eGFR thresholds that trigger interruption of IP.

The procedure of dose adjustment based on renal function (eGFR) during the study is detailed in Section 7.2.2.

6. Study Population

The study population will comprise patients diagnosed with AD who have completed an eligible originating study. Study investigator(s) will review patient records and screening test results to determine that the patient meets all inclusion and exclusion criteria to qualify for participation in the study.

Prospective approval of protocol deviations to recruitment and enrollment criteria, also known as protocol waivers or exemptions, are not permitted.

6.1. Inclusion Criteria

Patients are eligible to be included in the study only if they meet the following criteria:

[1] Have completed the final active treatment visit for an originating study (i.e., Visit 8, Week 16 of Studies JAHL or JAHM) eligible to enroll patients directly into Study JAHN.

[2] Are able to read, understand, and give documented (electronic or paper signature) informed consent.

[3] Are male or nonpregnant, nonbreastfeeding female patients, except

a. Male patients must agree to use 2 forms of birth control (one must be highly effective, see below) while engaging in sexual intercourse with female partners of childbearing potential while enrolled in the study and for at least 4 weeks following the last dose of IP.

b. Female patients of childbearing potential must agree to use 2 forms of birth control when engaging in sexual intercourse with a male partner while enrolled in the study and for at least 4 weeks following the last dose of IP.

The following birth control methods are considered acceptable (the patient should choose 2 methods to be used with her male partner, and 1 must be highly effective):

- Highly effective birth control methods: oral, injectable, or implanted hormonal contraceptives (combined estrogen/progesterone or progesterone only, associated with inhibition of ovulation); intrauterine device or intrauterine system (e.g., progestin-releasing coil); or vasectomized male (with appropriate post-vasectomy documentation of the absence of sperm in the ejaculate)
- Effective birth control methods: condom with a spermicidal foam, gel, film, cream, or suppository; occlusive cap (diaphragm or cervical/vault caps) with a spermicidal foam, gel, film, cream, or suppository; or oral hormonal contraceptives

Note: When local guidelines concerning highly effective or effective methods of birth control differ from the above, the local guidelines must be followed.

c. Females of nonchildbearing potential are not required to use birth control and they are defined as:

- women ≥60 years of age or women who are congenitally sterile, or
- women ≥40 and <60 years of age who have had a cessation of menses for ≥12 months and a follicle-stimulating hormone test from baseline in originating study, confirming nonchildbearing potential (≥40 mIU/mL or ≥40 IU/L), or
- women who are surgically sterile (i.e., have had a hysterectomy or bilateral oophorectomy or tubal ligation).

6.2. Exclusion Criteria

Patients will be excluded from the study if they meet **any** of the following criteria:

[4] Are currently enrolled in any other clinical trial involving an IP or any other type of medical research judged not to be scientifically or medically compatible with this study.

[5] Have significant uncontrolled cerebro-cardiovascular (e.g., myocardial infarction [MI], unstable angina, unstable arterial hypertension, severe heart failure, or cerebrovascular accident), respiratory, hepatic, renal, gastrointestinal, endocrine, hematologic, neuropsychiatric disorders, or abnormal laboratory values that developed during a previous baricitinib study that, in the opinion of the investigator, pose an unacceptable risk to the patient if IP continues to be administered.

[6] Have a known hypersensitivity to baricitinib or any component of this IP.

[7] Had IP permanently discontinued at any time during a previous baricitinib study, except for patients who had IP discontinued during originating study due to rescue with an oral systemic AD therapy (e.g., corticosteroid, cyclosporine, methotrexate)

[8] Had temporary IP interruption at the final study visit of a previous baricitinib study and, in the opinion of the investigator, this poses an unacceptable risk for the patient's participation in the study.

[9] Have any other condition that, in the opinion of the investigator, renders the patient unable to understand the nature, scope, and possible consequences of the study or precludes the patient from following and completing the protocol.

[10] Are unwilling or unable to comply with the use of a data collection device to directly record data from the subject.

6.3. Lifestyle Restrictions

Not applicable.

6.4. Screen Failures

Individuals who do not meet the criteria for participation in this study (screen failure) may not be rescreened.

7. Treatments

7.1. Treatments Administered

Patients in this long-term extension study will receive either baricitinib 1-mg, baricitinib 2-mg, baricitinib 4-mg, or placebo administered orally once a day. Patients with renal impairment who received a baricitinib dose of 2-mg daily in the originating study will continue to receive a 2-mg daily dose of baricitinib in Study JAHN. Patients with renal impairment randomized to baricitinib 4-mg in JAHN will receive baricitinib 2-mg; patients with renal impairment randomized to baricitinib 2-mg or 1-mg in Study JAHN will not have their dose modified. Table JAHN.3 shows the treatment regimens.

Table JAHN.3. Treatment Regimens

Assigned Treatment	IP Supplied in JAHN	Dose
Baricitinib 4-mg QD[a]	Baricitinib 4-mg tablets Placebo to match 2-mg tablets Placebo to match 1-mg tablets	3 tablets per day
Baricitinib 2-mg QD	Baricitinib 2-mg tablets Placebo to match 4-mg tablets Placebo to match 1-mg tablets	3 tablets per day
Baricitinib 1-mg QD	Baricitinib 1-mg tablets Placebo to match 4-mg tablets Placebo to match 2-mg tablets	3 tablets per day
Placebo QD	Placebo to match 4-mg tablets Placebo to match 2-mg tablets Placebo to match 1-mg tablets	3 tablets per day

Abbreviations: eGFR = estimated glomerular filtration rate; IP = investigational product; QD = once daily.

[a] The maximum baricitinib dose for patients with renal impairment, defined as eGFR <60 mL/min/1.73 m^2, will be 2-mg QD.

The investigator or his/her designee is responsible for the following:

- explaining the correct use of the investigational agent(s) to the patient
- verifying that instructions are followed properly
- maintaining accurate records of IP dispensing and collection
- at the end of the study, returning all unused medication to Lilly, or its designee, unless the sponsor and sites have agreed that all unused medications are to be destroyed by the site, as allowed by local law.

Patients will be instructed to take 3 tablets daily: 1 tablet matching appearance of baricitinib 4-mg, 1 tablet matching appearance of baricitinib 2-mg, and 1 tablet matching appearance of baricitinib 1-mg. All 3 tablets should be taken together, at the same time each day.

7.1.1. Packaging and Labeling

The sponsor (or its designee) will provide the following IPs:

- tablets containing 4-mg baricitinib
- tablets containing 2-mg baricitinib
- tablets containing 1-mg baricitinib
- placebo tablets to match baricitinib 4-mg, 2-mg, and 1-mg tablets.

Packaging for each dose will include 3 tablets per day. Each tablet has a distinctive shape and color, 4-mg versus 2-mg versus 1-mg, and each tablet strength has a matching placebo. Each active dose package will contain the appropriate active tablet strength, and corresponding placebo tablets for the other strengths, as noted in Table JAHN.3.

Investigational product tablets will be provided in cartons containing 4 blister packs.

Clinical trial materials will be labeled according to the country's regulatory requirements.

7.2. Method of Treatment Assignment

At entry into JAHN, patients who meet all criteria for enrollment will be randomized or assigned treatment by a computer-generated random sequence using an IWRS. Patients originally assigned to placebo, 1-mg baricitinib or 2-mg baricitinib in studies JAHL or JAHM and classified as nonresponders will be randomized in a 1:1 ratio to 2-mg baricitinib or 4-mg baricitinib and will be stratified by disease severity at baseline of JAHN (IGA 0, 1, 2 versus IGA 3 versus IGA 4). All other patients will be assigned to treatment in JAHN matching their prior assignment from JAHL or JAHM. Assignment to treatment groups will be determined by a computer-generated random sequence using an IWRS.

At Week 52, patients eligible for the withdrawal and downtitration substudy will be assigned to treatment by a computer-generated random sequence using an IWRS. Rerandomization will follow a 1:1:1 ratio allocation and will be stratified by disease severity (IGA 0, 1 versus IGA 2).

The IWRS will be used to assign blister packs, each containing double-blind IP tablets, to each patient, starting at Visit 1 (Week 0), and at each visit up to and including Visit 21 (Week 184). Site personnel will confirm that they have located the correct blister packs by entering a confirmation number found on the blister packs into the IWRS.

7.2.1. Selection and Timing of Doses

The IP (3 tablets from blister pack) should be taken once daily without regard to food and, if possible, at approximately the same time every day, usually at the start of the patient's day, to aid patient compliance.

7.2.2. Dose Adjustment for Renal Impairment

The rationale of dose adjustment for patients with documented renal impairment (defined as screening in the originator study eGFR ≥40 to <60 mL/min/1.73 m^2) is detailed in Section 5.5.1.

The dose adjustment for renal impairment will be managed by IWRS to ensure maintenance of the treatment blind. Patients who received dose adjustment for renal impairment in their originator study will continue to receive dose adjustment in JAHN.

Patients with documented renal impairment (defined as screening in the originator study eGFR ≥40 to <60 mL/min/1.73 m^2), who are randomized to the 4-mg active treatment arm will receive a dose of 2-mg by the IWRS. For patients randomized to the 2–mg dose or 1–mg dose, there will be no dose adjustment based on renal function.

No dose adjustment will be made for patients with screening eGFR ≥60 mL/min/1.73 m^2 in the originator study. These patients who are randomized to active treatment will receive their assigned dose, either baricitinib 4–mg, 2–mg, or 1–mg, respectively.

During the study, for patients with documented renal impairment when the subsequent eGFR falls <30 mL/min/1.73 m^2, IP will be withheld until their eGFR becomes ≥40 mL/min/1.73 m^2, whereupon the IP dosing may resume. For patients with screening eGFR ≥60 mL/min/1.73 m^2, when the subsequent eGFR falls to <40 mL/min/1.73 m^2, IP will be withheld until their eGFR becomes ≥50 mL/min/1.73 m^2, whereupon the IP dosing may resume (see Section 8.1.1).

7.3. Blinding

This is a double-blind study. To preserve the blinding of the study, a minimum number of Lilly personnel will see the randomization table and treatment assignments before the study is complete. All study assessments will be performed by study personnel who are blinded to the patient's treatment group. Except in clinical circumstances where unblinding is required, the patients, investigators, Lilly study team, and any personnel interacting directly with patients or investigative sites will remain blinded to baricitinib and placebo assignment until after completion of the double-blinded treatment period. It is expected that the need for unblinding a patient's treatment prior to completion of the double-blinded treatment period will be extremely rare. Every effort should be made to preserve the blind unless there is a compelling reason that knowledge of the specific treatment would alter the medical care of the patient. In case of an emergency, the investigator has the sole responsibility for determining if unblinding of a patient's treatment assignment is warranted for medical management of the event. Patient safety must always be the first consideration in making such a determination. If a patient's treatment assignment is unblinded, Lilly must be notified immediately. If the investigator decides that unblinding is warranted, it is the responsibility of the investigator to promptly document the decision and rationale and notify Lilly as soon as possible.

Emergency unblinding for AEs may be performed through the IWRS. This option may be used ONLY if the patient's well-being requires knowledge of the patient's treatment assignment. All unblinding events are recorded and reported by the IWRS. If an investigator, site personnel performing assessments, or the patient is unblinded, the patient must be discontinued from the study. In cases where there are ethical reasons to have the patient remain in the study, the investigator must obtain specific approval from a Lilly clinical research physician for the patient to continue in the study.

The processes to maintain blinding during the interim analysis conducted by the Data Monitoring Committee (DMC) are described in Section 10.3.7.

7.4. Dosage Modification

Not applicable.

7.5. Preparation/Handling/Storage/Accountability

All IPs (used and partially used) will be returned to the sponsor or destroyed at site level with the sponsor's written approval. In some cases, sites may destroy the material if, during the investigative site selection, the evaluator has verified and documented that the site has appropriate facilities and written procedures to dispose of clinical trial materials.

Follow storage and handling instructions on the IP packaging.

7.6. Treatment Compliance

Patient compliance with IP will be assessed at Visit 2 through Visit 22 and at Early Termination during the treatment period by counting returned tablets.

A patient will be considered significantly noncompliant if he or she misses more than 20% of the prescribed doses of IP during the study, unless the patient's IP 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 medication during the study.

Patients will be counseled by study staff on the importance of taking the IP as prescribed, as appropriate.

Patients' compliance will be further defined in the statistical analysis plan (SAP).

7.7. Concomitant Therapy

All concomitant medications, whether prescription or over-the-counter, must be recorded on the Concomitant Medication electronic case report form (eCRF). Patients will be instructed to consult the investigator or other appropriate study personnel at the site before taking any new medications or supplements during the study.

7.7.1. Topical Corticosteroids and Other Topical Treatments

Prior to use of TCS, it is recommended that increased frequency of emollient use is attempted at least twice a day in an effort to control symptoms. If symptoms are still not controlled, and topical treatments other than emollients are needed, then the investigator should start with low-to-medium-potency TCS options.

Choice of Topical Treatment

- Triamcinolone cream 0.1% and/or hydrocortisone 2.5% ointment. Where possible both of these treatments will be supplied by the sponsor during the first 2 years of the treatment period (dispensed at Visits 1-15), and use should be recorded via weight of

returned tube as indicated in the SOA (Section 2). In the event of these specific TCS being unavailable during the first 2 years, an alternate, equivalent-potency TCS may be provided by the sponsor. Following Visit 15, patients may independently continue to use their TCS of choice as directed by their investigator, as per clinical practice, but these will not be provided by the sponsor and weight will not be recorded.

  - In the event that the sponsor is unable to supply TCS during the first 2 years, commercially available triamcinolone 0.1% cream and/or hydrocortisone 2.5% ointment may be supplied by the sites. Where providing triamcinolone 0.1% cream and/or hydrocortisone 2.5% ointment is not possible, an equivalent-potency TCS cream and/or ointment that is in line with local practices can be supplied. Refer to Appendix 6 for guidance on potency equivalence.
  - If the TCS supplied by the sponsor during the first 2 years is not considered suitable for an individual patient, an equivalent-potency TCS cream and/or ointment that is in line with local practices can be supplied by the sites. Refer to Appendix 6 for guidance on potency equivalence.

- Investigators may also select to use TCNIs and/or crisaborole where approved, although use of either during the study is neither encouraged nor provided. If TCNIs are prescribed, use should be limited to problem areas only (e.g., face, neck, skin folds, genital areas, etc).
- On the days of study visits, topical therapy should not be applied before the patient has undergone all study procedures and clinical evaluations in order to allow adequate assessment of skin dryness.
- Use of any topical therapy will be documented in the CRF.

In patients who do not improve sufficiently with the provided topical therapy after 7 days, a higher potency TCS may be used (Appendix 6) and IP may continue. Higher potency TCS will not be supplied centrally by the sponsor. It is recommended that if a patient reaches "clear" to "almost clear" skin while receiving topical therapy, then medium- or high-potency TCS and TCNI should be stopped, and low–potency TCS should be used once daily for 7 days then stopped. Following cessation of TCS, if a patient again experiences worsening and unacceptable symptoms of AD then TCS can be re-initiated at the discretion of the investigator.

The protocol states that in certain circumstances TCS will be provided; however, the decision to dispense TCS is ultimately at the investigator's discretion.

7.7.2. Other Permitted Medications and Procedures

Treatment with concomitant AD therapies during the study is permitted only as described below.

- Daily use of emollients is required as background treatment. If daily applications are missed, it will not be considered a protocol violation.
  - Patients should not apply emollients on the day of their study visit prior to the procedures to allow adequate assessment of skin dryness.

- Antihistamines are allowed.
- Intra-articular or soft tissue (bursa, tendons, and ligaments) corticosteroid injections are allowed.
- Intranasal or inhaled steroid use is allowed.
- Topical anesthetics and topical and systemic anti-infective medications are allowed.
- Nonlive seasonal vaccinations and/or emergency vaccination, such as rabies or tetanus vaccinations, are allowed.

Any changes of these concomitant medications must be recorded in the Concomitant Therapy of Special Interest eCRF.

Treatment with concomitant therapies for other medical conditions such as diabetes and hypertension is permitted during the study.

7.7.3. Prohibited Medications and Procedures

Prohibited Medications and Procedures Not Requiring Interruption of Investigational Product

The following therapies will not be allowed during the course of the study and, if taken by or administered to the patient, the prohibited therapy must be discontinued.

- phototherapy including PUVA (psoralen and ultraviolet A), ultraviolet B, tanning booth, and excimer laser
- bleach baths

Prohibited Medications Requiring Temporary Interruption of Investigational Product

The following therapies will not be allowed during the course of the study and, if taken by or administered to the patient, temporary interruption of investigational product is required.

- live vaccines (including Bacillus Calmette-Guérin [BCG] or herpes zoster)
    - For BCG vaccination, investigational product should be temporarily interrupted for 12 weeks
    - For herpes zoster vaccination, investigational product should be temporarily interrupted for 4 weeks
- probenecid: if a patient is inadvertently started on probenecid, IP should be temporarily interrupted, and can be resumed after patient has discontinued probenecid. If a patient is not able to discontinue probenecid, then IP should be permanently discontinued

- systemic corticosteroids may be used for the treatment of an AE (for example, worsening of existing condition, such as asthma flare). Investigational product may be restarted if systemic corticosteroids were used for a short duration (<30 days). If used for >30 days, sponsor approval to restart investigational product is required.

Prohibited Medications Requiring Permanent Discontinuation of Investigational Product

- systemic corticosteroids used for the treatment of AD
- any other systemic therapy, investigational or commercial (approved or off-label use), used for the treatment of AD except for antihistamines
- other JAK inhibitors (e.g., tofacitinib and ruxolitinib)
- systemic immunosuppressive/immunomodulatory substances, including, but not limited to, cyclosporine, methotrexate, mycophenolate mofetil, interferon-γ, azathioprine, or biologic agents

Note: In the event that these prohibited medications were inadvertently used, agreement and documentation to continue investigational product must be sought from sponsor**.**

7.8. Treatment after the End of the Study

7.8.1. Continued Access

Baricitinib will not be made available to patients after conclusion of the study. Patients will be referred to their local treatment centers for continued therapy as clinically indicated.

8. Discontinuation Criteria

8.1. Discontinuation from Study Treatment

8.1.1. Temporary Interruption from Study Treatment

In some circumstances, it may be necessary to temporarily interrupt treatment as a result of AEs or abnormal laboratory values that may have an unclear relationship to IP. For example, IP should be temporarily interrupted if the patient experiences a cardiovascular AE considered to be related to study treatment, graded as moderate (Grade 2 according to Common Terminology Criteria for Adverse Events [CTCAE] Version 3.0), and that does not resolve promptly with supportive care. Except in cases of emergency, it is recommended that the investigator consult with Lilly (or its designee) before temporarily interrupting therapy for reasons other than those predefined in this section.

For the abnormal laboratory findings and clinical events (regardless of relatedness) listed in Table JAHN.4, specific guidance is provided for temporarily interrupting treatment and when treatment may be restarted. Retest frequency and timing of follow-up laboratory tests to monitor the abnormal finding is at the discretion of the investigator. Investigational product that was temporarily interrupted because of an AE or abnormal laboratory value not specifically covered in Table JAHN.4 may be restarted at the discretion of the investigator.

Table JAHN.4. Criteria for Temporary Interruption of Investigational Product

Hold IP if the Following Laboratory Test Results or Clinical Events Occur:	IP May Be Resumed When:
WBC count <2000 cells/µL (<2.00x10^3/µL or <2.00 GI/L)	WBC count ≥2500 cells/µL (≥2.50x10^3/µL or ≥2.50 GI/L)
ANC <1000 cells/µL (<1.00x10^3/µL or <1.00 GI/L)	ANC ≥1200 cells/µL (≥1.20x10^3/µL or ≥1.20 GI/L)
Lymphocyte count <500 cells/µL (<0.50x10^3/µL or <0.50 GI/L)	Lymphocyte count ≥750 cells/µL (≥0.75x10^3/µL or ≥0.75 GI/L)
Platelet count <75,000/µL (<75x10^3/µL or <75 GI/L)	Platelet count ≥100,000/µL (≥100x10^3/µL or ≥100 GI/L)
eGFR <40 mL/min/1.73 m^2 (from serum creatinine) for patients with originating study screening eGFR ≥60 mL/min/1.73 m^2	eGFR ≥50 mL/min/1.73 m^2
eGFR <30 mL/min/1.73 m^2 (from serum creatinine) for patients with originating study screening eGFR ≥40 to <60 mL/min/1.73 m^2	eGFR ≥40 mL/min/1.73 m^2
ALT or AST >5 x ULN	ALT and AST return to <2 x ULN, and IP is not considered to be the cause of enzyme elevation
Hemoglobin <8 g/dL (<80.0 g/L)	Hemoglobin ≥10 g/dL (≥100.0 g/L)
Symptomatic herpes zoster	All skin lesions have crusted and are resolving
Infection that, in the opinion of the investigator, merits the IP being interrupted	Resolution of infection

Abbreviations: ALT = alanine aminotransferase; ANC = absolute neutrophil count; AST = aspartate aminotransferase; eGFR = estimated glomerular filtration rate; IP = investigational product; ULN = upper limit of normal; WBC = white blood cell.

Although temporary interruption of IP is not a requirement at times of increased potential risk of VTE (venous thromboembolic event; e.g., surgery, significant air travel, or other situations involving prolonged immobilization), following appropriate VTE prophylaxis guidelines is recommended to help manage the elevated risk under these circumstances.

For specific guidance on temporary interruption of IP after use of a prohibited medication, please refer to Section 7.7.3 (Prohibited Medications and Procedures).

Lastly, IP should be temporarily interrupted for suicidal ideation or any suicide-related behaviors as assessed by the following patient responses on the Columbia-Suicide Severity Rating Scale (C-SSRS):

- A "yes" answer to Question 4 (Active Suicidal Ideation with Some Intent to Act, Without Specific Plan) **or**
- A "yes" answer to Question 5 (Active Suicidal Ideation with Specific Plan and Intent) on the "Suicidal Ideation" portion of the C-SSRS **or**
- A "yes" answer to any of the suicide-related behaviors (actual attempt, interrupted attempt, aborted attempt, preparatory act or behavior) on the "Suicidal Behavior" portion of the C-SSRS.

NOTE: Prior to resumption of IP, it is recommended that the patient be assessed by a psychiatrist or appropriately trained professional to assist in deciding whether he or she should remain on IP and ultimately continue participation in the study. Patients do not necessarily have to have IP interrupted if they have self-injurious behavior that would be classified as nonsuicidal self-injurious behavior.

8.1.2. Permanent Discontinuation from Study Treatment

Investigational product should be permanently discontinued if the patient requests to discontinue IP.

Discontinuation of the IP for abnormal liver tests should be considered by the investigator when a patient meets one of the following conditions after consultation with the Lilly-designated medical monitor:

- alanine aminotransferase (ALT) or aspartate aminotransferase (AST) >8 x upper limit of normal (ULN)
- ALT or AST >5 x ULN for more than 2 weeks
- ALT or AST >3 x ULN and total bilirubin level (TBL) >2 x ULN or international normalized ratio (INR) >1.5
- ALT or AST >3 x ULN with the appearance of fatigue, nausea, vomiting, right upper-quadrant pain or tenderness, fever, and/or rash

- alkaline phosphatase (ALP) >3 x ULN
- ALP >2.5 x ULN and TBL >2 x ULN
- ALP >2.5 x ULN with the appearance of fatigue, nausea, vomiting, right quadrant pain or tenderness, fever, and/or rash

NOTE: Patients who are discontinued from IP due to a hepatic event or liver test abnormality should have additional hepatic safety data collected via the hepatic safety eCRF.

Investigational product should be permanently discontinued if any of the following laboratory abnormalities are observed:

- white blood cell count <1000 cells/μL ($1.00x10^3$/μL or 1.00 GI/L)
- ANC <500 cells/μL ($0.50x10^3$/μL or 0.50 GI/L)
- lymphocyte count <200 cells/μL ($0.20x10^3$/μL or 0.20 GI/L)
- hemoglobin <6.5 g/dL (<65.0 g/L).

NOTE: Temporary interruption rules (see Section 8.1.1) must be followed where applicable. For laboratory values that meet permanent discontinuation thresholds, IP should be discontinued. However, if in the opinion of the investigator the laboratory abnormality is due to intercurrent illness such as cholelithiasis or another identified factor, laboratory tests may be repeated. Only when the laboratory value meets resumption thresholds (Table JAHN.4) following the resolution of the intercurrent illness or other identified factor may the investigator restart IP, after consultation with the Lilly-designated medical monitor.

In addition, patients will be discontinued from IP in the following circumstances:

- pregnancy
- malignancy (except for successfully treated basal or squamous cell skin carcinoma)
- hepatitis B virus (HBV) DNA is detected with a value above limit of quantitation or 2 sequential tests return a value of below the limit of quantitation (see Section 9.4.7).
- certain prohibited medications are taken per Section 7.7.3 (Prohibited Medications and Procedures).
- develop a VTE

NOTE: Patients who develop a VTE may have additional follow-up and testing recommended (see Appendix 7). Patients discontinuing from the IP prematurely for any reason should complete AE and other follow-up procedures per Section 2 (Schedule of Activities), Section 9.2 (Adverse Events), and Section 9.4 (Safety) of this protocol.

8.1.3. Discontinuation of Inadvertently Enrolled Patients

If the sponsor or investigator identifies a patient who did not meet enrollment criteria and was inadvertently enrolled, then the patient should be discontinued from study treatment unless there are extenuating circumstances that make it medically necessary for the patient to continue on study treatment. If the investigator and the sponsor clinical research physician (CRP) agree that it is medically appropriate to continue, the investigator must obtain documented approval from the sponsor CRP to allow the inadvertently enrolled patient to continue in the study with or without treatment with IP. Safety follow-up is as outlined in Section 2 (Schedule of Activities), Section 9.2 (Adverse Events), and Section 9.4 (Safety) of the protocol.

8.2. Discontinuation from the Study

Patients may choose to withdraw from the study for any reason at any time, and the reason for early withdrawal will be documented.

Some possible reasons that may lead to permanent discontinuation include the following:

- enrollment in any other clinical trial involving an IP or enrollment in any other type of medical research judged not to be scientifically or medically compatible with this study
- participation in the study needs to be stopped for medical, safety, regulatory, or other reasons consistent with applicable laws, regulations, and good clinical practice (GCP)
- study termination may occur in a specific country or region when baricitinib is approved for the treatment of atopic dermatitis and becomes reimbursed or commercially available in that country or region, or a negative regulatory opinion is received in that country or region.
- investigator decision
    - the investigator decides that the patient should be discontinued from the study
    - if the patient, for any reason, requires treatment with another therapeutic agent that has been demonstrated to be effective for treatment of the study indication, discontinuation from the study occurs prior to introduction of the new agent
- patient decision
    - the patient requests to be withdrawn from the study

Patients discontinuing from the study prematurely for any reason should complete AE and other safety follow-up per Section 2 (Schedule of Activities), Section 9.2 (Adverse Events), and Section 9.4 (Safety) of this protocol.

8.3. Lost to Follow-Up

A patient will be considered lost to follow-up if he or she repeatedly fails to return for scheduled visits and is unable to be contacted by the study site. Site personnel are expected to make

diligent attempts to contact patients who fail to return for a scheduled visit or were otherwise unable to be followed up by the site.

9. Study Assessments and Procedures

Section 2 lists the Schedule of Activities, with the study procedures and their timing (including tolerance limits for timing).

Appendix 3, Appendix 4, and Appendix 5 list the laboratory tests that will be performed for this study.

Unless otherwise stated in the subsections below, all samples collected for specified laboratory tests will be destroyed within 60 days of receipt of confirmed test results. Certain samples may be retained for a longer period, if necessary, to comply with applicable laws, regulations, or laboratory certification standards.

9.1. Efficacy Assessments

9.1.1. Investigator's Global Assessment

Validated Investigator's Global Assessment for Atopic Dermatitis (vIGA-AD): The IGA used in this study, the vIGA-AD (referred to as the IGA throughout the protocol) measures the investigator's global assessment of the patient's overall severity of their AD, based on a static, numeric 5-point scale from 0 (clear skin) to 4 (severe disease). The score is based on an overall assessment of the degree of erythema, papulation/induration, oozing/crusting, and lichenification.

9.1.2. Eczema Area and Severity Index Scores

The EASI assesses extent of disease at 4 body regions and measures 4 clinical signs:(1) erythema, (2) induration/papulation, (3) excoriation, and (4) lichenification each on a scale of 0 to 3. The EASI confers a maximum score of 72. The EASI evaluates 2 dimensions of AD: disease extent and clinical signs (Hanifin et al. 2001).

Body surface area affected by AD will be derived from data collected as part of the EASI assessment.

9.1.3. SCORing Atopic Dermatitis

The SCORAD index uses the rule of nines to assess disease extent and evaluates 6 clinical characteristics to determine disease severity: (1) erythema, (2) edema/papulation, (3) oozing/crusts, (4) excoriation, (5) lichenification, and (6) dryness. The SCORAD index also assesses subjective symptoms of pruritus and 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).

9.1.4. Health Outcomes and Quality-of-Life Measures

The patient self-reported questionnaires will be administered via either an electronic patient diary or an electronic tablet and in countries where the questionnaires have been translated into the native language of the region and linguistically validated.

9.1.4.1. Patient-Oriented Eczema Measure

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 past 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 to 28, with higher total scores indicating greater disease severity (Charman et al. 2004).

9.1.4.2. Itch Numeric Rating Scale

The Itch Numeric Rating Scale (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).

9.1.4.3. Atopic Dermatitis Sleep Scale

The Atopic Dermatitis Sleep Scale (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 representing "not at all" to 4 representing "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.

9.1.4.4. Skin Pain Numeric Rating Scale

Skin Pain NRS is a patient-administered, 11-point horizontal scale anchored at 0 and 10, with 0 representing "no pain" and 10 representing "worst pain imaginable." Overall severity of a patient's skin pain is indicated by selecting the number that best describes the worst level of skin pain in the past 24 hours.

9.1.4.5. Patient Global Impression of Severity

The Patient Global Impression of Severity - Atopic Dermatitis (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."

9.1.4.6. Dermatology Life Quality Index

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 "not at all," "a lot," and "very much," with corresponding scores of 1, 2, and 3, respectively, and unanswered ("not relevant") responses scored as 0. Scores range from 0 to 30, with higher scores indicating greater impairment of QoL. A 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).

9.1.4.7. European Quality of Life – 5 Dimensions – 5 Levels

The European Quality of Life–5 Dimensions–5 Levels (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 visual analog scale (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 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]).

9.1.4.8. Work Productivity and Activity Impairment Questionnaire – Atopic Dermatitis

The Work Productivity and Activity Impairment Questionnaire – Atopic Dermatitis (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.

9.1.5. Appropriateness of Assessments

All assessments utilized in this study are standard, widely used, and generally recognized as reliable, accurate, and relevant, with the exception of ADSS and Skin Pain NRS, which are currently being developed and validated according to regulatory guidance.

9.2. Adverse Events

Investigators are responsible for monitoring the safety of patients who have entered this study and for alerting Lilly or its designee to any event that seems unusual, even if this event may be considered an unanticipated benefit to the patient.

The investigator is responsible for the appropriate medical care of patients during the study.

Investigators must document their review of each laboratory safety report.

The investigator remains responsible for following, through an appropriate health care option, AEs that are serious or otherwise medically important, considered related to the IP or the study, or that caused the patient to discontinue the IP before completing the study. The patient should

be followed until the event resolves, stabilizes with appropriate diagnostic evaluation, or is reasonably explained. The frequency of follow-up evaluations of the AE is left to the discretion of the investigator.

Lack of drug effect is not an AE in clinical studies, because the purpose of the clinical study is to establish treatment effect.

After the informed consent form (ICF) is signed, study site personnel will record via eCRF the occurrence and nature of each patient's preexisting conditions, including clinically significant signs and symptoms of the disease under treatment in the study. In addition, site personnel will record any change in the condition(s) and any new conditions as AEs. Investigators should record their assessment of the potential relatedness of each AE to IP via eCRF.

The investigator will interpret and document whether or not an AE has a reasonable possibility of being related to study treatment, study device, or a study procedure, taking into account the disease, concomitant treatment, or pathologies. A "reasonable possibility" means that there is a cause-and-effect relationship between the IP, study device, and/or study procedure and the AE. The investigator answers yes/no when making this assessment.

Planned surgeries and nonsurgical interventions should not be reported as AEs unless the underlying medical condition has worsened during the course of the study.

If a patient's IP is discontinued as a result of an AE, study site personnel must report this to Lilly or its designee via eCRF, clarifying, if possible, the circumstances leading to any dosage modifications or discontinuations of treatment.

9.2.1. Serious Adverse Events

An SAE is any AE from this study that results in one of the following outcomes:

- death
- initial or prolonged inpatient hospitalization
- a life-threatening experience (i.e., 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. Examples of such medical events include allergic bronchospasm requiring intensive treatment in an emergency room or at home, blood dyscrasias, or convulsions that do not result in inpatient hospitalization, or the development of drug dependency or drug abuse

All AEs occurring after signing the ICF are recorded in the eCRF and assessed for serious criteria. The SAE reporting to the sponsor begins after the patient has signed the ICF and has received IP. However, if an SAE occurs after signing the ICF, but prior to receiving IP, the SAE

should be reported to the sponsor as per SAE reporting requirements and timelines if it is considered reasonably possibly related to study procedure.

Study site personnel must alert Lilly or its designee of any SAE within 24 hours of investigator awareness of the event via a sponsor-approved method. If alerts are issued via telephone, they are to be immediately followed with official notification on study-specific SAE forms. This 24-hour notification requirement refers to the initial SAE information and all follow-up SAE information. Patients with a serious hepatic AE should have additional data collected using the hepatic safety eCRF.

Pregnancy (during maternal or paternal exposure to IP) does not meet the definition of an AE. However, to fulfill regulatory requirements any pregnancy should be reported following the SAE process to collect data on the outcome for both mother and fetus.

Investigators are not obligated to actively seek AEs or SAEs in subjects once they have discontinued and/or completed the study (the patient summary CRF has been completed). However, if the investigator learns of any SAE, including a death, at any time after a subject has been discharged from the study, and he or she considers the event reasonably possibly related to the study treatment or study participation, the investigator must promptly notify Lilly.

9.2.1.1. Suspected Unexpected Serious Adverse Reactions

Suspected unexpected serious adverse reactions (SUSARs) are serious events that are not listed in the IB and that the investigator identifies as related to IP or procedure. United States 21 CFR 312.32 and European Union Clinical Trial Directive 2001/20/EC and the associated detailed guidance or national regulatory requirements in participating countries require the reporting of SUSARs. Lilly has procedures that will be followed for the recording and expedited reporting of SUSARs that are consistent with global regulations and the associated detailed guidance.

9.2.2. Adverse Events of Special Interest

Adverse events of special interest will include the following:

- infections (including tuberculosis, herpes zoster, or opportunistic infections)
- malignancies
- hepatic events (see Section 9.4.8)
- Major adverse cardiovascular events (MACE) (see Section 9.4.9)
- thrombotic events (such as deep vein thrombosis and pulmonary embolism)

Sites will provide details on these AEs as instructed on the eCRF and may be asked for additional description by Lilly.

9.2.3. Complaint Handling

Lilly collects product complaints on IPs and drug delivery systems used in clinical studies in order to ensure the safety of study participants, monitor quality, and facilitate process and product improvements.

Patients will be instructed to contact the investigator as soon as possible if he or she has a complaint or problem with the IP so that the situation can be assessed.

9.3. Treatment of Overdose

Refer to the IB.

9.4. Safety

Any clinically significant findings from electrocardiogram testing, physical examination, vital signs measurements, or laboratory measurements that result in a diagnosis and that occur after the patient receives the first dose of study treatment should be reported to Lilly or its designee as an AE via eCRF.

9.4.1. Vital Signs

For each patient, vital sign measurements should be conducted according to the Schedule of Activities (Section 2).

9.4.2. Physical Exam

For each patient, physical examinations will be conducted according to the Schedule of Activities (Section 2). Physical examinations throughout the study should include a symptom-directed physical examination. A complete physical examination may be conducted at the investigator's discretion at any time a patient presents with physical complaints.

9.4.3. Laboratory Tests

For each patient, laboratory tests detailed in Appendix 3, Appendix 4, Appendix 5, and Appendix 7 should be conducted according to the Schedule of Activities (Section 2). With the exception of laboratory test results that may unblind the study, Lilly or its designee will provide the investigator with the results of laboratory tests analyzed by a central vendor, if a central vendor is used for the clinical trial.

9.4.4. Hospital Anxiety Depression Scale

The Hospital Anxiety Depression Scale (HADS) is a 14-item self-assessment scale that determines the levels of anxiety and depression that a patient experienced over the past week. The HADS utilizes a 4-point Likert scale (e.g., 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).

9.4.5. Columbia-Suicide Severity Rating Scale

The C-SSRS captures the occurrence, severity, and frequency of suicidal ideation and/or behavior during the assessment period. The scale includes suggested questions to solicit the type of information needed to determine if suicidal ideation and/or behavior occurred. The C-SSRS is administered by an appropriately trained health care professional with at least 1 year of patient care/clinical experience. The tool was developed by the National Institute of Mental Health trial

group for the purpose of being a counterpart to the Columbia Classification Algorithm of Suicide Assessment categorization of suicidal events. For this study, the scale has been adapted (with permission from the scale authors) to include only the portion of the scale that captures the occurrence of the 11 preferred ideation and behavior categories.

The nonleading AE collection should occur prior to the collection of the C-SSRS. If a suicide-related event is discovered *during the C-SSRS* but was not captured during the nonleading AE collection, sites should not change the AE form. If an event is serious or leads to discontinuation, this is an exception where the SAE and/or AE leading to discontinuation should be included on the AE form and the process for reporting SAEs should be followed.

9.4.6. Self-Harm Supplement and Follow-Up Form

Suicide-related events (behavior and/or ideations) will be assessed and evaluated at every visit with the administration of the C-SSRS and the Self-Harm Supplement 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.

9.4.7. Hepatitis B Virus DNA Monitoring

Hepatitis B virus DNA testing will be performed in enrolled patients who tested positive for anti-hepatitis B core antibody (HBcAb) at screening for the originator study.

Patients who are HBcAb positive and HBV DNA negative (undetectable) at screening (Visit 1) for the originator study will require HBV DNA monitoring every 2 to 4 months and at the patients' last visit, regardless of their hepatitis B surface antibody (HBsAb) status.

The following actions should be taken in response to HBV DNA test results:

- If a single result is obtained with a value "below limit of quantitation," the test should be repeated within approximately 2 weeks. If the repeat test result is "target not detected," then patients will continue in JAHN and monitoring will resume as per the study schedule (Section 2).
- If the patient has 2 or more test results with a value "below limit of quantitation" or a test result above the limit of quantitation, the patient will be permanently discontinued from IP (see Section 8.1.2) and should be referred to a hepatology specialist.

9.4.8. Hepatic Safety Monitoring and Data Collection

If a study patient experiences elevated ALT ≥3 x ULN, ALP ≥2 x ULN, or elevated TBL ≥2 x ULN, liver testing (Appendix 4) should be repeated within 3 to 5 days including ALT, AST, ALP, TBL, direct bilirubin, gamma-glutamyl transferase, and creatine kinase to confirm the abnormality and to determine if it is increasing or decreasing. If the abnormality persists or worsens, clinical and laboratory monitoring should be initiated by the investigator in consultation

with the study medical monitor. Monitoring of ALT, AST, TBL, and ALP should continue until levels normalize or return to approximate baseline levels.

Discontinuation criteria of IPs, either temporary interruption or permanent discontinuation, due to abnormal ALT, AST, TBL, or ALP, are detailed in Section 8.1.

Additional safety data should be collected via the hepatic eCRF if 1 or more of the following conditions occur:

- elevation of serum ALT to ≥5 x ULN on 2 or more consecutive blood tests
- elevated serum TBL to ≥2 x ULN (except for cases of known Gilbert's syndrome)
- elevation of serum ALP to ≥2 x ULN on 2 or more consecutive blood tests
- patient discontinued from treatment due to a hepatic event or abnormality of liver tests
- hepatic event considered to be a SAE

See Section 8.1 for a description of hepatic laboratory values that are criteria for temporary interruption or permanent discontinuation of IP.

9.4.9. Safety Monitoring

Lilly will periodically review evolving aggregate safety data within the study by appropriate methods. Additionally, a DMC will oversee the conduct of this study and will periodically review safety data (Section 10.3.7.1). In the event that safety monitoring uncovers an issue that needs to be addressed by unblinding at the group level, only members of the data monitoring board (an advisory group for this study formed to protect the integrity of data; refer to Interim Analyses section [Section 10.3.7]) can conduct additional analyses of the safety data.

The Lilly CRP will monitor safety data throughout the course of the study. Lilly will review SAEs within time frames mandated by company procedures. The Lilly CRP will, as is appropriate, consult with the functionally independent Global Patient Safety therapeutic area physician or clinical scientist and periodically review trends in safety data and laboratory analytes. Any concerning trends in frequency or severity noted by an investigator and/or Lilly (or designee) may require further evaluation.

All deaths and SAE reports will be reviewed in a blinded manner by Lilly during the clinical trial. These reports will be reviewed to ensure completeness and accuracy but will not be unblinded to Lilly during the clinical trial. If a death or a clinical AE is deemed serious, unexpected, and possibly related to IP, only Lilly Global Patient Safety will be unblinded for regulatory reporting and safety monitoring purposes. These measures will preserve the integrity of the data collected during this trial and minimize any potential for bias while providing for appropriate safety monitoring.

Investigators will monitor vital signs and carefully review findings that may be associated with cardiovascular event and VTEs (Appendix 7). Adverse event reports and vital signs will be collected at each study visit. The cardiovascular monitoring plan includes the following:

- regular monitoring of lipid levels
- potential MACE (cardiovascular death, MI, stroke), other cardiovascular events (such as hospitalization for unstable angina, hospitalization for heart failure, serious arrhythmia, resuscitated sudden death, cardiogenic shock, coronary revascularization such as coronary artery bypass graft or percutaneous coronary intervention), venous thrombotic events, and noncardiovascular deaths will be identified by the investigative site or through medical review and will be sent to a blinded Clinical Event Committee for adjudication at regular intervals.

9.5. Pharmacokinetics

Not applicable.

9.6. Pharmacodynamics

Not applicable.

9.7. Pharmacogenetics

Not applicable.

9.8. Biomarkers

Biomarker research is performed to address questions of relevance to drug disposition, target engagement, pharmacodynamics, mechanism of action, variability of patient response (including safety), and clinical outcome. Sample collection is incorporated into clinical studies to enable examination of these questions through measurement of biomolecules including DNA, ribonucleic acid, proteins, lipids, and other cellular elements.

Blood samples for non-pharmacogenetic biomarker research will be collected at the times specified in the Schedule of Activities (Section 2) where local regulations allow.

Samples will be used for research on the drug target, disease process, variable response to baricitinib, pathways associated with AD, mechanism of action of baricitinib, and/or research method or in validating diagnostic tools or assay(s) related to AD.

All samples will be coded with the patient number. These samples and any data generated can be linked back to the patient only by the investigator site personnel.

Samples will be retained for a maximum of 15 years after the last patient visit for the study, or for a shorter period if local regulations and ethical review boards (ERBs) impose shorter time limits, at a facility selected by Lilly. This retention period enables use of new technologies, response to regulatory questions, and investigation of variable response that may not be observed until later in the development of baricitinib or after baricitinib becomes commercially available.

9.9. Medical Resource Utilization and Health Economics

Health Economics will be evaluated in this study utilizing the EQ-5D-5L and WPAI-AD (Section 9.1.4). Medical Resource Utilization parameters will not be evaluated in this study.

10. Statistical Considerations

10.1. Sample Size Determination

It is anticipated that 95% of enrolled patients will complete Studies JAHL, JAHM, and JAIY and roll over into Study JAHN. Therefore, planned enrollment into Study JAHN from the originating Studies JAHL, JAHM, and JAIY will be approximately 1425 patients. Of these patients, approximately 765 will start Study JAHN on baricitinib 4-mg, 547 on baricitinib 2-mg, 34 on baricitinib 1-mg, and 79 on placebo. Patients who are considered nonresponders at entry into Study JAHN will be randomized 1:1 to either baricitinib 4-mg or baricitinib 2-mg and stratified by IGA 0, 1, 2, IGA 3 and IGA 4. This study is meant to evaluate patients' long-term response of baricitinib and the sample sizes are not determined to detect differences between baricitinib and placebo in a statistically powered manner. Additional patients may enroll from addenda or other studies.

Patients at Week 52 will be stratified by responder status (IGA 0 or 1 versus IGA 2) when entering the randomized withdrawal and downtitration substudy. It is estimated that there will be approximately 600 patients entering into the randomized withdrawal and downtitration substudy. The substudy is meant to evaluate the change in clinical response after treatment withdrawal or downtitration and does not account for whether the sample size is sufficient to detect a difference between baricitinib and placebo. Maintenance of treatment benefit is defined as response of IGA 0, 1, or 2.

10.2. Populations for Analyses

Unless otherwise specified, the efficacy and health outcome analyses will be conducted on the modified intent-to-treat population, defined as all randomized patients, even if the patient does not receive the correct treatment, or otherwise did not follow the protocol and who received at least 1 dose of the IP in Study JAHN. Unless otherwise specified, safety analyses will be conducted on the safety population, which is defined as all enrolled patients who received at least 1 dose of the IP they were randomized to in Study JAHN. Patients will be analyzed for efficacy, health outcomes, and safety according to the treatment to which they were assigned. Further details will be described in the SAP, including, but not limited to, additional populations for the randomized withdrawal and downtitration substudy.

10.3. Statistical Analyses

10.3.1. General Statistical Considerations

Statistical analysis of this study will be the responsibility of Lilly or its designee. A detailed SAP describing the statistical methodologies will be developed by Lilly or its designee.

Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol, and the justification for making the change, will be described in the clinical study report. Additional exploratory analyses of the data will be conducted as deemed appropriate.

All tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, unless otherwise stated. Adjustment for multiple comparisons will not be carried out.

Primary and secondary discrete efficacy variables will be descriptively summarized by treatment group in terms of frequencies and percentages. Treatment comparisons of discrete efficacy variables between treatment groups may be made using a logistic regression analysis with disease severity (IGA 0 or 1 versus IGA 2), and treatment group in the model. Other factors may be included in the model. If the logistic regression model is performed, then the p-value from the logistic model, percentages, difference in percentages, and 100(1-alpha)% confidence interval (CI) of the difference in percentages using the Newcombe–Wilson method (Newcombe 1998) without continuity correction will be reported. The p-value from the Fisher exact test will also be produced.

Continuous efficacy variables will be descriptively summarized by treatment group in terms of number of patients, mean, standard deviation, median, minimum, and maximum. When evaluating these continuous measures over time, a restricted maximum likelihood-based mixed model for repeated measures (MMRM) may be used. The model will include treatment, baseline severity, visit, and treatment-by-visit-interaction as fixed categorical effects and baseline score and baseline score-by-visit-interaction as fixed continuous effects. Other factors may be included in the model. An unstructured (co)variance structure will be used to model the between- and within-patient errors. If this analysis fails to converge, other structures will be tested. The Kenward–Roger method will be used to estimate the degrees of freedom. Type III sums of squares for the least-square means (LSM) will be used for the statistical comparison; 95% CI will also be reported. Further details on the use of MMRM will be described in the SAP.

Treatment comparisons of continuous efficacy and health outcome variables may also be made using analysis of covariance (ANCOVA) with disease severity, treatment group, and baseline value in the model. Other factors may be included in the model. If the ANCOVA is performed, then Type III tests for the LSM will be used for statistical comparison between treatment groups. The LSM difference, standard error, p-value, and 100(1-alpha)% CI may also be reported. The method used to handle missing data will be specified in the SAP.

All safety data will be descriptively summarized by treatment groups. For categorical events, Fisher exact test may be used to perform comparisons between each baricitinib dose and the placebo group. Fisher exact test may also be used for the discontinuation, and other categorical safety data for between-treatment group comparisons. Continuous vital signs, body weight, and other continuous safety variables including laboratory variables will be analyzed by an ANCOVA with treatment and baseline value in the model. Shift tables for categorical safety analyses (e.g., “high” or “low” laboratory results) will also be produced.

Time-to-an-event analysis may be done and would be analyzed using a Cox proportional hazards model with treatment and stratification variables in the model. Hazard ratio with CIs may be reported. Kaplan–Meier curves may also be produced. Diagnostic tests for checking the validity of proportional hazards assumption may be done and these would be described in more detail in the SAP. If the assumption of proportional hazards is not justified, nonproportionality may be

modeled by stratification, as the most likely variable that interacts with time is categorical, that is, disease severity.

Missing data imputation:

1. Nonresponder imputation (NRI): All patients who either discontinue the study treatment or discontinue the study for any reason at any time will be defined as nonresponders for the NRI analysis for categorical variables such as IGA 0/1 or EASI 50/75/90 after rescue or discontinuation and onward.
2. MMRM: Continuous variables such as EASI and SCORAD scores will be assumed to be missing after rescue or discontinuation and then a MMRM analysis will be performed.
3. Last observed carried forward: An additional analysis may be done that uses the last observed value on or prior to discontinuation or rescue therapy. This may then be analyzed using a logistic model for categorical variables or ANCOVA for continuous variables as described above.

Additional sensitivity analyses for the primary and key secondary endpoints such as tipping point analyses as well as a reference-based multiple imputation method may be done and will be specified in the SAP.

10.3.2. Treatment Group Comparability

10.3.2.1. Patient Disposition

All patients who discontinue from the study or the study treatment will be identified, along with their reason for discontinuation. Reasons for discontinuation from the study will be summarized by treatment group. This will be done for Study JAHN as well as the randomized withdrawal and downtitration substudy. Additional summaries that may be produced for other populations will be documented in the SAP. No formal statistical comparisons will be made among treatment groups.

10.3.2.2. Patient Characteristics

Demographic and baseline characteristics will be summarized descriptively by treatment group for Study JAHN and for the randomized withdrawal and downtitration substudy. Additional summaries that may be produced for other populations will be documented in the SAP. Descriptive statistics including number of patients, mean, standard deviation, median, minimum, and maximum will be provided for continuous measures, and frequency counts and percentages will be tabulated for categorical measures. No formal statistical comparisons will be made among treatment groups unless otherwise stated.

10.3.2.3. Concomitant Therapy

Concomitant medications will be descriptively summarized by treatment group in terms of frequencies and percentages. The medications will be coded accordingly.

10.3.2.4. Treatment Compliance

Treatment compliance with the randomly assigned study medication will be evaluated at every clinic visit through the counts of returned study drug tablets. A patient will be considered significantly noncompliant if he or she misses more than 20% of the prescribed doses during the

study, that is, compliance <80%, unless the patient's IP is withheld by the investigator. Similarly, a patient will be considered significantly noncompliant if he or she is judged by the investigator to have intentionally or repeatedly taken more than the prescribed amount of medication, that is, compliance ≥120%.

10.3.3. Efficacy Analyses

Analyses will be conducted in Study JAHN from Week 0 to 52 as well as for the randomized withdrawal and downtitration substudy. Discrete efficacy variables will be descriptively summarized by treatment group in terms of frequencies and percentages. Treatment comparisons of categorical variables, such as IGA, will be analyzed using a logistic regression model described above. Other dichotomous secondary endpoints will be analyzed similarly. Nonresponder imputation will be used as described above. Additionally, time to retreatment may be analyzed using Cox proportion hazard model along with Kaplan–Meier curves. If the assumption of proportional hazards is not justified, analysis will proceed using the proportion of patients who meet criteria for loss of treatment benefit. Continuous efficacy variables will be descriptively summarized by treatment group in terms of number of patients, mean, standard deviation, median, minimum, and maximum. For continuous variables, data after discontinuation or rescue will be assumed to be missing and data may be analyzed using a MMRM model as described in Section 10.3.1.

10.3.4. Safety Analyses

All safety data will be descriptively summarized by treatment groups and analyzed using the safety population, unless otherwise stated.

Treatment-emergent adverse events (TEAEs) are defined as AEs that either first occurred or worsened in severity after the first dose of study treatment in Study JAHN. The number of TEAEs as well as the number and percentage of patients who experienced at least 1 TEAE will be summarized using Medical Dictionary for Regulatory Activities for each system organ class (or a body system) and each preferred term by treatment group. Serious adverse events and AEs that lead to study drug discontinuation will also be summarized by treatment group and treatment period. Fisher exact test may be used to perform comparisons between treatment groups. Further details will be given in the SAP.

All clinical laboratory results will be descriptively summarized by treatment group and treatment period. Individual results that are outside of normal reference ranges will be flagged in data listings. Quantitative clinical hematology, chemistry, and urinalysis variables obtained at baseline to postbaseline visits will be summarized as changes from baseline by treatment group and treatment period and may be analyzed using ANCOVA with treatment and baseline value in the model. Categorical variables, including the incidence of abnormal values and incidence of adverse events of special interest, will be summarized by frequency and percentage of patients in corresponding categories. Shift tables will be presented for selected measures.

Observed values and changes from baseline (predose or screening if missing) for vital signs and physical characteristics will be descriptively summarized by treatment group and treatment

period. Change from baseline to postbaseline in vital signs, and body weight may be analyzed using ANCOVA with treatment and baseline value in the model.

The incidence and average duration of study drug interruptions may be summarized and compared descriptively among treatment group and treatment period. Various techniques may be used to estimate the effects of study drug interruptions on safety measures. Further analyses may be performed and will be planned in the SAP.

Data collected after initiation of rescue therapy will be summarized as appropriate.

10.3.5. Pharmacokinetic/Pharmacodynamic Analyses

Not applicable.

10.3.6. Other Analyses

The health outcome measures will be analyzed using methods described for continuous or categorical data as described for efficacy measures in Section 10.3.3. More detailed analytical methods will be described in the SAP.

10.3.6.1. Subgroup Analyses

To assess whether the treatment effect is similar across subgroups, a logistic model will be used and will include treatment, stratification variables, the subgroup variable (e.g., sex), and the subgroup by treatment interaction. If the interaction is statistically significant at alpha=0.10, the nature of the interaction will be explored, that is, within each subgroup the treatment effect will be estimated. Similarly, for continuous variable, the MMRM model will include additional variables for subgroup and the subgroup by treatment interaction.

Subgroups to be evaluated may include region, baseline severity, sex, age, race, prior therapy, etc. Further definitions for the levels of the subgroup variables, the analysis methodology, and any additional subgroup analyses will be defined in the SAP. All subgroup analyses will be treated as exploratory.

10.3.7. Interim Analyses

10.3.7.1. Data Monitoring Committee

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

The DMC will be authorized to review unblinded results of analyses by treatment group prior to database lock, including study discontinuation data, AEs including SAEs, clinical laboratory data, vital sign data, etc. 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. 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. However, the

study will not be stopped for positive efficacy results, and there is no planned futility assessment. Hence, no alpha is spent. As an extension of the DMC for the originating studies, interim analyses will be conducted for the DMC review on an approximate semiannual basis (this timing may be modified based on recruitment rates in originating studies). Furthermore, details of the DMC will be documented in a DMC charter and interim analysis plan contained in the SAP.

Besides DMC members, a limited number of pre-identified individuals may gain access to the limited unblinded data, as specified in the unblinding plan, prior to the final database lock, in order to initiate work for regulatory submission. Interim locks may be conducted at various timepoints to support regulatory activities and scientific disclosures. The timing of the data lock(s) for the analysis of the efficacy data from the sub-study will be determined by the retreatment rates. Information that may unblind the study during the analyses will not be reported to study sites or blinded study team until the study has been unblinded.

Unblinding details are specified in a separate unblinding plan document.

10.3.7.2. Adjudication Committee

A blinded Clinical Event Committee will adjudicate potential MACE (cardiovascular death, MI, stroke), other cardiovascular events (such as hospitalization for unstable angina, hospitalization for heart failure, serious arrhythmia, resuscitated sudden death, cardiogenic shock, coronary revascularization such as coronary artery bypass graft or percutaneous coronary intervention), venous thrombotic events, and noncardiovascular deaths. Details of membership, operations, recommendations from the Committee, and the communication plan will be documented in the Charter.

11. References

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

Bieber T. Atopic dermatitis. *Ann Derm*. 2010;22(2):125-137. Bieber T, Straeter B. Off-label prescriptions for atopic dermatitis in Europe. *Allergy*. 2015;70(1):6-11.

Charman CR, Venn AJ, Williams HC. The Patient-Oriented Eczema Measure. *Arch Dermatol.* 2004;140:1513-1519.

Clark JD, Flanagan ME, Telliez JB. Discovery and development of Janus kinase (JAK) inhibitors for inflammatory diseases. *J Med Chem*. 2014;57(12):5023-5038.

Divekar R, Kita H. Recent advances in epithelium-derived cytokines (IL-33, IL-25, and thymic stromal lymphopoietin) and allergic inflammation. *Curr Opin Allergy Clin Immunol*. 2015;15(1):98-103.

Dougados M, van der Heijde D, Chen YC, Greenwald M, Drescher E, Liu J, Beattie S, Witt S, de la Torre I, Gaich C, Rooney T, Schlichting D, de Bono S, Emery P. Baricitinib in patients with inadequate response or intolerance to conventional synthetic DMARDs: results from the RA-BUILD study. *Ann Rheum Dis*. 2017;76(1):88-95.

EuroQol Group. EQ-5D-5L User Guide. Version 2.1. April 2015. Available at: http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-5L_UserGuide_2015.pdf. Published April 2015. Accessed March 1, 2017.

Fleischmann R, Schiff M, van der Heijde D, Ramos-Remus C, Spindler A, Stanislav M, Zerbini CA, Gurbuz S, Dickson C, de Bono S, Schlichting D, Beattie S, Kuo WL, Rooney T, Macias W, Takeuchi T. Baricitinib, methotrexate, or combination in patients with rheumatoid arthritis and no or limited prior disease-modifying antirheumatic drug treatment. *Arthritis Rheumatol*. 2017;69(3):506-517.

Fridman JS, Scherle PA, Collins R, Burn TC, Li Y, Li J, Covington MB, Thomas B, Collier P, Favata MF, Wen X, Shi J, McGee R, Haley PJ, Shepard S, Rodgers JD, Yeleswaram S, Hollis G, Newton RC, Metcalf B, Friedman SM, Vaddi K. Selective inhibition of JAK1 and JAK2 is efficacious in rodent models of arthritis: preclinical characterization of INCB028050. *J Immunol*. 2010;184(9):5298-5307.

Genovese MC, Kremer J, Zamani O, Ludivico C, Krogulec M, Xie L, Beattie SD, Koch AE, Cardillo TE, Rooney TP, Macias WL, de Bono S, Schlichting DE, Smolen JS. Baricitinib in patients with refractory rheumatoid arthritis. *N Engl J Med*. 2016;374(13):1243-1252.

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.

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.

Krakowski AC, Eichenfield LF, Dohil MA. Management of atopic dermatitis in the pediatric population. *Pediatrics*. 2008;122(4):812-824.

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

Langenbruch A, Radtke M, Franzke N, Ring J, Foelster-Holst R, Augustin M. Quality of health care of atopic eczema in Germany: results of the national health care study AtopicHealth. *J Eur Acad Dermatol Venereol*. 2014;28(6):719-726.

Li M, Messaddeq N, Teletin M, Pasquali JL, Metzger D, Chambon P. Retinoid X receptor ablation in adult mouse keratinocytes generates an atopic dermatitis triggered by thymic stromal lymphopoietin. *Proc Natl Acad Sci U S A*. 2005;102(41):14795-14800.

Moniaga CS, Jeong SK, Egawa G, Nakajima S, Hara-Chikuma M, Jeon JE, Lee SH, Hibino T, Miyachi Y, Kabashima K. Protease activity enhances production of thymic stromal lymphopoietin and basophil accumulation in flaky tail mice. *Am J Pathol*. 2013;182(3):841-851.

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.

Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med*. 1998 Apr 30;17(8):873-90. Erratum in: *Stat Med* 1999 May 30;18(10):1293.

Nomura T and Kabashima K. Advances in atopic dermatitis in 2015. *J Allergy Clin Immunol*. 2015;138(6):1548-1555.

Palmer CN, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, Goudie DR, Sandilands A, Campbell LE, Smith FJ, O'Regan GM, Watson RM, Cecil JE, Bale SJ, Compton JG, DiGiovanna JJ, Fleckman P, Lewis-Jones S, Arseculeratne G, Sergeant A, Munro CS, El Houate B, McElreavey K, Halkjaer LB, Bisgaard H, Mukhopadhyay S, McLean WH. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet.* 2006;38:441-446.

Papp KA, Menter MA, Raman M, Disch D, Schlichting DE, Gaich C, Macias W, Zhang X, Janes JM. A randomized phase 2b trial of baricitinib, an oral Janus kinase (JAK) 1/JAK2 inhibitor, in patients with moderate-to-severe psoriasis. *Br J Dermatol*. 2016;174(6):1266-1276.

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

Schram ME, Spuls PI, Leeflang MM, Lindeboom R, Bos J, Schmitt J. EASI, (objective) SCORAD and POEM for atopic eczema:responsiveness and minimal clinically important difference. *Allergy*. 2012;67(1):99-106.Simpson EL. Comorbidity in atopic dermatitis. *Curr Dermatol Rep*. 2012;1(1):29-38.

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

Stalder JF, Taïeb A, Atherton DJ, et al. for the European Task Force on Atopic Dermatitis. Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. *Dermatology*. 1993;186(1):23-31.

Tadicherla S, Ross K, Shenefelt PD, Fenske NA. Topical Corticosteroids in Dermatology. *J Drugs Dermatol.append* 2009;8(12):1093-1105.

Taylor PC, Keystone EC, van der Heijde D, Weinblatt ME, del Carmen Morales L, Gonzaga JR, Yakushin S, Ishii T, Emoto K, Beattie S, Arora V, Gaich CL, Rooney T, Schlichting D, Macias WL, de Bono S, Tanaka. Baricitinib versus placebo or adalimumab in rheumatoid arthritis. *N Engl J Med.* 2017;376:652-662.

Tuttle KR, Brosius FC, Adler SG, Kretzler M, Mehta RL, Tumlin JA, Liu J, Silk ME, Cardillo TE, Duffin KL, Haas JV, Macias WL, Janes JM. Baricitinib in diabetic kidney disease: results from a phase 2, multicenter, randomized, double-blind, placebo-controlled study. Paper presented at: Annual Meeting of the American Diabetes Association; 5-9 June 2015; Boston, MA.

Vaddi K, Luchi M. JAK inhibition for the treatment of rheumatoid arthritis: a new era in oral DMARD therapy. *Expert Opin Investig Drugs*. 2012;21(7):961-973.

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.

[WHO] World Health Organization. Model Prescribing Information: Drugs used in skin diseases. 1997; Geneva.

Wilson SR, Thé L, Batia LM, Beattie K, Katibah GE, McClain SP, Pellegrino M, Estandian DM, Bautista DM. The epithelial cell-derived atopic dermatitis cytokine TSLP activates neurons to induce itch. *Cell*. 2013;155(2):285-295.

Yano C, Saeki H, Ishiji T, Ishiuji Y, Sato J, Tofuku Y, Nakagawa H. Impact of disease severity on sleep quality in Japanese patients with atopic dermatitis. *J Dermatol Sci*. 2013;72(2):195-197.

Ziegler SF, Roan F, Bell BD, Stoklasek TA, Kitajima M, Han H. The biology of thymic stromal lymphopoietin (TSLP). *Adv Pharmacol*. 2013;66:129-155.

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

12. Appendices

Appendix 1. Abbreviations and Definitions

Term	Definition
AD	atopic dermatitis
ADSS	Atopic Dermatitis Sleep Scale
AE	adverse event: Any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product that does not necessarily have a causal relationship with this treatment. An adverse event can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not related to the medicinal (investigational) product.
ALP	alkaline phosphatase
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
blinding/masking	A single-blind study is one in which the investigator and/or his staff are aware of the treatment but the patient is not, or vice versa, or when the sponsor is aware of the treatment but the investigator and/his staff and the patient are not. A double-blind study is one in which neither the patient nor any of the investigator or sponsor staff who are involved in the treatment or clinical evaluation of the subjects are aware of the treatment received.
CI	confidence interval
complaint	A complaint is any written, electronic, or oral communication that alleges deficiencies related to the identity, quality, purity, durability, reliability, safety or effectiveness, or performance of a drug or drug delivery system.
CRP	clinical research physician: Individual responsible for the medical conduct of the study. Responsibilities of the CRP may be performed by a physician, clinical research scientist, global safety physician or other medical officer.
C-SSRS	Columbia–Suicide Severity Rating Scale
CSR	clinical study report
DLQI	Dermatology Life Quality Index
DMC	Data Monitoring Committee
EASI	Eczema Area and Severity Index
eCOA	electronic clinical outcome assessment

eCRF	electronic case report form
eGFR	estimated glomerular filtration rate
enroll	The act of assigning a patient to a treatment. Patients who are enrolled in the trial are those who have been assigned to a treatment.
Enter	Patients entered into a trial are those who sign the informed consent form directly or through their legally acceptable representatives.
EQ-5D-5L	European Quality of Life–5 Dimensions–5 Levels
ERB	ethical review board
ETV	early termination visit
FDA	the Food and Drug Administration
GCP	good clinical practice
HADS	Hospital Anxiety Depression Scale
HBcAb	anti-hepatitis B core antibody
HBV	hepatitis B virus
IB	Investigator's Brochure
ICF	informed consent form
ICH	International Council for Harmonisation
IGA	Investigator's Global Assessment
IL	interleukin
INR	international normalized ratio
investigational product (IP)	A pharmaceutical form of an active ingredient or placebo being tested or used as a reference in a clinical trial, including products already on the market when used or assembled (formulated or packaged) in a way different from the authorized form, or marketed products used for an unauthorized indication, or marketed products used to gain further information about the authorized form.
IWRS	interactive web-response system
JAK	Janus kinase
LSM	least squares mean
MACE	major adverse cardiovascular events
MI	myocardial infarction

MMRM	mixed model for repeated measures
NRI	nonresponder imputation
NRS	numeric rating scale
PK	pharmacokinetic
POEM	Patient-Oriented Eczema Measure
PRO/ePRO	patient-reported outcome/electronic patient-reported outcome
QD	once daily
QoL	quality of life
RA	rheumatoid arthritis
SAE	serious adverse event
SAP	statistical analysis plan
SCORAD	SCORing Atopic Dermatitis
screen	The act of determining if an individual meets minimum requirements to become part of a pool of potential candidates for participation in a clinical study.
SUSAR	suspected unexpected serious adverse reaction
TBL	total bilirubin level
TCNI	topical calcineurin inhibitor
TCS	topical corticosteroid(s)
TEAE	treatment-emergent adverse event: An untoward medical occurrence that emerges during a defined treatment period, having been absent pretreatment, or worsens relative to the pretreatment state, which does not necessarily have to have a causal relationship with this treatment.
TSLP	thymic stromal lymphopoietin
TYK2	tyrosine kinase 2
ULN	upper limit of normal
VAS	visual analog scale
vIGA-AD	Validated Investigator's Global Assessment for Atopic Dermatitis (referred to as the IGA throughout the protocol)
VTE	venous thromboembolic event (deep vein thrombosis or pulmonary embolism)
WPAI-AD	Work Productivity and Activity Impairment Questionnaire – Atopic Dermatitis

Appendix 2. Study Governance Considerations

Appendix 2.1. Regulatory and Ethical Considerations, Including the Informed Consent Process

Appendix 2.1.1. Informed Consent

The investigator is responsible for ensuring:

- that the patient understands the nature of the study, the potential risks and benefits of participating in the study, and that their participation is voluntary.
- that informed consent is given by each patient. This includes obtaining the appropriate signatures and dates on the informed consent form (ICF) prior to the performance of any protocol procedures and prior to the administration of investigational product (IP).
- answering any questions the patient may have throughout the study and sharing in a timely manner any new information that may be relevant to the patient's willingness to continue his or her participation in the trial.
- that a copy of the ICF is provided to the participant or the participant's legal representative and is kept on file.
- that the medical record includes a statement that written informed consent was obtained before the participant was enrolled in the study and the date the written consent was obtained. The authorized person obtaining the informed consent must also sign the ICF.

Appendix 2.1.2. Ethical Review

The investigator must give assurance that the ethical review board (ERB) was properly constituted and convened as required by International Council for Harmonisation (ICH) guidelines and other applicable laws and regulations.

Documentation of ERB approval of the protocol and the ICF and Assent Form must be provided to Lilly before the study may begin at the investigative site(s). Lilly or its representatives must approve the ICF, including any changes made by the ERBs, before it is used at the investigative site(s). All ICFs must be compliant with the ICH guideline on good clinical practice (GCP).

The study site's ERB(s) should be provided with the following:

- the current Investigator's Brochure (IB) and updates during the course of the study
- ICF and Assent Form
- relevant curricula vitae

Appendix 2.1.3. Regulatory Considerations

This study will be conducted in accordance with:

- consensus ethics principles derived from international ethics guidelines, including the Declaration of Helsinki and Council for International Organizations of Medical Sciences International Ethical Guidelines
- applicable ICH GCP Guidelines
- applicable laws and regulations

Some of the obligations of the sponsor will be assigned to a third party.

Appendix 2.1.4. Investigator Information

Physicians with a specialty in dermatology will participate as investigators in this clinical trial.

Appendix 2.1.5. Protocol Signatures

The sponsor's responsible medical officer will approve the protocol, confirming that, to the best of his or her knowledge, the protocol accurately describes the planned design and conduct of the study.

After reading the protocol, each principal investigator will sign the protocol signature page and send a copy of the signed page to a Lilly representative.

Appendix 2.1.6. Final Report Signature

Lilly will select a qualified investigator(s) from among investigators participating in the design, conduct, and/or analysis of the study to serve as the clinical study report (CSR) coordinating investigator. If this investigator is unable to fulfill this function, another investigator will be chosen by Lilly to serve as the CSR coordinating investigator.

The CSR coordinating investigator will sign the final CSR for this study, indicating agreement that, to the best of his or her knowledge, the report accurately describes the conduct and results of the study.

The sponsor's responsible medical officer and statistician will approve the final CSR for this study, confirming that, to the best of his or her knowledge, the report accurately describes the conduct and results of the study.

Appendix 2.2. Data Quality Assurance

To ensure accurate, complete, and reliable data, Lilly or its representatives will do the following:

- provide instructional material to the study sites, as appropriate
- sponsor start-up training to instruct the investigators and study coordinators. This training will give instruction on the protocol, the completion of the case report forms (CRFs), and study procedures.
- make periodic visits to the study site

- be available for consultation and stay in contact with the study site personnel by mail, telephone, and/or fax
- review and evaluate CRF data and use standard computer edits to detect errors in data collection
- conduct a quality review of the database

In addition, Lilly or its representatives will periodically check a sample of the patient data recorded against source documents at the study site. The study may be audited by Lilly or its representatives, and/or regulatory agencies at any time. Investigators will be given notice before an audit occurs.

The investigator will keep records of all original source data. This might include laboratory tests, medical records, and clinical notes. If requested, the investigator will provide the sponsor, applicable regulatory agencies, and applicable ERBs with direct access to original source documents.

Appendix 2.2.1. Data Capture System

An electronic data capture system will be used in this study. The site maintains a separate source for the data entered by the site into the sponsor-provided electronic data capture system.

Electronic patient-reported outcome (ePRO) measures (e.g., a rating scale) and electronic clinical outcome assessments (eCOAs) are entered into an ePRO/eCOA instrument at the time that the information is obtained. In these instances, where there is no prior written or electronic source data at the site, the ePRO/eCOA instrument record will serve as the source.

If ePRO/eCOA records are stored at a third party site, investigator sites will have continuous access to the source documents during the study and will receive an archival copy at the end of the study for retention.

Any data for which the ePRO/eCOA instrument record will serve to collect source data will be identified and documented by each site in that site's study file.

Case report form data will be encoded and stored in InForm. Data managed by a central vendor, such as laboratory test data, will be stored electronically in the central vendor's database system. Data will subsequently be transferred from the central vendor to the Lilly data warehouse.

Data managed by a central vendor, such as laboratory test data, will be stored electronically in the central vendor's database system. Data will subsequently be transferred from the central vendor to the Lilly data warehouse.

Data from complaint forms submitted to Lilly will be encoded and stored in the global product complaint management system.

Appendix 2.3. Study and Site Closure

Appendix 2.3.1. Discontinuation of Study Sites

Study site participation may be discontinued if Lilly, the investigator, or the ERB of the study site judges it necessary for medical, safety, regulatory, or other reasons consistent with applicable laws, regulations, and GCP.

Appendix 2.3.2. Discontinuation of the Study

The study will be discontinued if Lilly judges it necessary for medical, safety, regulatory, or other reasons consistent with applicable laws, regulations, and GCP. Study termination may occur in a specific country or region when baricitinib is approved for the treatment of AD and becomes reimbursed or commercially available in that country or region, or a negative regulatory opinion is received in that country or region.

Appendix 3. Clinical Laboratory Tests

Hematology[a,b]	**Clinical Chemistry**[a,b]
Hemoglobin	**Serum Concentrations of:**
Hematocrit	Sodium
Erythrocyte count (RBC)	Potassium
Absolute reticulocyte count	Total bilirubin
Mean cell volume	Direct bilirubin
Mean cell hemoglobin	Alkaline phosphatase
Mean cell hemoglobin concentration	Alanine aminotransferase (ALT)
Leukocytes (WBC)	Aspartate aminotransferase (AST)
Platelets	Blood urea nitrogen (BUN)
	Creatinine
Absolute counts of:	Cystatin C
Neutrophils, segmented	Uric acid
Neutrophils, juvenile (bands)	Calcium
Lymphocytes	Glucose
Monocytes	Albumin
Eosinophils	Total protein
Basophils	Estimated glomerular filtration rate (eGFR)[c]
	Creatine phosphokinase (CPK)
Urinalysis[a,b,d]	
Color	**Lipids**[a,e]
Specific gravity	Total cholesterol
pH	Low-density lipoprotein
Protein	High-density lipoprotein
Glucose	Triglycerides
Ketones	
Bilirubin	**Other Tests**[a]
Urobilinogen	Exploratory storage samples (serum, plasma, and mRNA)
Blood	Pregnancy Test[f]
Leukocyte esterase	HBV DNA[g]
Nitrite	Serum immunoglobulin (IgE)

Abbreviations: HBcAb = anti-hepatitis B core antibody; HBV = hepatitis B virus; mRNA = messenger ribonucleic acid; RBC = red blood cell; WBC = white blood cell.

a Assayed by sponsor-designated laboratory.

b Unscheduled or repeat blood chemistry, hematology, and urinalysis panels may be performed at the discretion of the investigator, as needed.

c eGFR from serum creatinine calculated by the central laboratory using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) Creatinine 2009 equation.

d Microscopic examination of sediment performed only if abnormalities are noted on the routine urinalysis.

e Fasting lipid profile. Patients should not eat or drink anything except water for 12 hours prior to test. If a patient attends these visits in a nonfasting state, this will not be considered a protocol violation.

f For all women of childbearing potential, local urine pregnancy tests will be performed at all visits.

g HBV DNA testing will be done in those patients who are HBcAb+ at screening.

Appendix 4. Hepatic Monitoring Tests for Treatment-Emergent Abnormality

Selected tests may be obtained in the event of a treatment-emergent hepatic abnormality and may be required in follow-up with patients in consultation with the Lilly, or its designee, CRP.

Hepatic Monitoring Tests

Hepatic Hematology[a]	**Haptoglobin[a]**
Hemoglobin	
Hematocrit	**Hepatic Coagulation[a]**
RBC	Prothrombin time
WBC	Prothrombin time, INR
Neutrophils, segmented	
Lymphocytes	**Hepatic Serologies[a,b]**
Monocytes	Hepatitis A antibody, total
Eosinophils	Hepatitis A antibody, IgM
Basophils	Hepatitis B surface antigen
Platelets	Hepatitis B surface antibody
	Hepatitis B Core antibody
Hepatic Chemistry[a]	Hepatitis C antibody
Total bilirubin	Hepatitis E antibody, IgG
Direct bilirubin	Hepatitis E antibody, IgM
Alkaline phosphatase	
ALT	**Anti-nuclear antibody[a]**
AST	
GGT	**Alkaline Phosphatase Isoenzymes[a]**
CPK	
	Anti-smooth muscle antibody (or anti-actin antibody)[a]

Abbreviations: ALT = alanine aminotransferase; AST = aspirate aminotransferase; CPK = creatinine phosphokinase; GGT = gamma-glutamyl transferase; Ig = immunoglobulin; INR = international normalized ratio; RBC = red blood cells; WBC = white blood cells.

[a] Assayed by Lilly-designated or local laboratory.

[b] Reflex/confirmation dependent on regulatory requirements and/or testing availability.

Appendix 5. Liver Function Testing and Hepatic Safety Monitoring

Liver Function Testing and Hepatic Safety Monitoring

Analyte	Exclusion Criteria	Additional Hepatic Testing	Hepatic eCRF Reporting	Temporary Interruption of IP	Permanent Discontinuation of IP after Consultation with the Lilly-Designated Medical Monitor
Protocol section	Section 6.2	Section 9.4.8	Section 9.4.8	Section 8.1.1	Section 8.1.2
ALT/ AST	≥2 x ULN	ALT ≥3 x ULN	ALT ≥5 x ULN on ≥2 consecutive tests	≥5 x ULN	• >8 x ULN • >5 x ULN for >2 weeks • >3 x ULN AND TBL >2 x ULN or INR >1.5 • >3 x ULN with symptoms[a]
ALP	≥2x ULN	≥2x ULN	≥2x ULN on ≥2 consecutive tests	No applicable criteria	• >3 x ULN • >2.5 x ULN AND TBL >2 x ULN • >2.5 x ULN with symptoms[a]
TBL	≥1.5 x ULN	≥2 x ULN	≥2 x ULN (excluding Gilbert's syndrome)	No applicable criteria	• ALT or AST >3 x ULN AND TBL >2 x ULN • ALP >2.5 x ULN AND TBL >2 x ULN

Abbreviations: ALP = alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase; INR = international normalized ratio; IP = investigational product; TBL = total bilirubin level; ULN = upper level of normal.

[a] Fatigue, nausea, vomiting, right upper-quadrant pain or tenderness, fever, and/or rash.

Appendix 6. Classification of Potency for Topical Corticosteroids

Potency	Class	Topical Corticosteroid	Formulation
Ultra high	I	Clobetasol propionate	Cream 0.05%
		Diflorasone diacetate	Ointment 0.05%
High	II	Amcinonide	Ointment 0.1%
		Betamethasone dipropionate	Ointment 0.05%
		Desoximetasone	Cream or ointment 0.025%
		Fluocinonide	Cream, ointment, or gel 0.05%
		Halcinonide	Cream 0.1%
	III	Betamethasone dipropionate	Cream 0.05%
		Betamethasone valerate	Ointment 0.1%
		Diflorasone diacetate	Cream 0.05%
		Triamcinolone acetonide	Ointment 0.1%
Moderate	IV	Desoximetasone	Cream 0.05%
		Fluocinolone acetonide	Ointment 0.025%
		Fludroxycortide	Ointment 0.05%
		Hydrocortisone valerate	Ointment 0.2%
		Triamcinolone acetonide	Cream 0.1%
	V	Betamethasone dipropionate	Lotion 0.02%
		Betamethasone valerate	Cream 0.1%
		Fluocinolone acetonide	Cream 0.025%
		Fludroxycortide	Cream 0.05%
		Hydrocortisone butyrate	Cream 0.1%
		Hydrocortisone valerate	Cream 0.2%
		Triamcinolone acetonide	Lotion 0.1%
Low	VI	Betamethasone valerate	Lotion 0.05%
		Desonide	Cream 0.05%
		Fluocinolone acetonide	Solution 0.01%
	VII	Dexamethasone sodium phosphate	Cream 0.1%
		Hydrocortisone	Lotion, cream, or ointment 2.5%
		Hydrocortisone acetate	Cream 1%
		Methylprednisolone acetate	Cream 0.25%

Source: WHO (1997) and Tadicherla et al. (2009).

Appendix 7. Monitoring Tests for Confirmed VTE

Selected tests may be obtained in patients with a venous thromboembolic event (VTE) in consultation with Eli Lilly and Company, its designee, or the clinical research physician. The choice and optimal timing of these tests will be directed by the patient's management and may require ongoing follow-up after study discontinuation.

Protein C Functional
Protein S Clottable
Antithrombin III
APC Resistance
PT
APTT
Fibrinogen
Cardiolipin Antibodies
PT Gene
Factor VIII C Assay
Hexagonal Phase Phospholipid Neutralization
C-Reactive Protein
PTT Incubated Mixing
Dilute Russell Viper Venom
Platelet Neutralization
Factor V Leiden
MTHFR
Thrombin Time
Reptilase
Fibrinogen Antigen
Protein C Immunologic
Protein S Immunologic
Heparin fXa Inhibition

Abbreviations: APC = activated protein C; APTT = activated partial thromboplastin time; fXa = factor Xa; MTHFR = methylene tetrahydrofolate reductase; PT = prothrombin time; PTT = partial thromboplastin time.

Appendix 8. Protocol Amendment I4V-MC-JAHN(c) Summary - A Phase 3 Multicenter, Double-Blind Study to Evaluate the Long-Term Safety and Efficacy of Baricitinib in Adult Patients with Atopic Dermatitis

Overview

Protocol I4V-MC-JAHN, A Phase 3 Multicenter, Double-Blind Study to Evaluate the Long-Term Safety and Efficacy of Baricitinib in Adult Patients with Atopic Dermatitis, has been amended. This amendment is substantial based on the criteria set forth in Article 10(a) of Directive 2001/20EC of the European Parliament and the council of the European Union. The new protocol is indicated by amendment (c) and will be used to conduct the study in place of any preceding version of the protocol.

The overall changes and rationale for the changes made to this protocol are described in the following table. Other minor typographical corrections and minor formatting changes not affecting content have been made in the document, and these changes are not identified in the amendment summary table or with strikethrough/underscore.

Amendment Summary for Protocol I4V-MC-JAHN Amendment (c)

Section # and Name	Description of Change	Brief Rationale
Title, Synopsis, Section 2, Table 1 - Schedule of Activities, and Section 4	Included an additional six visits to reflect the 2-year extension to the study. Added additional exploratory endpoint reflecting extension of treatment period 2 up to 200 weeks.	This will allow for additional long-term safety information to be collected and provide patients the opportunity to continue study treatment until the anticipated approval of baricitinib in this indication. This endpoint was added to reflect that additional efficacy analyses may be performed beyond Week 104 up to 200 weeks. The exact timing of analyses will be determined by the retreatment rates within the sub-study.
Section 5.1 and Figure JAHN 1	Reflected the change of the extension of treatment period 2 by approximately two years. Figure was updated to include the additional visits.	Updated language to incorporate additional patient visits. Additional visits were incorporated into Figure JAHN 1.
Section 5.1.3, 5.1.4, and 5.4	Updated visits that were added.	Reflected the extension of patient visits in treatment period 2.
Section 7.2	Updated week visit of when IP will be last distributed.	Reflected change in visit schedule (from Week 104 to 184).
Section 7.6	Treatment compliance updated.	The calculation of treatment

Section # and Name	Description of Change	Brief Rationale
		compliance was updated to include visits 17-22/ET.
Section 7.7.1	Updated language to reflect that TCS will not be provided or reimbursed by the sponsor for the additional visits (TCS supplied by the sponsor only during the first 2 years).	The purpose for this extension is to allow patients the opportunity to continue to receive baricitinib while allowing clinicians and patients freedom to use TCS product/dosing form (e.g., cream, ointment) that is preferred and is available as part of local standard of care.
Section 7.7.3	Removed leukotriene inhibitors from prohibited medications. Removed allergen immunotherapy from prohibited medications.	Updated to align with JAIY protocol. Evidence suggests leukotriene inhibitors have little impact to AD. Due to the extended duration of the study, the sponsor anticipates some patients may require allergen immunotherapy for allergic conditions.
Section 8.1.1, 8.1.2, and 9.4.9	Patients will be permanently discontinued after one VTE instead of two.	Discontinuation criteria for VTEs were updated to align with current safety protocols within the Atopic dermatitis and Systemic Lupus Erythematosus (SLE) programs for baricitinib.
Section 9.4.9	Updated cardiovascular monitoring plan.	This language was updated to reflect accurate details of the cardiovascular monitoring plan, as designated by the originating studies.
Section 8.2, and Appendix 2.3.2	Added study discontinuation criterion.	In line with the 2-year extension, providing patients the opportunity to continue study treatment until the anticipated approval of baricitinib, in this indication. The discontinuation criterion was updated to reflect the possibility of study termination following the potential approval of baricitinib for the treatment of Atopic Dermatitis or its dismissal due to negative opinion within a given country.
Section 10.1	Updated sample size estimates for individual doses.	Sample size estimates for individual doses were corrected from the previous protocol version. There was no change in the overall sample size estimate
Section 10.3.7.1	Updated wording related to Interim locks.	This study will provide long-term safety data to support regulatory submissions. The wording related to study interim data locks was updated to make it clearer that multiple interim locks may occur to support these regulatory activities. In addition, the timing of the data lock(s) for the

Section # and Name	Description of Change	Brief Rationale
		analysis of the efficacy data from the sub-study will also be determined by the retreatment rates within the sub-study.

Revised Protocol Sections

Note:	Deletions have been identified by ~~strikethroughs~~. Additions have been identified by the use of underscore.

A ~~2~~4-year Phase 3 multicenter, double-blind long-term extension study to evaluate the safety and efficacy of placebo, baricitinib 1-mg, baricitinib 2-mg, and baricitinib 4-mg in adult patients with atopic dermatitis, including a blinded randomized treatment withdrawal and randomized downtitration.

1. Synopsis

Summary of Study Design:

Study I4V-MC-JAHN (JAHN) is a Phase 3, multicenter, double-blind study to evaluate the long-term safety and efficacy of baricitinib (1-mg once daily, 2-mg once daily, and 4-mg once daily) in adult patients with AD for ~~approximately 2~~ up to 4 years. The study population will include patients aged 18 years or older who completed an originating study (such as I4V-MC-JAHL, I4V-MC-JAHM, or 14V-MC-JAIY) and were eligible for enrollment into JAHN. There is a single substudy included that will evaluate treatment withdrawal and dose downtitration.

Treatment Arms and Duration:

The treatment study duration will be up to ~~2~~ 4 years. The study consists of 3 study periods and 1 substudy: randomized treatment withdrawal and downtitration.

Treatment Period: The full treatment period will last from Week 0/Visit 1 through Week ~~104~~200/Visit ~~16~~22. Patients will continue using emollients daily and topical corticosteroid (TCS) use will be permitted at the investigator's discretion and provided automatically at the time of rescue or retreatment.

Treatment Period 2: Week 52 (Visit 8) through Week ~~104~~ 200 (Visit ~~16~~22):

Post-Treatment Follow-Up Period: This period spans from the last treatment visit at Week ~~104~~200/Visit 22~~16~~, or early termination visit to approximately 28 days following the last dose of IP.

2. Schedule of Activities

Table JAHN.1. I4V-MC-JAHN Schedule of Activities

	Treatment Period 1							Treatment Period 2															
	Screening and Baseline Period																						**PTFU Period**
Visit number	**1[a]**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16 ~~/ET~~**	**17**	**18**	**19**	**20**	**21**	**22/ ET[s]**	**801[b]**
Weeks from entry into JAHN	**0**	**4**	**8**	**16**	**24**	**36**	**48**	**52**	**56**	**60**	**64**	**68**	**76**	**84**	**92**	**104**	**120**	**136**	**152**	**168**	**184**	**200**	**~~108~~204**
Visit tolerance interval (days) from entry into JAHN	0 to 56 from last visit of originating study[c]	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±7	±7	±7	±7	±7	±7	28 ± 4 after last dose
Procedures																							
Inclusion and exclusion criteria review	X[d]																						
Informed consent	X[e]																						
Abbreviated demographics	X																						
Clinical Assessments																							
Weight	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Vital signs (BP and pulse)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Symptom-directed physical examination[f]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Adverse events	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Concomitant medication	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ePRO (patient diary) dispensed	X	X	X				X	X[g]	X[g]	X[g]	X[g]												
ePRO (patient diary) returned	X	X	X	X				X	X[g]	X[g]	X[g]	X[g]				~~X~~[h]						X[h]	
Rerandomization[i, j]	X							X															
IWRS	X[j]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
IP dispensed[j]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
IP returned and compliance assessed		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Dispense TCS (as needed)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	~~X~~							

	Treatment Period 1							Treatment Period 2															PTFU Period
	Screening and Baseline Period																						
Visit number	**1[a]**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16 ~~/ET~~**	**17**	**18**	**19**	**20**	**21**	**22/ ET[s]**	**801[b]**
Weeks from entry into JAHN	**0**	**4**	**8**	**16**	**24**	**36**	**48**	**52**	**56**	**60**	**64**	**68**	**76**	**84**	**92**	**104**	**120**	**136**	**152**	**168**	**184**	**200**	**~~108~~204**
Visit tolerance interval (days) from entry into JAHN	0 to 56 from last visit of originating study[c]	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±7	±7	±7	±7	±7	±7	28 ± 4 after last dose
Weigh (tube with cap) and record returned TCS (as needed)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X						X[s]	X[t]
Physician-Assessed Efficacy Measures																							
IGA	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
EASI	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
SCORAD	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Health Outcomes Measures and Other Questionnaires[k]																							
POEM	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
DLQI	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
WPAI-AD	X	X		X	X	X	X	X	X			X	X	X	X	X						X[s]	X[t]
EQ-5D-5L	X	X		X	X	X	X	X	X			X	X	X	X	X						X[s]	X[t]
Itch NRS	X	X	X	X			X	X	X	X	X	X				~~X~~[h]						X[h]	
Skin Pain NRS	X	X	X	X			X	X	X	X	X	X				~~X~~[h]						X[h]	
ADSS	X	X	X	X			X	X	X	X	X	X				~~X~~[h]						X[h]	
PGI-S-AD	X	X	X	X			X	X	X	X	X	X				~~X~~[h]						X[h]	
HADS	X	X	X	X	X	X	X	X	X		X		X	X	X	X						X[s]	X[t]
C-SSRS[l] and Self-Harm Supplement Form	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Self-Harm Follow-up Form[m]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Laboratory Assessments																							
Clinical chemistry[n]	X[o]	X	X	X	X	X		X			X		X		X	X	X	X	X	X	X	X	X
Hematology	X[o]	X	X	X	X	X		X			X		X		X	X	X	X	X	X	X	X	X
Lipids (fasting)[p]	X[o]			X		X		X			X		X			X	X	X	X	X	X	X	
Urinalysis	X[o]	X		X		X		X			X		X		X	X						X[s]	X[t]
HBV DNA[q]	X[o]			X		X		X			X		X		X	X	X	X	X	X	X	X	X
Urine pregnancy[r]	X[o]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

	Treatment Period 1							Treatment Period 2															PTFU Period
	Screening and Baseline Period																						
Visit number	**1[a]**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16 ~~/ET~~**	**17**	**18**	**19**	**20**	**21**	**22/ ETs**	**801[b]**
Weeks from entry into JAHN	**0**	**4**	**8**	**16**	**24**	**36**	**48**	**52**	**56**	**60**	**64**	**68**	**76**	**84**	**92**	**104**	**120**	**136**	**152**	**168**	**184**	**200**	**~~108~~204**
Visit tolerance interval (days) from entry into JAHN	0 to 56 from last visit of originating study[c]	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±4	±7	±7	±7	±7	±7	±7	28 ± 4 after last dose
Serum immunoglobulin (IgE)				X		X		X								X						Xs	
Stored serum and plasma samples for exploratory analysis				X		X		X								X						Xs	
Stored blood for RNA analysis				X		X		X								X						Xs	

b Patients who have discontinued IP but remain in the study for more than 28 days without IP can combine their Visit ~~16~~22/ET with Visit 801 (follow-up visit).

s An ET visit should be conducted if a patient discontinues from the study before Week 200. Early termination visit activities do not need to be duplicated if occurring at the time of a scheduled visit. Weighing of TCS, collection of WPAI-AD, EQ-5D-5L, HADS, urinalysis, serum IgE exploratory storage samples, RNA and biomarker samples should only be performed at the ET visit if it occurs at or before Week 104.

t For V801, the weighing of TCS, and collection of WPAI-AD, EQ-5D-5L, HADS, and urinalysis should only be performed if it occurs at or before Week 104.

NOTE: Patients completing V16 and planning to sign ICF for amendment (c) can participate as long as they have not completed a V801.

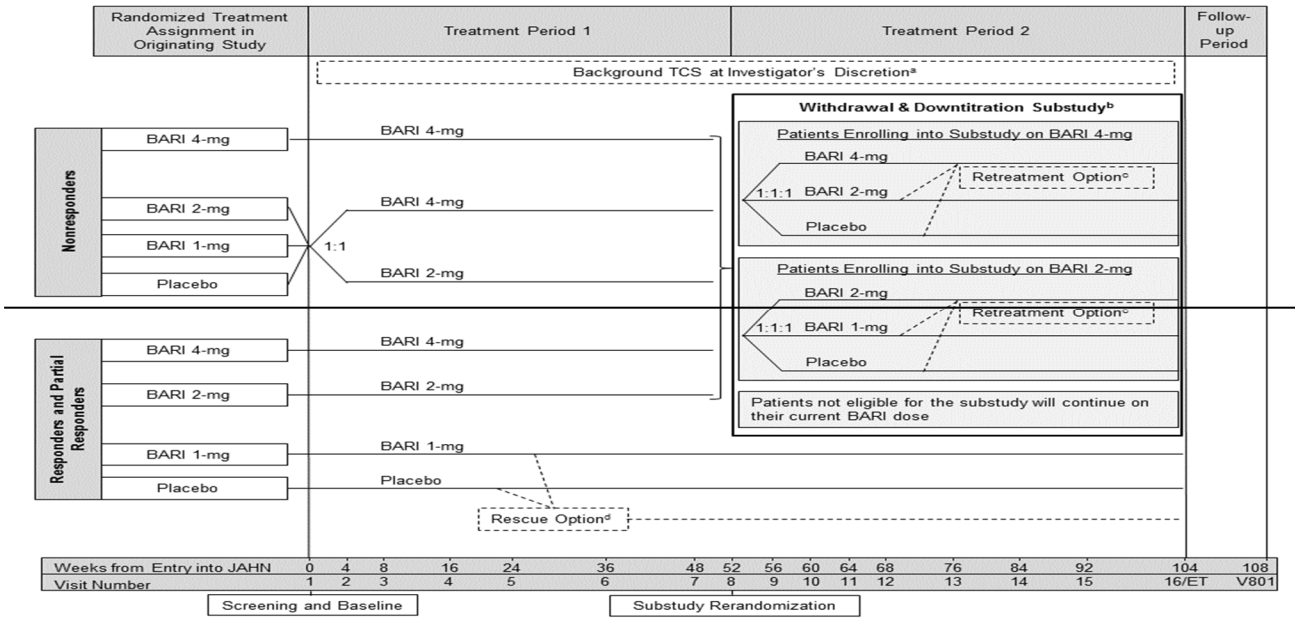
4. Objectives and Endpoints

Table JAHN.2. Objectives and Endpoints

Exploratory objectives 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 2 score, Skin Pain NRS, PGI-S-AD. Patients continuing on placebo as responders will be assessed during the long-term extension for relevant efficacy endpoints. Assessments of efficacy may be performed beyond Week 104 up to Week 200. The timing of the data lock(s) for the analysis of the efficacy data from the sub-study will be determined by the retreatment rates (see Section 10.3.7).

5.1 Overall Design

Study I4V-MC-JAHN (JAHN) is a Phase 3, multicenter, double-blind study to evaluate the long-term safety and efficacy of daily baricitinib 1-mg, 2-mg, and 4-mg in patients with AD for approximately ~~2~~ 4 years. The study will consist of 3 study periods and 1 substudy: randomized treatment withdrawal and downtitration. Patients entering Study JAHN will be classified as "Responders and Partial Responders" or "Nonresponders." Figure JAHN.1 illustrates the study design. The full visit schedule is outlined in the Study Schedule of Activities (Section 2). Patients who completed originating Studies I4V-MC-JAHL (JAHL) and I4V-MC-JAHM (JAHM) may be eligible for enrollment into Study JAHN; there may also be additional studies developed that will be eligible to enroll patients directly into Study JAHN.



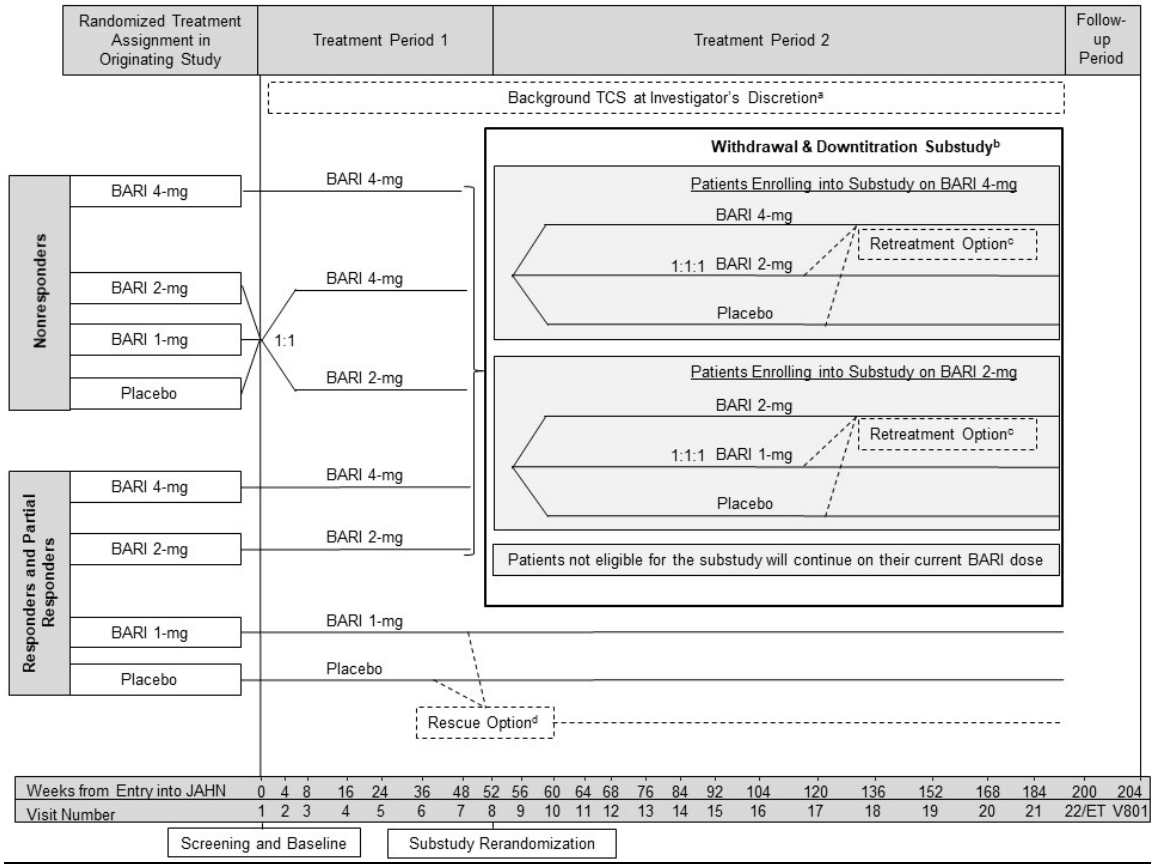


Figure JAHN.1. Illustration of study design for clinical Protocol I4V-MC-JAHN.

5.1.3. Treatment Period 2: 52 to ~~104~~ 200 Weeks

5.1.4. Post-treatment Follow-up Period

Patients who complete the study through Visit ~~16~~ 22(Week ~~104~~ 200) will have a post-treatment follow-up visit (Visit 801) approximately 28 days after the last dose of IP.

5.4 Scientific Rationale for Study Design

Atopic dermatitis is a chronic, relapsing inflammatory skin disorder (Bieber 2010). Although several therapies are available to patients for the treatment of flares, there are very limited options for long-term management. Most patients rely on chronic use of emollients and intermittent use of TCS or topical calcineurin inhibitors (TCNIs) to regulate skin inflammation in response to flares. In patients with moderate to severe disease, long-term disease control is not always achieved. In order to address this important medical need for a safe and effective long-term therapy, this study will allow for a longer treatment period (~~approximately 2~~ up to 4 years).

7.2 Method of Treatment Assignment

The IWRS will be used to assign blister packs, each containing double-blind IP tablets, to each patient, starting at Visit 1 (Week 0), and at each visit up to and including Visit ~~15~~21 (Week ~~92~~184). Site personnel will confirm that they have located the correct blister packs by entering a confirmation number found on the blister packs into the IWRS.

7.6 Treatment Compliance

Patient compliance with IP will be assessed at Visit 2 through Visit 22 ~~16~~ and at Early Termination during the treatment period by counting returned tablets.

7.7.1 Topical Corticosteroids and Other Topical Treatments

Choice of Topical Treatment

- Triamcinolone cream 0.1% and/or hydrocortisone 2.5% ointment. Where possible both of these treatments will be supplied by the sponsor during the first 2 years of the treatment period (dispensed at Visits 1-15), and use should be recorded via weight of returned tube as indicated in the SOA (Section 2). In the event of these specific TCS being unavailable during the first 2 years, an alternate, equivalent-potency TCS may be provided by the sponsor. Following Visit 15, patients may independently continue to use their TCS of choice as directed by their investigator, as per clinical practice, but these will not be provided by the sponsor and weight will not be recorded.
    - In the event that the sponsor is unable to supply TCS during the first 2 years, commercially available triamcinolone 0.1% cream and/or hydrocortisone 2.5% ointment may be supplied by the sites. Where providing triamcinolone 0.1% cream and/or hydrocortisone 2.5% ointment is not possible, an equivalent-potency TCS cream and/or ointment that is in line with local practices can be supplied. Refer to Appendix 6 for guidance on potency equivalence.
    - If the TCS supplied by the sponsor during the first 2 years is not considered suitable for an individual patient, an equivalent-potency TCS cream and/or ointment that is in line with local practices can be supplied by the sites. Refer to Appendix 6 for guidance on potency equivalence.

7.7.3 Prohibited Medications and Procedures

Prohibited Medications and Procedures Not Requiring Interruption of Investigational Product

The following therapies will not be allowed during the course of the study and, if taken by or administered to the patient, the prohibited therapy must be discontinued.

- ~~leukotriene inhibitors (e.g., montelukast [Singulair], zafirlukast [Accolate], and zileuton [Zyflo])~~
- ~~allergen immunotherapy~~

- phototherapy including PUVA (psoralen and ultraviolet A), ultraviolet B, tanning booth, and excimer laser
- bleach baths

8.1.1 Temporary Interruption from Study Treatment

Table JAHN.4. Criteria for Temporary Interruption of Investigational Product

Hold IP if the Following Laboratory Test Results or Clinical Events Occur:	IP May Be Resumed When:
WBC count <2000 cells/μL (<2.00x10^3/μL or <2.00 GI/L)	WBC count ≥2500 cells/μL (≥2.50x10^3/μL or ≥2.50 GI/L)
ANC <1000 cells/μL (<1.00x10^3/μL or <1.00 GI/L)	ANC ≥1200 cells/μL (≥1.20x10^3/μL or ≥1.20 GI/L)
Lymphocyte count <500 cells/μL (<0.50x10^3/μL or <0.50 GI/L)	Lymphocyte count ≥750 cells/μL (≥0.75x10^3/μL or ≥0.75 GI/L)
Platelet count <75,000/μL (<75x10^3/μL or <75 GI/L)	Platelet count ≥100,000/μL (≥100x10^3/μL or ≥100 GI/L)
eGFR <40 mL/min/1.73 m^2 (from serum creatinine) for patients with originating study screening eGFR ≥60 mL/min/1.73 m^2	eGFR ≥50 mL/min/1.73 m^2
eGFR <30 mL/min/1.73 m^2 (from serum creatinine) for patients with originating study screening eGFR ≥40 to <60 mL/min/1.73 m^2	eGFR ≥40 mL/min/1.73 m^2
ALT or AST >5 x ULN	ALT and AST return to <2 x ULN, and IP is not considered to be the cause of enzyme elevation
Hemoglobin <8 g/dL (<80.0 g/L)	Hemoglobin ≥10 g/dL (≥100.0 g/L)
Symptomatic herpes zoster	All skin lesions have crusted and are resolving
Infection that, in the opinion of the investigator, merits the IP being interrupted	Resolution of infection
~~Clinical features of VTE (such as deep vein thrombosis or pulmonary embolism) are present[a]~~	~~After evaluation and institution of appropriate treatment of VTE[b]~~

Abbreviations: ALT = alanine aminotransferase; ANC = absolute neutrophil count; AST = aspartate aminotransferase; eGFR = estimated glomerular filtration rate; IP = investigational product; ULN = upper limit of normal; ~~VTE = venous thromboembolic event;~~ WBC = white blood cell.

~~[a] Evaluate promptly and institute appropriate treatment. If upon evaluation VTE is ruled out and no other temporary or permanent discontinuation criteria are met, then IP may be resumed.~~

~~[b] If after evaluation and institution of treatment the investigator deems that the patient is still at significant risk, or if this would constitute a second VTE for the patient, then IP should be discontinued permanently.~~

Although temporary interruption of IP is not a requirement at times of increased potential risk of VTE (venous thromboembolic event; e.g., surgery, significant air travel, or other situations involving prolonged immobilization), ~~the~~ following appropriate VTE prophylaxis guidelines ~~are~~ is recommended to help manage the elevated risk under these circumstances.

8.1.2 Permanent Discontinuation from Study Treatment

- develop a ~~second~~ VTE

NOTE: Patients who develop a VTE may have additional follow-up and testing recommended (see Appendix 7).

8.2 Discontinuation from the Study

- study termination may occur in a specific country or region when baricitinib is approved for the treatment of atopic dermatitis and becomes reimbursed or commercially available in that country or region, or a negative regulatory opinion is received in that country or region.

9.4.9 Safety Monitoring

Investigators will monitor vital signs and carefully review findings that may be associated with cardiovascular event~~s~~ and VTEs (Appendix 7). Adverse event reports and vital signs will be collected at each study visit. The cardiovascular monitoring plan includes the following:

- potential MACE (cardiovascular death, MI, stroke), other cardiovascular events (such as hospitalization for unstable angina, hospitalization for heart failure, serious arrhythmia, resuscitated sudden death, cardiogenic shock, coronary revascularization such as coronary artery bypass graft or percutaneous coronary intervention~~s~~), venous thrombotic events, and noncardiovascular deaths will be identified by the investigative site or through medical review and will be sent to a blinded Clinical Event Committee for adjudication at regular intervals.

10.1 Sample Size Determination

It is anticipated that ~~90%~~ 95% of enrolled patients will complete Studies JAHL, JAHM, and JAIY and roll over into Study JAHN. Therefore, planned enrollment into Study JAHN from the originating Studies JAHL, JAHM, and JAIY will be approximately 1425 patients. Of these patients, approximately ~~465~~765 will start Study JAHN on baricitinib 4-mg, ~~348~~547 on baricitinib 2-mg, ~~58~~34 on baricitinib 1-mg, and ~~97~~79 on placebo. Patients who are considered nonresponders at entry into Study JAHN will be randomized 1:1 to either baricitinib 4-mg or baricitinib 2-mg and stratified by IGA 0, 1, 2, IGA 3 and IGA 4. This study is meant to evaluate patients' long-term response of baricitinib and the sample sizes are not determined to detect differences between baricitinib and placebo in a statistically powered manner. Additional patients may enroll from addenda or other studies.

10.3.7.1 Data Monitoring Committee

Besides DMC members, a limited number of pre-identified individuals may gain access to the limited unblinded data, as specified in the unblinding plan, prior to the ~~interim or~~ final database lock, in order to initiate work for regulatory submission. Interim locks may be conducted at various timepoints to support regulatory activities and scientific disclosures. The timing of the data lock(s) for the analysis of the efficacy data from the sub-study will be determined by the retreatment rates. Information that may unblind the study during the analyses will not be reported to study sites or blinded study team until the study has been unblinded.

Appendix 2.3.2. Discontinuation of the Study

The study will be discontinued if Lilly judges it necessary for medical, safety, regulatory, or other reasons consistent with applicable laws, regulations, and GCP. Study termination may occur in a specific country or region when baricitinib is approved for the treatment of AD and becomes reimbursed or commercially available in that country or region, or a negative regulatory opinion is received in that country or region.

Leo Document ID = aa36d560-7ca0-4d7d-b51c-b18cc3da1f51

Approver: PPD

Approval Date & Time: 27-Nov-2019 01:30:10 GMT

Signature meaning: Approved

Approver: PPD

Approval Date & Time: 27-Nov-2019 02:33:07 GMT

Signature meaning: Approved

Approver: PPD

Approval Date & Time: 27-Nov-2019 14:09:49 GMT

Signature meaning: Approved

1. Statistical Analysis Plan: I4V-MC-JAHN: A Phase 3 Multicenter Double-Blind Study to Evaluate the Long-Term Safety and Efficacy of Baricitinib in Adult Patients with Atopic Dermatitis

BREEZE-AD3

Baricitinib (LY3009104) Atopic Dermatitis

Study I4V-MC-JAHN (NCT03334435) is a 2-year Phase 3 multicenter, double-blind long-term extension study to evaluate the safety and efficacy of placebo, baricitinib 1 mg, baricitinib 2 mg, and baricitinib 4 mg in adult patients with atopic dermatitis, including a blinded randomized treatment withdrawal and randomized down-titration.

Eli Lilly and Company
Indianapolis, Indiana USA 46285
Protocol I4V-MC-JAHN
Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 12 April 2019
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly: 03 May 2019
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly: 29 July 2019
Statistical Analysis Plan Version 4 electronically signed and approved by Lilly: 23 January 2020
Statistical Analysis Plan Vesion 5 electronically signed and approved by Lilly on date provided below.

Approval Date: 19-May-2020 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

Table of Contents

Appendix **Page**

3. Revision History

Statistical Analysis Plan (SAP) Version 1 is based on Protocol I4V-MC-JAHN(a) (JAHN[a]) and was approved prior to the first unblinding. The analyses for Treatment Period 1 of Addendum I4V-MC-JAHN(7) (JAHN[7]) are described in Appendix 1. The analyses for the Withdrawal and Down-titration Substudy of I4V-MC-JAHN (JAHN), for both the main protocol JAHN(a) and addendum JAHN(7), will be described in subsequent versions of this SAP. SAP Version 2 was approved prior to the Week 16 database lock (DBL) and includes the following changes:

- *Section 5.1* – separated "responders" into 2 groups: "responders" and "partial responders"
- *Section 6.2.1* – added verbiage to distinguish patients from originating studies: monotherapy studies (I4V-MC-JAHL [JAHL]/I4V-MC-JAHM [JAHM]) and combo therapy study (I4V-MC-JAIY [JAIY])
- *Section 6.2* – updated model to include region; added methods for time to 4-point itch reduction and time to partial response, and aligned Cyclosporine wording to that of originating studies.
- *Section 6.4* – updated censoring rule, updated sensitivity analyses
- *Section 6.8.2* – Baseline Disease Characteristics: Cyclosporine wording updated to "…prior to entry into originating study" and added topical calcineurin inhibitor (TCNI) and vaccine
- *Section 6.9.1* – Background Therapy: added 2 new analyses "Summary of AD Background Therapy" and "Summary of time to initial use of AD background therapy post baseline" (for responders and partial responders)

SAP Version 3 was approved prior to the DBL that supports the EU and Japan submissions and includes the following changes:

- *Section 6.2.1* – a Week 36 efficacy evaluable set was defined for Week 36 efficacy and health outcome analyses.
- *Section 6.2.2* – clarifed that for step 2 of the Week 16 daily diary window construction, the visit anchor date will be on or before a scheduled visit occuring after Week 16.
- *Sections 6.2.3 and 6.11.1; Tables JAHN.6.6 and JAHN.6.8* – clarified that efficay and health outcome analyses for the EU and Japan DBL will be done for Weeks 0 through 24, and for Week 36 only descriptive statistics will be included.
- *Section 6.2.3* – analyses for safety, for the EU and Japan DBL, will include data for Weeks 0 through 52 entered up to the database cutoff date.
- *Table JAHN.6.2* – added modified baseline observation carried forward (mBOCF) imputation method for nonresponder modified intent-to-treat (mITT) continuous sensitivity analyses.
- *Section 6.7* – clarified that a by-patient listing of disposition and of treatment assignments will be provided for the intent-to-treat (ITT) population.
- *Section 6.9.1* was edited and reorganized for clarity.
- *Section 6.11.3* – added sensitivity analyses for categorical outcomes using JAHN baseline instead of the baseline from the respective originating study.
- *Tables JAHN.6.6 and JAHN.6.8* – added the following analyses:

  - proportion of patients with a response of Investigator's Global Assessment (IGA) [0,1], analysis using JAHN as the baseline value
  - proportion of patients with a 4-point improvement from JAHN baseline in Itch numeric rating scale (NRS) in subgroup of patients who had baseline Itch NRS ≥4
  - proportion of patients with a 4-point improvement from baseline of originating study in Skin Pain in subgroup of patients who had baseline Skin Pain NRS ≥4
  - proportion of patients with a 1-point improvement from baseline of originating study in a subgroup of patients with baseline item score of ≥1 for the following Atopic Dermatitis Sleep Scale (ADSS) item scores:
    - Item 1 score
    - Item 2 score
    - Item 3 score
  - proportion of patients with a 2-point improvement from baseline of originating study in item 2 score of ADSS in a subgroup of patients with baseline item 2 score of ≥2
  - proportion of patients with a 4-point improvement in Patient-Oriented Eczema Measure (POEM) score from baseline of originating study in a subgroup of patients with baseline POEM ≥4
  - change from baseline in Hospital Anxiety Depression Scale (HADS) total score
  - proportion of patients with HADS domain (anxiety and depression) score of <8 in a subgroup of patients with baseline HADS domain (anxiety and depression) score of ≥ 8
  - proportion of patient with HADS anxiety or depression domain score of <8 in a subgroup of patients with baseline HADS anxiety or depression doman score of ≥8
  - proportion of patients with Dermatology Life Quality Index (DLQI) (0,1); and
  - proportion of patients with a 4-point improvement from baseline in originating study in a subgroup of patients with baseline DLQI score ≥4
- *Section 6.14.1* – defined exposure ranges for the EU and Japan DBL.
- *Section APP.1.2.5; Tables APP.1.3 and APP.1.4* – clarified that for the EU and Japan DBL the efficacy and health outcome analyses will include descriptive statistics for Weeks 0 through 36 only.
- *Table APP.1.2* – removed the mixed -model repeated measures (MMRM)/mBOCF imputation methods for continuous endpoints.
- *Section APP.1.2.8* – clarifed that the by-patient listing for treatment assignments will be provided for the ITT population.
- *Section APP.1.2.15* clarified that the open-label safety analyses will be pooled with the baricitinib 2-mg treatment group from Study JAHN.

SAP Version 4 was approved prior to the DBL that supports the US submission. This version of the SAP pertains exclusively to the US submissions DBL and does not contain information related to any other prior or upcoming DBLs. This version is based on Protocol I4V-MC-JAHN(c) (JAHN[c]) and was approved prior to the first unblinding. The analyses for treatment Period 1 of Addendum I4V-MC-JAHN(7.1) (JAHN[7.1]) are described in Appendix 1. This includes the following changes:

- *Section 5.1* – changed the duration of Treatment Period 2 in accordance with Protocol amendment JAHN(c).
- *Figure JAHN.5.1.* – changed to reflect the change in Study Period 2 timeline.
- *Section 6.1* – updated sample size for the study.
- *Section 6.2.1* – added the following populations:
  - Week 24 efficacy evaluable set for patients originating from study JAIY
  - Weeks 56, 60, 64, and 68 efficacy evaluable set for patients entering downtitration and withdrawal substudy (Period 2)
  - safety population for Period 2
- *Table JAHN.6.1* – updated to reflect separate treatment groups for Period 1 and Period 2. Added treatment groups for Treatment Period 2.
- *Section 6.2.2* – the following changes were incorporated:
  - Baseline and postbaseline measures are separately defined for Period 1 and Period 2.
  - Week 52 baseline is defined.
  - Baseline for safety is defined.
  - Daily diary window calculation for Period 2 is added.
- *Section 6.2.3* – the following changes were incorporated:
  - Treatment comparisons for patients originating from combination therapy study and for patients entering Period 2 are added.
  - Analysis of covariance (ANCOVA) is added as the primary analysis method for continuus endpoints for Period 2.
  - Deleted analysis using Cox proportional hazards model.
- *Section 6.3* – added week 52 baseline values for Period 2 covariate adjustments.
- *Section 6.4* – the following changes were incorporated:
  - added retreatment after downtitration or early withdrawal as an intercurrent event
  - added Period 2 censoring rule
  - deleted section on mBOCF method for missing data
- *Table JAHN.6.2* was updated as follows:
  - deleted placebo multiple imputation (pmi) method for missing data handling for IGA 0,1 endpoint
  - modified last observation carried forward (mLOCF) added as an imputation method
  - deleted mBOCF imputation method
  - added imputation methods used in Period 2
- *Section 6.7* – added patient disposition calculation for Period 2
- *Section 6.8* – added patient compliance calculation for Period 2
- *Section 6.9.1* – added background therapy calculation for Period 2
- *Section 6.7* – added patient disposition calculation for Period 2
- *Table JAHN.6.3* – added tretament groups for Period 2
- *Table JAHN.6.4* – added variables relapse and rebound
- *Table JAHN.6.5* – added analysis for the following:
  - patients originating from combination therapy study
  - patients entering Period 2
  - changed time point analyzed for monotherapy studies in accordance with the US DBL

  - added analysis for endpoints rebound and relapse
- *Tables JAHN.6.7* – added analysis corresponding to patients originating from combination therapy study
- *Section 6.8* – added safety analysis for patients entering Period 2
- *Appendix 1* – aligned with protocol amendment JAHN(7.1)
- *Section APP.1.2.5; Table APP.1.3 and APP.1.4* – clarified that for the EU and Japan DBLs the efficacy and health outcome analyses will include descriptive statistics for Weeks 0 through 36 only
- *Figure APP.1.2.* – changed to reflect the change in Study Period 2 timeline.
- *Section APP.1.2.8* clarifed that the by-patient listing for treatment assignments will be provided for the ITT population
- *Section APP.1.2.3.2.1* – defined Week 24 efficacy evaluable set for open label patients
- *Table APP.1.3* – changed analysis time period to be Week 0 to 16 and Week 24
- *Table APP.1.4* – added analysis for the following endpoints:
  - 4-point improvement in Skin Pain NRS
  - 1.5-point improvement in ADSS Item 2 score
  - 4-point improvement in POEM score
  - DLQI (0,1) response
  - 4-point improvement in DLQI score

SAP Version 5 was approved prior to the 4-month safety update (20 May 2020). The overall changes incorporated in Version 5 are as follows:

- *Section 6.4.4.* – section added to include missing data imputation method for missingness due to COVID-19.
- *Appendix 2* – appendix 2 added to include the list of efficacy analyses planned for the 4-month safety update database lock.

4. Study Objectives

4.1. Primary Objective

The primary objective of this study is to estimate the effect of long-term therapy with baricitinib on responders and partial responders at entry of Study JAHN as assessed by the proportion of patients with a response of IGA 0 or 1 at Weeks 16, 36, and 52.

In particular, the associated estimand for this objective is to measure the effect of long-term therapy with baricitinib on responders and partial responders at entry of Study JAHN as assessed by the proportion of patients with a response of IGA 0 or 1 at Weeks 16, 36, and 52 assuming that treatment response disappears after patients discontinue from study or treatment. See Sections 6.4.1 and 6.11.1 on how this estimand handles outcomes after the occurrence of any intercurrent event through nonresponder imputation (NRI).

4.2. Secondary Objectives

The secondary objectives of this study are:

Objectives	Endpoints
Weeks 0-52	
Baricitinib Patients at Entry to Study JAHN To evaluate the effect of increasing or maintaining baricitinib dose on clinical measures and patient-reported outcomes.	• Proportion of patients with a response of Investigator's Global Assessment (IGA) 0, 1, or 2 assessed at Weeks 16, 36, and 52 • Proportion of patients with a response of IGA 0 or 1 assessed at Weeks 16, 36, and 52 (nonresponders) • Proportion of patients achieving response of ≥75% improvement from baseline in Eczema Area and Severity Index score (EASI75) from baseline of originating study assessed at Weeks 16, 36, and 52 • Proportion of patients with a 4-point improvement from baseline of originating study in Itch Numeric Rating Scale (NRS) at Week 16
Placebo Nonresponders at Entry to Study JAHN To evaluate the effect of starting baricitinib 2-mg versus 4-mg on clinical measures and patient-reported outcomes.	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at Weeks 4, 16, 24, 52 • Proportion of patients with a response of IGA 0 or 1 assessed at Weeks 4, 16, 24, 52 • Proportion of patients achieving response of EASI75 from baseline of originating study assessed at Weeks 4, 16, 24, 52 • Proportion of patients with a 4-point improvement from baseline of originating study in Itch NRS at Week 16
Weeks 52-104	
All Patients Entering the Substudy To evaluate the change in clinical response after treatment withdrawal or downtitration. **Patients Entering the Substudy with IGA 0 or 1** To evaluate the change in clinical response after	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at 16 weeks after rerandomization (Week 68) and Week 104 • Proportion of patients with a response of IGA 0 or 1 assessed at 16 weeks after rerandomization (Week 68) and Week 104

Objectives	Endpoints
treatment withdrawal or downtitration.	• Proportion of patients with a response of EASI75 from baseline of originating study assessed at 16 weeks after rerandomization (Week 68) and Week 104 • Time to retreatment (time to IGA ≥3)
Patients Retreated during Substudy To evaluate the ability to recapture efficacy based on clinical measures after experiencing a loss of treatment benefit:	• Proportion of patients with a response of IGA 0, 1, or 2 within 16 weeks of retreatment • Proportion of patients with a response of IGA 0 or 1 within 16 weeks of retreatment • Proportion of patients with a response of EASI75 from baseline of originating study within 16 weeks of retreatment
Patients Not Entered into the Substudy To evaluate the effect of maintaining baricitinib dose on clinical measures.	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at Week 104 • Proportion of patients with a response of IGA 0 or 1 assessed at Week 104 • Proportion of patients with a response of EASI75 from baseline of originating study assessed at Week 104

4.3. Exploratory Objectives

The exploratory objectives may include evaluating the response to baricitinib treatment regimens on clinical measures and patient-reported outcomes (PROs).

These endpoints may include dichotomous endpoints or change from baseline for the following measures: IGA, Eczema Area and Severity Index (EASI), SCORing Atopic Dermatitis (SCORAD), POEM, DLQI, Work Productivity and Activity Impairment (WPAI-AD), European Quality of Life – 5 Dimensions 5 Levels (EQ-5D-5L), Itch NRS, ADSS Item 2 score, Skin Pain NRS, Patient Global Impression of Severity (PGI-S-AD).

Additionally, assessment of background topical corticosteroids (TCS), such as the weight of sponsor-provided TCS and number of days not using TCS, will be analyzed. Patients continuing on placebo as responders will be assessed during the long-term extension for relevant efficacy endpoints.

Time to 4-point itch reduction during the first 14 days after initiation of treatment will be assessed.

Summary of IGA response (0,1) at Week 16 for the proportion of nonresponders with originating study treatment assignment in placebo, baricitinib 1 mg or baricitinib 2 mg with IGA response (3,4) at the start of study JAHN (Week 0).

5. Study Design

5.1. Summary of Study Design

Study JAHN is a Phase 3, multicenter, double-blind study to evaluate the long-term safety and efficacy of daily baricitinib 1 mg, 2 mg, and 4 mg in patients with atopic dermatitis (AD) for approximately 2 years. The study will consist of 3 study periods and 1 substudy: randomized treatment withdrawal and downtitration. Patients entering Study JAHN will be classified as "responders," "partial responders" or "nonresponders" based on the following criteria:

- **Responders (IGA of 0 or 1):** patients entering Study JAHN who have an IGA score of 0 or 1 and were not rescued in the originating study
- **Partial Responders (IGA of 2):** patients entering Study JAHN who have an IGA score of 2 and were not rescued in the originating study
- **Nonresponders:** any patient who does not meet the "rsponder" or "partial responder" definitions

Figure JAHN.5.1 illustrates the study design. The full visit schedule is outlined in the Study Schedule of Activities in the Protocol. Patients who completed originating studies, such as Studies JAHL, JAHM or JAIY may be eligible for enrollment into Study JAHN. This study also includes an open-label addendum for which patients do not have to complete an originating study.

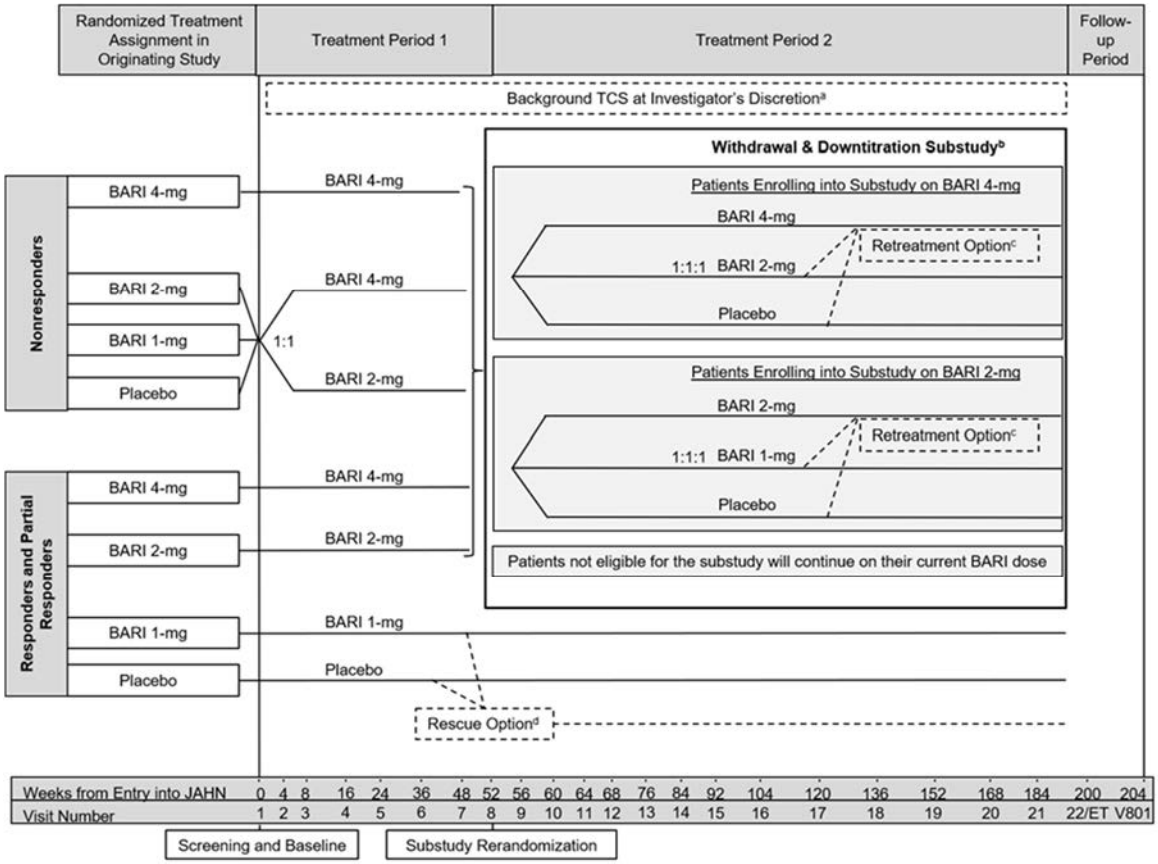
Patients completing Studies JAHL, JAHM or JAIY will have had approximately 16 weeks of treatment with investigational product (IP) (baricitinib or placebo) and will enter Study JAHN at Visit 1; the final visit of Study JAHL, JAHM or JAIY will be the first visit of Study JAHN unless washout of a concomitant medication was required in which case a new Visit 1 is required. If oral systemic AD treatments were administered for rescue in the originating study, then a minimum washout of 4 weeks is required prior to study drug initiation in Study JAHN (Visit 1). Patients will have been using emollients daily in their originating study and thus will continue during Study JAHN participation.

Daily diary dispensation/collection will continue through the first 4 months of treatment in JAHN and then dispensation will resume at Week 48 for an additional 5 months of collection to capture PROs prior to and during the withdrawal and downtitration substudy (except for patients known not to be in the substudy).

Study JAHN will consist of 4 periods:

- Screening and baseline period: Screening should occur during the last visit of the originating study unless the patient received oral systemic AD treatment as rescue therapy during the originating study and needs additional time to complete the required 4-week washout period or the sponsor gave approval.
- Treatment Period 1: Treatment period from Week 0 to Week 52. Patients will be classified as responders, partial responders or nonresponders at the entry of Study JAHN based on the criteria given in Section 5.1.

- Treatment Period 2: Treatment period from Week 52 to 200 and will contain a randomized withdrawal and downtitration substudy. At Week 52, all patients will be evaluated for substudy eligibility and the criteria are given in the Protocol. Patients not entered in the substudy will continue on their current treatment.
- Posttreatment follow-up period: Period from Week 200 (Visit 22) or Early Termination Visit (ETV) to approximately 28 days after the last dose of IP.



Abbreviations: BARI = baricitinib; IGA = Investigator's Global Assessment; TCS = topical corticosteroids.

[a] Background TCS may be initiated or reinitiated at any time during the study, following the guidelines in the Protocol (Section 7.7.1) and will be provided as part of rescue or retreatment any time a patient's IGA score becomes ≥3 as described in the Protocol (Sections 5.1.2.1 and 5.1.3.1).

[b] Eligible patients will be rerandomized in the withdrawal and downtitration substudy as described in the Protocol (Section 5.1.3.1). Patients who do not enroll in the substudy will remain on their treatment.

[c] Patients enrolled in the substudy will automatically be retreated if their IGA score becomes ≥3 as described in the Protocol (Section 5.1.3.1).

[d] Rescue is available as described in the Protocol (Section 5.1.2.1).

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

5.2. Method of Assignment to Treatment

At entry into Study JAHN, patients who meet all criteria for enrollment will be randomized or assigned treatment by a computer-generated random sequence using an interactive web-response system (IWRS). Patients originally assigned to placebo, 1-mg baricitinib or 2-mg baricitinib in Studies JAHL or JAHM, or patients originally assigned to placebo or 2-mg baricitinib in Study JAIY and classified as nonresponders, will be randomized in a 1:1 ratio to 2-mg baricitinib or 4-mg baricitinib and will be stratified by disease severity at baseline of JAHN (IGA 0, 1, 2 versus IGA 3 versus IGA 4). All other patients will be assigned to treatment in Study JAHN matching their prior assignment from Studies JAHL, JAHM or JAIY.

Additional addendum patients will be permitted to enroll directly into Study JAHN on Baricitinib 2-mg. These patients will not be blinded to treatment.

At Week 52, patients eligible for the withdrawal and downtitration substudy will be assigned to treatment by a computer-generated random sequence using an IWRS. Rerandomization will follow a 1:1:1 ratio allocation and will be stratified by disease severity (IGA 0, 1 versus IGA 2). Patients entering the substudy on baricitinib 4 mg will be rerandomized to baricitinib 4 mg, baricitinib 2 mg, and placebo; patients entering the substudy on baricitinib 2 mg will be rerandomized to baricitinib 2 mg, baricitinib 1 mg, and placebo.

The IWRS will be used to assign blister packs, each containing double-blind IP tablets, to each patient, starting at Visit 1 (Week 0), and at each visit up to and including Visit 21 (Week 184). Site personnel will confirm that they have located the correct blister packs by entering a confirmation number found on the blister packs into the IWRS.

This study will be conducted internationally in multiple sites. Table JAHN.5.1 describes how regions were defined for stratification in the originating studies.

Table JAHN.5.1. Geographic Regions for Stratification

Region	Country
Europe	Austria, Czech Republic, Denmark, France, Germany, Hungary, Italy, Poland, Spain, Switzerland
Japan	Japan
Rest of World	Argentina, Australia, India, Israel, Korea, Mexico, Russia, Taiwan

6. A Priori Statistical Methods

6.1. Determination of Sample Size

It is anticipated that 95% of enrolled patients will complete Studies JAHL, JAHM and JAIY and roll over into Study JAHN. Therefore, planned enrollment into Study JAHN from the originating Studies JAHL, JAHM, and JAIY will be approximately 1425 patients. Patients who are nonresponders at entry into Study JAHN and were not randomized to baricitinib 4 mg will be randomized 1:1 to either baricitinib 4 mg or baricitinib 2 mg and stratified by IGA 0, 1, 2 versus IGA 3 versus IGA 4. This study is intended to evaluate patients' long-term response of baricitinib and the sample sizes are not determined to detect differences between baricitinib and an appropriate comparator in a statistically powered manner. Additional patients may enroll from addenda or other studies.

Patients at Week 52 will be stratified by responder status (IGA 0 or 1 versus IGA 2) when entering the randomized withdrawal and downtitration substudy. It is estimated that there will be approximately 600 patients entering the randomized withdrawal and downtitration substudy. The substudy is meant to evaluate the change in clinical response after treatment withdrawal or downtitration and does not account for whether the sample size is sufficient to detect a difference between baricitinib and placebo. Maintenance of treatment benefit is defined as response of IGA 0, 1, or 2.

6.2. General Considerations

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

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly). The 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 would be 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.

Data collected at early termination visits will be mapped to the closest scheduled visit if it falls within the visit window as discussed in Section 6.2.2. For by-visit summaries, only visits in which a measure was scheduled to be collected will be summarized. Any unscheduled visit data will be included at the patient-level listings. However, the data will still be used in other analyses, including categorical analyses for safety.

6.2.1. Analysis Populations

Intent-to-treat (ITT) population: The ITT population analysis set is defined as all enrolled patients in Study JAHN.

Modified intent-to-treat (mITT) population: The mITT population analysis set is defined as all randomized patients who received at least one dose of the IP in Study JAHN.

The efficacy and health outcome analyses will be conducted on subsets of the mITT population, which is dependent on the objectives. For all analyses of patients from originating studies, responders, partial responders and nonresponders will be distinguished by their originating studies at entry of Study JAHN – monotherapy (JAHL/JAHM) and combination therapy (JAIY). Patients will be analyzed according to the treatment to which they were randomized.

The analysis of the primary objective will be conducted on the mITT population who are responders and partial responders (Section 5.1). Additionally, these endpoints will also be evaluated by parsing this group into responder mITT and partial responder mITT groups.

The populations for the secondary objectives (nonresponders who were on a baricitinib dose in originating study and nonresponders who were on placebo in originating study) will also be comprised of patients meeting the mITT definition. Exploratory objectives will also be conducted on the same populations.

For the DBL that supports the US submission (subsequently referred to as the US DBL), the population originating from combination therapy study JAIY, Week 24 efficacy and health outcome analyses will be conducted on the **Week 24 efficacy evaluable set**. This subset of the mITT Population is anchored on the database cut-off date for the US DBL (13 December 2019). Specifically, a patient will be included in the Week *x* efficacy evaluable set if their Week *x* visit has occurred or their expected Week *x* visit date plus a 15-day buffer is on or prior to the database cutoff date. The expected Week *x* visit date will be calculated as follows: date of first dose date + (*x* weeks * 7 days) + 15 days.

For the US DBL, for Treatment Period 2 (downtitration and withdrawal), Week 56, Week 60, Week 64, and Week 68 efficacy analyses will be conducted on the Week 56, Week 60, Week 64, and Week 68 efficacy evaluable sets respectively. The efficacy evaluable sets are a subset of the mITT population and are anchored on the database cut-off date for the US DBL.

Safety population: For the purpose of Study JAHN alone, the safety population is defined as all patients who received at least 1 dose of IP in Study JAHN and who did not discontinue from the study for the reason 'Lost to Follow-up' at the first postbaseline visit in Study JAHN. For the safety analysis of Treatment Period 2, patients with at least 1 dose of IP in Treatment Period 2 will be included in the safety analysis of Treatment Period 2.

Safety analyses will be performed using the safety population and will be analyzed according to the following treatment groups (or cohorts) defined in Table JAHN.6.1. Patients will be analyzed according to the dosing regimen to which they were assigned.

Table JAHN.6.1. Safety Population Treatment Groups

Treatment Groups for Treatment Period 1 (Week 0-52)	**Definition**
PBO	Placebo at entry to Study JAHN followed to end of Treatment Period 1 or censored at treatment change
BARI 1 mg	BARI 1 mg followed to end of Treatment Period 1 or censored at dose or treatment change
PBO/BARI 1 mg to BARI 2 mg	PBO/BARI 1 mg switching to BARI 2 mg at entry to Study JAHN followed to end of Treatment Period 1 or censored at dose or treatment change
BARI 2 mg	BARI 2 mg followed to end of Treatment Period 1 or censored at dose or treatment change
PBO/BARI 1 mg/BARI 2 mg to BARI 4 mg	PBO/BARI 1 mg/BARI 2 mg switching to BARI 4 mg at entry to Study JAHN followed to end of Treatment Period 1 or censored at dose or treatment change
BARI 4 mg	BARI 4 mg followed to end of Treatment Period 1 or censored at dose or treatment change
Pooled BARI	BARI 1 mg/BARI 2 mg/BARI 4 mg at entry to Study JAHN followed to end of Treatment Period 1 or censored at dose or treatment change
Treatment Groups for Treatment Period 2 (Week 52-104)	**Definition**
BARI 2 mg to PBO	Randomized to PBO in Treatment Period 2 (previously treated with BARI 2 mg) followed to end of study or censored at retreatment with BARI 2 mg
BARI 2 mg to BARI 1 mg	Randomized to BARI 1 mg in Treatment Period 2 (previously treated with BARI 2 mg) followed to end of study or censored at retreatment with BARI 2 mg
BARI 2 mg to BARI 2 mg	Randomized to stay on BARI 2 mg in Treatment Period 2 followed to end of study
BARI 4 mg to PBO	Randomized to PBO in Treatment Period 2 (previously treated with BARI 4 mg) followed to end of study or censored at retreatment with BARI 4 mg
BARI 4 mg to BARI 2 mg	Randomized to BARI 2 mg in Treatment Period 2 (previously treated with BARI 4 mg) followed to end of study or censored at retreatment with BARI 4 mg
BARI 4 mg to BARI 4 mg	Randomized to stay on BARI 4 mg in Treatment Period 2 followed to end of study
BARI 2 mg	Not part of the substudy, stayed on BARI 2 mg in Treatment Period 2 followed to end of study (this treatment group includes the open-label BARI 2 mg addendum patients)
BARI 4 mg	Not part of substudy, stayed on BARI 4 mg in Treatment Period 2 followed to end of study

Safety Population Treatment Groups

Treatment Group for Treatment Periods 1 and 2 (Week 0-104) (for Box Plots Only)	Definition
PBO	Placebo at entry to Study JAHN followed to end of study or censored at treatment change
BARI 1 mg	BARI 1 mg followed to end of study or censored at dose or treatment change
BARI 2 mg to PBO	BARI 2 mg in Treatment Period 1, rerandomized to PBO in Treatment Period 2 followed to end of study or censored at retreatment with BARI 2 mg
BARI 2 mg to BARI 1 mg	BARI 2-mg in Treatment Period 1, rerandomized to BARI 1 mg in Treatment Period 2 followed to end of study or censored at retreatment with BARI 2 mg
BARI 2 mg to BARI 2 mg	BARI 2 mg in Treatment Period 1, rerandomized to stay on BARI 2 mg in Treatment Period 2 followed to end of study
BARI 4 mg to PBO	BARI 4 mg in Treatment Period 1, rerandomized to PBO in Treatment Period 2 followed to end of study or censored at retreatment with BARI 4 mg
BARI 4 mg to BARI 2 mg	BARI 4 mg in Treatment Period 1, rerandomized to BARI 2 mg in Treatment Period 2 followed to end of study or censored at retreatment with BARI 4 mg
BARI 4 mg to BARI 4 mg	BARI 4 mg in Treatment Period 1, rerandomized to stay on BARI 4 mg in Treatment Period 2 followed to end of study
BARI 2 mg	BARI 2 mg in Treatment Period 1, did not enter the substudy and stayed on BARI 2 mg in Treatment Period 2 followed to end of study (this treatment group includes the open label BARI 2 mg addendum patients)
BARI 4 mg	BARI 4 mg in Treatment Period 1, did not enter the substudy and stayed on BARI 4mg in Treatment Period 2 followed to end of study

Abbreviations: BARI = baricitinib; PBO = placebo.

An all-baricitinib exposure, defined as patients valid for the safety population with at least 1 dose of baricitinib anytime during Study JAHN, will be analyzed for the final analysis (final datalock).

Follow-up population: The follow-up population is defined as patients who entered the follow-up period and completed the follow-up visit.

6.2.2. Definition of Baseline and Postbaseline Measures

Baseline

Period 1 (Weeks 0 to 52)

The baseline value for efficacy variables measured at scheduled visits is defined as the last nonmissing measurement on or prior to the date of first study drug administration in the originating study (expected at Week 0, Visit 2 of the originating study). For patients who directly enroll into Study JAHN without having completed an originating study, baseline value

for efficacy variables measured at scheduled visits is defined as the last nonmissing measurement on or prior to the date of first study drug administration in Study JAHN.

The baseline value for the daily diary assessments is the mean of the nonmissing assessments in the 7 days prior to the date of first study drug administration in the originating study (expected at Week 0, Visit 2).

If there are <4 nonmissing assessments in the baseline diary window, the interval lower bound can be extended up to 7 additional days, one day at a time, to obtain the most recent 4 nonmissing values. If there are not at least 4 nonmissing assessments in the baseline period, the baseline mean is missing.

Additionally, the Study JAHN baseline may be derived for some efficacy analyses (e.g., patients enrolled directly into Study JAHN without having completed an originating study). In this case, the Study JAHN baseline is defined as the last nonmissing measurement on or prior to the date of first study drug administration in Study JAHN.

For patients who are enrolled from an originating study and do not require washout, the baseline will be the last visit of the originating study. However, patients who received oral systemic AD treatment as rescue therapy during the originating study require additional time to compete the required 4-week washout and have an additional screening visit in Study JAHN (Visit 1 [Week 0]). Thus, for these patients the last nonmissing measurement on or after last visit of originating study and on or prior to first dose of study drug in Study JAHN will serve as the baseline.

For patients who do not require washout, baseline will be the last visit of the originating study. Patients who received oral systemic AD treatment as rescue therapy during the originating study require a 4-week washout, and will have an additional screening visit in Study JAHN (Visit 1 [Week 0]). The assessments collected at this visit will provide the Study JAHN baseline for these patients. If labs were not collected at this visit, then the most recent lab value collected at the last visit of the originating study will be used as baseline.

Period 2 (Study Weeks 52 to 68)

For analysis of endpoints assessed at scheduled visits in Period 2, baseline from originating study will be used. Change from baseline analysis will also explore changes from baseline using Week 52 baseline. In this case, the baseline for endpoints collected at scheduled visits is defined as the last nonmissing measurement on or prior to the date of the Week 52 visit in Study JAHN.

The baseline for daily diary assessments in Period 2 is the baseline from originating study as defined for Period 1. Analyses of some daily diary endpoints will also explore changes from baseline using Week 52 baseline as the defined baseline. The methodology for defining Week 52 as the baseline is the same as the methodology for defining Week 0 of the originating study described above anchoring on the first study drug administration of Period 2.

Safety Analyses

Safety analyses for treatment Period 1, and for Treatment Periods 1 and 2 combined (box plots of laboratory analytes and vital signs) will be conducted using the Study JAHN baseline, which is

defined as the last nonmissing scheduled (planned) measurement on or prior to the date of first study drug administration for continuous measures by-visit analyses and non-missing measurements on or prior to the date of first study drug administration in Study JAHN for all other analyses. Safety analyses for Treatment Period 2 will be conducted for adverse events (AEs) only. Baseline for Treatment Period 2 is defined as nonmissing measurements on or prior to the date of first study drug administration in Treatment Period 2 of Study JAHN.

Postbaseline

Period 1(Weeks 0 to 52)

Nonmissing postbaseline efficacy data collected at scheduled visits will be used for analyses. If an assessment is missing at a scheduled visit, an unscheduled postbaseline assessment can be used provided it falls within a ±4-day window of the scheduled visit date. If there is more than 1 unscheduled visit within the defined visit window and no scheduled visit assessment is available, the unscheduled visit closest to the scheduled visit date will be used. If 2 unscheduled visits of equal distance are available, then the latter of the 2 will be used. If there is no nonmissing measure collected at the scheduled visit or an unscheduled visit falling within the visit window, the assessment is missing for that scheduled visit.

Diary endpoints were collected up to Week 16 in Period 1 of Study JAHN. Postbaseline daily diary endpoints will be the mean of weekly visit windows (diary windows) anchored on day of first dose (Day 1) for Weeks 1 through 14 as follows:

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

Week 16 Daily Diary Window Construction

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

Step 1: If the Week 16 scheduled visit exists, the Week 16 diary interval is the 7 days prior to the Week 16 date provided that window has at least 4 nonmissing observations. If there are <4 nonmissing observations, the diary window's lower bound will be extended 1 day at a time (up to Day 99) to a maximum of 14 days prior to the Week 16 date until 4 nonmissing observations are obtained. If after extending this diary window's lower bound to 14 days there are <4 nonmissing observations, then go to Step 2.

Step 2: If the Week 16 scheduled visit does not exist, the 7 days prior to the last visit (scheduled or unscheduled) occurring after Day 105 and on or before a scheduled visit occuring after Week 16 will constitute the Week 16 diary window provided that window contains at least 4 nonmissing observations. If there are <4 nonmissing observations, the diary window's lower bound will be extended 1 day at a time (up to Day 99) to a maximum of 14 days prior to the unscheduled visit date until 4 nonmissing observations are obtained. If after extending this diary window's lower bound to 14 days there are less than 4 nonmissing observations then go to Step 3.

Step 3: If neither a Week 16 scheduled visit or an unscheduled visit to act as a surrogate for the Week 16 diary window is available, then the Week 16 window will be Day 106 to Day 112. If there are <4 nonmissing observations, the diary window's lower bound will be extended 1 day at a time to Day 99 until 4 nonmissing observations are obtained.

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

Week 15 Daily Diary Window Construction

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

Period 2 (Study Weeks 52-68)

Diary endpoints were collected up to Week 16 in Period 2 of JAHN. For Treatment Period 2, postbaseline daily diary endpoints will be the mean of weekly visit windows (diary windows) anchored on day of the Week 52 visit for Weeks 53 through 66 as follows:

Week of Period 2	Days in Period 2
53	1-7
54	8-14
55	15-21
56	22-28
57	29-35
58	36-42
59	43-49
60	50-56
61	57-63
62	64-70
63	71-77
64	78-84
65	85-91
66	92-98

Week 68 Daily Diary Window Construction

Week 68 daily diary window construction for Period 2 will follow the exact same steps as the window construction for Week 16 in Period 1.

Week 67 Daily Diary Window Construction

Week 67 daily diary window construction for Period 2 will follow the exact same steps as the window construction for Week 15 in Period 1.

Handling of Duplicate Diary Records

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

6.2.3. Analysis Methods

As few placebo patients are expected to be responders, no formal comparisons between treatment groups and placebo will be conducted for the responder and partial responder mITT populations. For the nonresponder mITT population, the following comparisons will be made for Period 1:

- for patients from originating monotherapy studies:
  - baricitinib 2 mg to baricitinib 4 mg versus baricitinib 2 mg to baricitinib 2 mg
  - baricitinib 1 mg to baricitinib 4 mg versus baricitinib 1 mg to baricitinib 2 mg
  - placebo to baricitinib 4 mg versus placebo to baricitinib 2 mg

- for patients from originating combination therapy study:
  - baricitinib 2 mg to baricitinib 4 mg versus baricitinib 2 mg to baricitinib 2 mg
    - placebo to baricitinib 4 mg versus placebo to baricitinib 2 mg

For Period 2, the following treatment comparisons will be performed:

- for patients on baricitinib 4 mg in Period 1:
  - placebo versus baricitinib 4 mg
  - baricitinib 2 mg versus baricitinib 4 mg
- for patients on baricitinib 2 mg in Period 1:
  - placebo versus baricitinib 2 mg
    - baricitinib 1 mg versus baricitinib 2 mg

For Period 1, the main analysis method of categorical efficacy variables will use a logistic regression analysis with region, baseline disease severity from originating study (IGA 3 or IGA 4), baseline value from originating study, and treatment group in the model. For Period 2, the main analysis method of categorical efficacy variables will use a logistic regression analysis but with region, baseline disease severity at Week 52, baseline value at Week 52, and treatment group in the model. In this case, the baseline disease severity at Week 52 can take values in 1 of 2 levels: 1) IGA of 0 or 1, and 2) IGA of 2. Firth's correction will be used in order to accommodate (potential) sparse response rates. The p-value for the odds ratio from the logistic regression model will be used for statistical inference, unless Firth's correction still results in quasiseparation. In that case, a Fisher's exact test p-value will be used for statistical inference. The difference in percentages and 100(1-alpha)% CI of the difference in percentages using the Newcombe-Wilson method without continuity correction will be reported.

For nonresponders, the main analysis method for all continuous efficacy variables will use MMRM analysis. The MMRM model will use a restricted maximum likelihood estimation. The model will include treatment, region, 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 assessment endpoints, the model term 'visit' will include each of 16 weeks. 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, followed by autoregressive (AR[1]), followed by compound symmetry will be used. The Kenward-Roger method will be used to estimate the degrees of freedom. Treatment least square means (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 an appropriate comparator (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. For responders and partial responders, descriptive statistics will be provided for all continuous efficacy and health outcomes variables.

Treatment comparisons for select continuous efficacy and health outcomes variables may also be made using ANCOVA for primary and secondary objectives. For Period 1 analysis, when an ANCOVA model is used, the model includes region, baseline disease severity, treatment group, and baseline value. Treatment LSM will be estimated within the framework of the ANCOVA

using type 3 sums of squares. Reported differences in LSM and associated p-values, standard errors, and 95% CI will be used for statistical inference. Analysis of covariance will be the primary method of analysis for continuous efficacy and health outcome variables for Period 2. In this Period, ANCOVA model includes region, baseline disease severity at Week 52, treatment group, and baseline value at Week 52. Baseline disease severity is defined previously.

Patients on baricitinib 2 mg or baricitinib 4 mg who are considered nonresponders (IGA >2) at entry into Study JAHN may respond with continued baricitinib treatment. Time to first response (IGA of ≤ 1) will be assessed using cumulative incidence functions for patients remaining on the same dose from the originating study. This will be used to determine when continued baricitinib treatment is futile. Missing IGA data will be replaced using the NRI rule.

For the US DBL all patients are expected to have either reached the Week 52 visit or have discontinued from the study. Fisher's exact test will be used to test for differences between each baricitinib dose (including those that are randomized to a different dose) and placebo in proportions of patients experiencing 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 or not the treatment group LSM changes from baseline are different from zero; the standard error for the LSM change will also be displayed. Differences in LSM will be displayed, with the p-value associated with the LSM comparison to placebo or appropriate comparator 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.

6.2.4. Derived Data

- age (year), derived using first dose date as the reference start date and July 1st of birth year and truncated to a whole-year (integer) age. Patients whose derived age is <18 will have the required minimum age of 18 at informed consent confirmed; reporting for age, age groups, and lab ranges, however, will be based on their derived age.
- age group (<65, ≥65 years old)
- age group (<65, ≥65 to <75, ≥75 to <85, ≥85 years old)
- body mass index (BMI) (kg/m^2) = Weight (kg)/(Height [cm]/100)2
- BMI category (<25 kg/m^2, ≥25 to <30 kg/m^2, ≥30 kg/m^2)
- the duration of AD from diagnosis (years) = ([Date of informed consent – Date of AD diagnosis] + 1) / 365.25
    - If year of onset is missing, duration of AD will be set as missing. 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 to <5 years, 5 to <10 years, 10 to <20 years, ≥20 years)
- diagnosis age (year), derived using diagnosis date as the reference start date and July 1st 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 from originating study
  - If a baseline value is missing, it will not be imputed and the change from baseline will not be calculated.
- percent change from baseline at Visit x:
  ([postbaseline measurement at Visit x – baseline measurement] / baseline measurement)*100
  - If a baseline value is missing, it will not be imputed and percent change from baseline will not be calculated.
- Weight (kg) = weight (lbs) * 0.454
- Weight category (<60 kg, ≥60 to <100 kg, ≥100 kg)
- Height (cm) = height (in) * 2.54
- cyclosporine inadequate efficacy prior to entry into originating study (yes, no)
  - Set to **yes** if the reason for discontinuation is inadequate response.
- cyclosporine intolerance prior to entry into originating study (yes, no)
  - Set to **yes** if the reasons for discontinuation are: intolerance to medication or contraindication (physician indicated cyclosporine was used and a contraindication was noted).
- cyclosporine contraindication (ineligible) prior to entry into originating study (yes, no)
  - Set to **yes** if cyclosporine was never used because of a contraindication.
- cyclosporine inadvisable prior to entry into originating study (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.
- TCNI inadequate efficacy prior to entry into originating study (yes, no)
  - Set to **yes** if the reason for discontinuation is inadequate response.
- TCNI intolerance prior to entry into originating study (yes, no)
  - Set to **yes** if the reasons for discontinuation are: intolerance to medication or contraindication (physician indicated TCNI was used and a contraindication was noted).
- TCNI contraindication (ineligible) prior to entry into originating study (yes, no)
  - Set to **yes** if TCNI was never used because of a contraindication
- TCNI inadvisable prior to entry into originating study (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. Adjustments for Covariates

The randomization to treatment groups in the originating studies is stratified by disease severity (IGA) and geographic region. Unless otherwise specified, the statistical analysis models will adjust for these stratification variables. For Period 1, the covariates used in the logistic model for categorical data will include the baseline value from the originating study. For Period 2, the covariates used in the logistic model for categorical data will include the baseline value from Week 52. The baseline value from Study JAHN will be used as a covariate in the ANCOVA model for continuous safety data. When an MMRM analysis is performed, baseline value from originating study and baseline-by-visit interactions will be included as covariates.

6.4. Handling of Dropouts or Missing Data

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

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

- application of the censoring rule (including after permanent study drug discontinuation or for responders and partial responders [on Placebo or baricitinib 1-mg] who are rescued to a higher dose [baricitinib 2 mg or baricitinib 4 mg])
- retreatment after downtitration or early withdrawal
- discontinuation
- missing an intermediate visit prior to discontinuation or rescue
- lost to follow-up

Noncensor intercurrent events are events that are not due to the application of any censoring rule (i.e., the last 3 items in the list above).

Note that as efficacy and health outcome data can accrue after a patient permanently discontinues study drug or is rescued to either baricitinib 2 mg or baricitinib 4 mg (for responders and partial responders on placebo or baricitinib 1 mg), a censoring rule will be applied to all efficacy and health outcome observations subsequent to these events depending on the estimand being addressed.

The *Period 1 censoring rule* will censor efficacy after permanent study drug discontinuation or for responders and partial responders (on placebo or baricitinib 1 mg) who are rescued to a higher dose (baricitinib 2 mg or baricitinib 4 mg). This censoring rule will be applied to all efficacy endpoints conducted for the responder, partial responder, and nonresponder mITT population (defined in Section 6.2.1).

A *Period 2 censoring rule* will censor efficacy data after permanent study drug discontinuation or after retreatment. This censoring rule will be applied to efficacy endpoints conducted for the Period 2 population corresponding to withdrawal and downtitration substudy.

Table JAHN.6.2 summarizes the various imputation techniques being used for select efficacy analyses. Sections 6.4.1 through 6.4.3 summarize the imputation methods.

Table JAHN.6.2. Imputation Techniques for Various Efficacy Variables

Population	Endpoints[a]	Imputation Method[b]
Responder and partial responder mITT Period 1	IGA(0,1), IGA(0,1,2), EASI75, 4-point Itch NRS improvement	NRI mLOCF
	Exploratory endpoints	NRI (categorical endpoints) mLOCF
Nonresponder mITT Period 1	IGA(0,1), IGA(0,1,2), EASI75, 4-point Itch NRS improvement	NRI mLOCF
	Exploratory endpoints	NRI (categorical endpoints), MMRM (continuous endpoints) mLOCF
mITT Period 2	IGA(0,1), IGA(0,1,2), EASI75	NRI mLOCF
	Exploratory endpoints	NRI (categorical endpoints) mLOCF

Abbreviations: EASI = Eczema Area and Severity Index; IGA = Investigator's Global Assessment for atopic dermatitis; mITT = modified intent-to-treat; mLOCF = modified last observation carried forward; MMRM = mixed-model repeated measures; NRI = nonresponder imputation; NRS = Numeric Rating Scale.

[a] Refer to Table JAHN.6.6 and Table JAHN.6.8.

[b] All analysis with censoring.

6.4.1. Nonresponder Imputation

An NRI method imputes missing values as nonresponses and can be justified based on the composite strategy for handling intercurrent events (ICH E9 R1). This imputation procedure assumes the effects of treatment disappear after an occurrence of the intercurrent event (described in Section 6.4). For analyses which utilize censoring, randomized patients without at least 1 postbaseline observation will be defined as nonresponders for all visits. As well, patients who are missing a value prior to discontinuation or were rescued to a higher dose (if censoring on the latter) (ie, the patient is missing an intermediate visit) will be imputed as nonresponders on that visit only.

6.4.2. Mixed-Model for Repeated Measures

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

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

MMRM is executed after application of any censoring rules.

6.4.3. Modified Last Observation Carried Forward

An mLOCF imputation technique replaces missing data with the most recent nonmissing postbaseline assessment. The specific modification to the LOCF is data after an intercurent event will not be carried forward thus the mLOCF is applied after the specified censoring rule is implemented. The mLOCF assumes the effect of treatment remain the same after the event that caused missing data as it was just prior to the missing data event. Analyses using mLOCF require a nonmissing baseline and at least 1 postbaseline measure otherwise the data is missing for analyses purposes. Analyses using mLOCF help ensure the number of randomized patients who were assessed postbaseline is maximized and is reasonable for this data as data directly prior to an intercurrent event (such as initiation of rescue therapy or drop out) is likely a nonefficacious response.

6.4.4. Missingness Due to COVID-19

For missingness that Covid-19 is identified as the cause by the study team, the following imputation methods will be applied.

- For continuous measures, if missing due to Covid-19,
    - when they are analyzed using ANCOVA for a single timepoint, mLOCF method will be used.
    - when they are analyzed using MMRM with repeated measures, no imputation will be done.
- For categorical variables, if missing due to Covid-19,
    - mLOCF method will be used. This applies to categorical variables that are collected categorically, as well as categorical variables that are derived from measures that are collected as continuous, such as EASI50, EASI75, and EASI 90.
    - for primary and key efficacy measures including IGA 0 or 1 and EASI75, multiple imputation method will be used. Seed = 123456 will be used for multiple imputation. Results from multiple imputations will be aggregated to generate the final estimates for statistical inferences. This will be the main imputation method for these primary and key efficacy measures for visits with missing data due to Covid-19.

All measures collected through ePRO diary device should not be impacted by Covid-19.

For a specific database lock, if Covid-19 is still ongoing at the time of the database lock, the list of visits with missing data due to Covid-19 identified by the study for that lock may not be complete and may be updated in future locks. The imputation for missingness due to Covid-19 for each database lock is based on the list we receive for that specific lock.

For the lock to occur on 20 May 2020, as there are very limited missing data due to Covid-19, the imputation method specified in this section (Section 6.4.4) will not be applied to this lock.

6.5. Multicenter Studies

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

6.6. Multiple Comparisons/Multiplicity

As this study is designed to assess the long-term efficacy and safety of baricitinib in patients with AD, no adjustments for multiple comparisons will be utilized in the statistical analyses for this study.

6.7. Patient Disposition

Period 1:

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.

Patient disposition through Weeks 16, 36, and 52 will be summarized using the mITT population. Frequency counts and percentages of patients who complete the study treatment visits or discontinue early from the study along with whether they completed follow-up or did not complete follow-up will be summarized separately by treatment group for patients who are responder mITT, partial responder mITT, and for patients who are nonresponder mITT, along with their reason for study discontinuation. Rescue status for the responder mITT and partial responder mITT population will also be summarized. Frequency counts and percentages of patients who complete the treatment or discontinue treatment early will also be summarized separately by treatment group for patients who are responder mITT, partial responder mITT, and for patients who are nonresponder mITT, along with their reason for treatment discontinuation and their rescue status.

A listing of patient disposition will be provided for the ITT population, with the extent of their participation in the study and the reason for discontinuation. A listing of the ITT population with their treatment assignments for both treatment periods will also be provided.

Period 2:

For Period 2, disposition for patients entering the downtitration and withdrawal substudy will be reported similarly as Period 1 above. Patients not entering the downtitraion and withdrawal substudy will be reported using listings.

6.8. Patient Characteristics

Patient characteristics including demographics, baseline characteristics and pre-existing conditions will be summarized descriptively by treatment group. Period 1 summaries will be presented for the responder mITT, partial responder mITT, and nonresponder mITT populations. Period 2 summaries will be presented for mITT patients in each treatment group in the downtitration and withdrawal substudy. No formal statistical comparisons will be made among treatment groups unless otherwise stated.

6.8.1. Demographics

Patient demographics will be based on entry into originating study and summarized as described above. The following demographic information will be included:

- age
- age group (<65 versus ≥65)
- age group (<65, ≥65 to <75, ≥75 to <85, ≥85)
- gender (male, female)
- race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Multiple)
- region (as defined in Table JAHN.5.1)
- country
- weight (kg)
- weight category (<60 kg, ≥60 to <100 kg, ≥100 kg)
- height (cm)
- BMI (kg/m^2)
- BMI category (<25 kg/m^2, ≥25 to <30 kg/m^2, ≥30 kg/m^2)

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

6.8.2. Baseline Disease Characteristics

Baseline disease characteristics will be reported where the baseline is from the originating study. The following baseline disease information (but not limited to only these) 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 to <5 years, 5 to <10 years, 10 to <20 years, ≥20 years)
- age at diagnosis (years)
- age group at diagnosis (<18 years, ≥18 to <50 years, ≥50 years)
- validated IGA for AD score
- EASI score
- SCORAD
- body surface area affected by AD
- HADS subscales
- POEM
- Itch NRS
- ADSS Item 2
- DLQI
- Skin Pain NRS
- PGI-S-AD
- prior therapy using originating study baseline (topical therapy only; systemic therapy only; topical and systemic therapy)
- use of cyclosporine prior to entry into originating study (yes, no)

- cyclosporine inadequate response prior to entry into originating study (yes, no)
- cyclosporine intolerance prior to entry into originating study (yes, no)
- cyclosporine contraindication (ineligible) prior to entry into originating study (yes, no)
- cyclosporine inadvisable prior to entry into originating study (yes, no)
- use of TCNI prior to entry into originating study (yes, no)
- TCNI inadequate response prior to entry into originating study (yes, no)
- TCNI intolerance prior to entry into originating study (yes, no)
- TCNI contraindication (ineligible) prior to entry into originating study (yes, no)
- TCNI inadvisable prior to entry into originating study (yes, no)
- vaccine prior to entry into originating study (yes, no)
- originating study 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: intrinsic (<200 kU/I) or extrinsic (≥200 kU/I)

6.8.3. Historical Illness and Preexisting Conditions

Historical illnesses are described in the originating studies and no additional summaries will be created for Study JAHN.

Preexisting conditions are defined as those conditions recorded in the Preexisting Conditions and Medical History electronic case report form (eCRF), or the Adverse Events eCRF with a start date prior to the first dose of the study drug and stop dates that are at or after the informed consent date or have no stop date (i.e., are ongoing). For events occurring on the day of the first dose of study drug, the date and time of the onset of the event will both be used to determine if the event was preexisting. Conditions with a partial or missing start date (or time if needed) will be assumed to be 'not preexisting' unless there is evidence, through comparison of partial dates, to suggest otherwise. Preexisting conditions will be categorized using the Medical Dictionary for Regulatory Activities (MedDRA®, most current available version) algorithmic standardized MedDRA queries or similar predefined lists of Preferred Terms (PTs) of interest. Frequency counts and percentages of patients with selected preexisting conditions will be summarized by treatment group. Analyses will be presented for the responder mITT, partial responder mITT and nonresponder mITT populations.

6.9. Treatment Compliance

Period 1 patient compliance with study medication will be assessed from Week 0 (Visit 2) to Week 52 (Visit 8) (or Early Termination if it occurs prior to Week 52), using the mITT population. Period 2 compliance will be summarized by treatment group up to Week 16 for the US filing.

All patients are expected to take 3 tablets daily from a blisterpack as described in Section 5.2. Each blisterpack 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

Descriptive statistics for percent compliance and noncompliance rate will be summarized for the mITT population by treatment group for both study periods. 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 for all mITT patient for Week 0 to 52, and Week 52 to 68 for patients entering the downtitration and withdrawal substudies.

6.9.1. Background Therapy

Background TCS therapy with low-potency and/or moderate-potency TCS (eg, hydrocortisone 2.5% ointment and/or triamcinolone 0.1% cream) may be used on active lesions. Such TCS therapy may be provided by the sponsor or the investigative sites. For more detail, see Section 7.7.1 in the protocol.

A summary of the reasons initial background therapy is used will be produced, as well as a summary of the proportion of patients with sponsor-provided TCS at each study visit and a summary of other AD therapy for the responder and partial responder mITT population. Also, a summary of time to initial use of AD background therapy postbaseline will be produced for the responder and partial responder mITT population.

The number of days TCS therapy is used for each patient is also collected on the diary device. The diary device is only dispensed up until Week 16 and is redispensed for patients who are eligible for the randomized withdrawal and downtitration substudy for an additional 16 weeks of the substudy. The proportion of time that the patients did not use TCS therapy will be summarized for these visit intervals (i.e., from Week 0 to Week 16 for the responder mITT, partial responder mITT, and nonresponder mITT patients provided with low or moderate TCS). For this analysis, the date of the first entry on the diary device will be used to signify the first day of TCS therapy use.

Descriptive statistics for drug accountability of sponsor-provided TCS (topical low and moderate corticosteroids) will also be provided, including the amount utilized (dispensed tube with cap weight minus the weight of the returned tube with cap) throughout the treatment period from Week 0 to Weeks 16, 36, and 52. The total amount utilized in grams for low and moderate potency will be summarized between visits (Week 0 through Week 4, Week 4 through Week 8,

Week 8 through Week 16, and so on) as well as throughout the treatment period from Week 0 through Week 52 (Period 1) and Week 52 through 68 (Period 2). The total amount utilized will also be presented for the aforementioned visit intervals, irrespective of potency. Analyses will be summarized for the responder mITT, partial responder mITT, and nonresponder mITT populations for Period 1. Additionally a summary of total amount used will be provided for the first 16 Weeks of Period 2 by treatment group for those that entered the randomized withdrawal and downtitration studies.

The dispensed weight of sponsor-provided TCS full tubes used for background therapy for the 2 different potencies (low and moderate) varies between countries due to different supply regions. Average weights of full tubes were used to determine the dispensed weights for each region. Returned tubes were weighed with cap (without the carton) to determine the amount of TCS in grams (g) used between visits. For low and moderate potency TCS, the dispensed tube weight with cap (without the carton) in Japan is 13.5 g. For countries supplied by European distributors (Austria, Czech Republic, Denmark, France, Germany, Hungary, Italy, Poland, Spain, and Switzerland), the dispensed weight of low potency TCS is 21 g and moderate potency TCS is 38 g. The remaining countries, supplied by US distributors (Argentina, Australia, India, Israel, Korea, Mexico, Russia, and Taiwan), the weight of low and moderate potency TCS is 40 g.

If a returned tube is not weighed in grams or the tube is not returned then the tube can be classified as partially used, fully used, unused, or unknown. Partially used TCS medication tubes will be considered to be 50% used whereas fully used and unused tubes will be considered as 100% and 0% used, respectively. When drug accountability is not performed for a particular tube of TCS 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.

6.10. Previous and Concomitant Therapy

Previous medications are described in the originating studies.

Summaries of concomitant medications used for AD will be based on the mITT population.

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

A summary of concomitant therapy including concomitant therapies of special interest will be provided for the mITT population.

6.11. Efficacy Analyses

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

Efficacy analyses will generally be analyzed according to the groups in Table JAHN.6.3 and patients will be analyzed according to the IP being analyzed in the objective.

Table JAHN.6.3. Efficacy Analyses: Comparisons or Treatment Paradigms

Treatment Period	Treatment Groups: Originating Study Treatment to JAHN Treatment	Comparison or Treatment Paradigms	Analysis
Treatment Period 1 (main JAHN protocol)	**Responders and Partial Responders:** PBO to PBO (PBO), BARI 1 mg to BARI 1 mg (BARI 1 mg), BARI 2 mg to BARI 2 mg (BARI 2 mg), BARI 4 mg to BARI 4 mg (BARI 4 mg)	PBO, BARI 1 mg, BARI 2 mg, BARI 4 mg; descriptive	Primary, Secondary, Exploratory
	Nonresponders: PBO to BARI 2 mg, PBO to BARI 4 mg, BARI 1 mg to BARI 2 mg, BARI 1 mg to BARI 4 mg, BARI 2 mg to BARI 2 mg, BARI 2 mg to BARI 4 mg, BARI 4 mg to BARI 4 mg	PBO to BARI 4 mg vs PBO to BARI 2 mg; BARI 1 mg to BARI 4 mg vs BARI 1 mg to BARI 2 mg; BARI 2 mg to BARI 4 mg vs BARI 2 mg to BARI 2 mg	Secondary, Exploratory
Treatment Period 2 (Randomized Withdrawal and Downtitration Study	**Treatment Groups:** Bari 4 mg at Week 52 Bari 4 mg to Bari 2 mg Bari 4 mg to PBO Bari 2 mg at Week 52 Bari 2 mg to Bari 1 mg Bari 2 mg to PBO	 Bari 4 mg to Bari 2 mg Bari 4 mg to PBO Bari 2 mg to Bari 1 mg Bari 2 mg to PBO	Secondary

Abbreviations: BARI = baricitinib; PBO = placebo.

Table JAHN.6.4 includes the descriptions and derivations of the primary, secondary, and exploratory efficacy outcomes.

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

Table JAHN.6.4. Description and Derivation of Primary, Secondary, and Exploratory Efficacy Outcomes

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
Validated Investigator's Global Assessment for AD (IGA)	The validated IGA 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: 0-4 0 represents "clear" 4 represents "severe"	Single item, missing if missing.
		▪ IGA (0,1) ▪ IGA (0,1,2)	▪ Observed score of 0 or 1 ▪ Observed score of 0, 1, or 2	Single item, missing if missing.
		Relapse	If IGA score at any visit after Week 52 is observed to be >2 then relapse = 'Y'	Missing if all post-Week 52 IGA assessments are missing
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 0-3 (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 0-72, and the final EASI score will be obtained by weight-averaging these 4 scores. Hence, the final EASI score will range from 0-72 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 0-3 and value from percentage involvement is on a scale of 0-6. 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.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
		▪ Change from baseline in EASI score ▪ Percent change from baseline in EASI score	Change from baseline: observed EASI score – baseline EASI score % change from baseline: $100 \times \frac{Observed\ score - Baseline}{Baseline}$	Missing if baseline or observed value is missing.
		EASI50	% Improvement in baseline ≥50%: % change from baseline ≤ -50	Missing if baseline or observed value is missing.
		EASI75	% Improvement in baseline ≥75%: % change from baseline ≤ -75	Missing if baseline or observed value is missing.
		EASI90	% Improvement in baseline ≥90%: % change from baseline ≤ -90	Missing if baseline or observed value is missing.
		Rebound	If EASI percentage change from originating study baseline at any visit after Week 52 is observed to be ≥125, then disease rebound = 'Y'	Missing if baseline or observed value is missing.
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%-100% involvement. The overall total percentage will be reported based off of all 4 body regions combined, after applying specific multipliers to the different body regions to account for the percent of the total BSA represented by each of the 4 regions.	BSA score	Use the percentage of skin affected for each region (0-100%) 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 (SCORAD) index uses the rule of nines to assess disease extent (head and neck 9%; upper limbs 9% each; lower limbs 18% each; anterior trunk 18%; back 18%; and genitals 1%). It evaluates 6 clinical characteristics to determine disease severity: 1) erythema,	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

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
	2) edema/papulation, 3) oozing/crusts, 4) excoriation, 5) lichenification, and 6) dryness on a scale of 0-3 (0=absence, 1=mild, 2=moderate, 3=severe). The SCORAD index also assesses subjective symptoms of pruritus and sleep loss in the last 72 hours on visual analogue scales (VAS) of 0-10 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).			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.
		SCORAD75	% Improvement in baseline ≥75%: % change from baseline ≤ -75	Missing if baseline or observed value is missing.
		SCORAD90	% Improvement in baseline ≥90%: % change from baseline ≤ -90	Missing if baseline or observed value is missing.

Abbreviations: AD = atopic dermatitis; N/A = not applicable.

Table JAHN.6.5. Description of Primary, Secondary, and Exploratory Efficacy Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Validated Investigator's Global Assessment for AD (IGA)	Proportion of patients with a response of IGA 0 or 1	mITT population from originating studies	Descriptive using NRI	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Primary (Weeks 16, 36, and 52)
		mITT population from originating combination therapy study	Descriptive using NRI	Week 24	Exploratory
		mITT population from originating combination therapy study	Descriptive using mLOCF	Weeks 0-24	Sensitivity
		mITT population from originating studies - nonresponders	Logistic regression using NRI	Weeks 0-16 (combination therapy study) Week 0-52 (monotherapy studies)	Secondary
		mITT population entering the downtitration and withdrawal study (Period 2)	Logistic regression using NRI	Week52 Week 56 Week 60 Week 64 Week 68	Secondary for Week 68 Exploratory for Weeks 56, 60, and 64
		mITT population entering the downtitration and withdrawal study (Period 2) with Week 52 IGA of 0 or 1	Logistic regression using NRI	Week 56 Week 60 Week 64 Week 68	Secondary for Week 68 Exploratory for Weeks 56, 60, and 64
	Proportion of patients with a response of IGA 0,1, or 2	mITT population from originating studies	Descriptive using NRI	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Secondary

Description of Primary, Secondary, and Exploratory Efficacy Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Validated Investigator's Global Assessment for AD (IGA)	Proportion of patients with a response of IGA 0,1, or 2	mITT population from originating combination therapy study mITT population from open-label study	Descriptive using NRI	Week 24	Exploratory
		mITT population from originating combination therapy study	Descriptive using mLOCF	Weeks 0-24	Sensitivity
		mITT Population from originating studies - nonresponders	Logistic regression using NRI	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Secondary for Week 68 Exploratory for Weeks 56, 60, and 64
		mITT population entering the downtitration and withdrawal study (Period 2)	Logistic regression using NRI	Week 56 Week 60 Week 64 Week 68	Secondary for Week 68 Exploratory for Weeks 56, 60, and 64
		mITT population entering the downtitration and withdrawal study (Period 2) with Week 52 IGA of 0 or 1	Logistic regression using NRI	Week 56 Week 60 Week 64 Week 68	Exploratory
	Time to retreatment (time to IGA ≥3)	mITT population entering the downtitration and withdrawal study (Period 2)	Descriptive	Weeks 52-68	Secondary

Description of Primary, Secondary, and Exploratory Efficacy Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Validated Investigator's Global Assessment for AD (IGA)	Proportion of patients with IGA >2 within 16 weeks after disease relapse	mITT population entering the downtitration and withdrawal study (Period 2)		Weeks 52-68	Exploratory
Eczema Area and Severity Index (EASI)	Change from baseline in EASI score Percent change from baseline in EASI score	mITT population from originating studies	Descriptive	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Exploratory
		mITT population from originating combination therapy study	Descriptive	Week 24	Exploratory
		mITT population from originating studies - nonresponders	MMRM	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Exploratory
		mITT population from originating combination therapy study	Descriptive using mLOCF	Weeks 0-24	Sensitivity
	Proportion of patients achieving EASI percent change from baseline < -75	mITT population from originating studies	Descriptive using NRI	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Secondary
		mITT population from originating combination therapy study	Descriptive using NRI	Week 24	Exploratory

Description of Primary, Secondary, and Exploratory Efficacy Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Eczema Area and Severity Index (EASI)	Proportion of patients achieving EASI percent change from baseline < -75	mITT population from originating studies - nonresponders	Logisitc regression using NRI	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Secondary
		mITT population from originating studies	Descriptive using mLOCF	Weeks 0-24 (combination therapy study) Week 0-52 (monotherapy studies)	Sensitivity
		mITT population entering the downtitration and withdrawal study (Period 2)	Logistic regression using NRI	Week 52 Week 56 Week 60 Week 64 Week 68	Secondary
		mITT population entering the downtitration and withdrawal study (Period 2) with Week 52 IGA of 0 or 1	Logistic regression using NRI	Week 52 Week 56 Week 60 Week 64 Week 68	Exploratory
	▪ Proportion of patient with disease rebound	mITT population entering the downtitration and withdrawal study (Period 2)	Logistic regression	Up to Week 68	Exploratory
	▪ Proportion of patients achieving EASI75 within 16 weeks of disease relapse	mITT population entering the downtitration and withdrawal study (Period 2)	Descriptive	Up to Week 68	Secondary

Description of Primary, Secondary, and Exploratory Efficacy Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Eczema Area and Severity Index (EASI)	▪ Proportion of patients achieving EASI50 ▪ Proportion of patients achieving EASI90	mITT population from originating studies	Descriptive using NRI	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Exploratory
		mITT population from originating combination therapy study	Descriptive using NRI	Week 24	Exploratory
		mITT population from originating studies – nonresponders	Logistic regression using NRI	Weeks 0-16 (combination therapy study) Week 0-52 (monotherapy studies)	Exploratory
Body Surface Area (BSA) Affected by AD	▪ BSA score ▪ Change from baseline in BSA score	mITT population from originating studies	Descriptive, as observed	Weeks 0-16 (combination therapy study) Week 0-52 (monotherapy studies)	Exploratory
		mITT population from originating combination therapy study	Descriptive, as observed	Week 24	Exploratory
		mITT population from originating studies – nonresponders	MMRM	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Exploratory

Description of Primary, Secondary, and Exploratory Efficacy Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
SCORing Atopic Dermatitis (SCORAD)	▪ SCORAD score ▪ Change from baseline in SCORAD score ▪ Percent change from baseline in SCORAD score	mITT population from originating studies	Descriptive, as observed	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Exploratory
		mITT population from originating combination therapy study	Descriptive, as observed	Week 24	Exploratory
		mITT population from originating studies – nonresponders	MMRM	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Exploratory
	▪ Proportion of patients achieving SCORAD75 ▪ Proportion of patients achieving SCORAD90	mITT population from originating studies	Descriptive using NRI	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Exploratory
		mITT population from originating combination therapy study	Descriptive using NRI	Week 24	Exploratory
		mITT population from originating studies – nonresponders	Logisitic regresion using NRI	Weeks 0-16 (combination therapy study) Weeks 0-52 (monotherapy studies)	Exploratory

Description of Primary, Secondary, and Exploratory Efficacy Analyses

Abbreviations: AD = atopic dermatitis; mITT = modified intent-to-treat; mLOCF = modified last observation carried forward; MMRM = mixed-model repeated measures; NRI = nonresponder imputation.

[a] Populations from originating studies for reporting periods of Weeks 0-52 are on the mITT population; they are reported by responder status from originating study: responders, partial responders, responders + partial responders and nonresponders. Populations for visitwise analyses for those who entered the randomized withdrawal and downtitration study are based on efficacy evaluable sets for each of Week 56, 60, 64, and 68 (defined in Section 6.2.1).

[b] All categorical endpoints use NRI, unless otherwise stated.

6.11.1. Primary Outcome and Methodology

The validated IGA for AD uses the clinical characteristics of erythema, papulation/induration, oozing/crusting, and lichenification to produce a single-item score ranging from 0 to 4. The primary analysis of the study is to estimate the effect of long-term therapy with baricitinib on responders and partial responders at the entry of Study JAHN by evaluating the proportion of patients with a response of IGA 0 or 1 assessed at Weeks 16, 36, and 52 using the mITT population, assuming that treatment response disappears after the patient is rescued or discontinued study or study treatment. This will serve as the primary estimand. In this estimand, missing data due to the application of the censoring rule and the occurrence of other noncensor intercurrent events will be imputed using the NRI method described in Section 6.4.1. A logistic regression analysis as described in Section 6.2.3 will be used for the comparisons. The odds ratio, the corresponding 95% CIs and p-value, as well as the treatment differences and the corresponding 95% CIs, will be reported.

6.11.2. Secondary and Exploratory Efficacy Analyses

There will be no adjustment for multiple comparisons for any analyses. The secondary and exploratory efficacy analyses are detailed in Table JAHN.6.4. Health outcomes analyses are described in Section 6.12.

For patients entering the downtitration and withdrawal substudy who need to be retreated, a summary of treatment responses containing proportion of patients with IGA (0, 1, 2), IGA (0, 1), and EASI75 response will be provided. A summary of the time to retreatment will also be provided.

Note that the exploratory efficacy analyses on a continuous outcome that is measured over time use a slightly different estimand which is to estimate the effect of long-term therapy with baricitinib on a specific endpoint evaluated at specified time points using the mITT population assuming all patients remained in their treatment throughout the defined period of the study.

6.11.3. Sensitivity Analyses

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

- analysis of select continuous outcomes with ANCOVA (Section 6.2.3), with missing data imputed using mLOCF (Section 6.4.3)
- analysis of select continuous outcomes in the randomized withdrawal and downtitration substudy using Week 52 as the baseline

6.12. Health Outcomes/Quality-of-Life Analyses

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

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

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

Table JAHN.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.

Table JAHN.6.6. Description and Derivation of Health Outcomes and Quality-of-Life Measures

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
Itch Numeric Rating Scale (NRS)	The Itch Numeric Rating Scale (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 weekly score.	Refer to Section 6.2.2 on how to derive the weekly 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.
Skin Pain Numeric Rating Scale (NRS)	Skin Pain NRS is a patient-administered, 11-point horizontal scale anchored at 0 and 10, with 0 representing "no pain" and 10 representing "worst pain imaginable." Overall severity of a patient's skin pain is indicated by selecting the number that best describes the worst level of skin pain in the past 24 hours. Refer to Section 6.2.2 for details on how to calculate the weekly score which will be used in the continuous analysis.	▪ Skin Pain NRS score	Single item; range 0-10. Refer to Section 6.2.2 on how to derive the weekly score.	Refer to Section 6.2.2 on how to derive the weekly 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.
Atopic Dermatitis Sleep Scale (ADSS)	The Atopic Dermatitis Sleep Scale (ADSS) is a 3-item, patient-administered questionnaire developed to assess the impact of itch on sleep including difficulty	▪ Item 1 score of ADSS ▪ Item 2 score of ADSS ▪ Item 3 score of ADSS	Single items: Item 1, range 0-4; Item 2, range 0-29; Item 3, range 0-4. Refer to Section 6.2.2 on how to derive the weekly score.	Refer to Section 6.2.2 on how to derive the weekly score.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
	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 0-29 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.	▪ 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.
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 1-5. Refer to Section 6.2.2 on how to derive the weekly score.	Refer to Section 6.2.2 on how to derive the weekly score.
		Change from baseline in PGI-S-AD	Change from baseline: observed PGI-S-AD score – baseline PGI-S-AD score	Missing if baseline or observed value is missing.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
Patient-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 1-7, range 0-28.	If a single question is left unanswered, then that question is scored as 0. If more than 1 question is unanswered, then the tool is not scored. If more than 1 response is selected, then the response with the highest score is used.
		Change from baseline in POEM score	Change from baseline: observed POEM score – baseline POEM score	Missing if baseline or observed value is missing.
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 (e.g., 0-3) for each question and is intended for ages 12-65 years (Zigmond and Snaith 1983; White et al. 1999). Scores for each domain (anxiety and depression) can range from 0-21, 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 7 anxiety questions, range 0-21; Depression domain score is sum of the 7 depression questions, range 0-21.	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.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
Dermatology Life Quality Index (DLQI)	The Dermatology Life Quality Index (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 DLQI total score of 0-1 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).	▪ 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, and 0 respectively.	N/A – partial assessments cannot be saved.
		• Personal relationships domain	Sum of Questions 8 and 9, range 0-6.	N/A – partial assessments cannot be saved.
		• Treatment domain	Question 10, range 0-3.	N/A – partial assessments cannot be saved.
		• DLQI total score	DLQI total score: sum of all 6 DLQI domain scores, range 0-30.	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.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
Work Productivity and Activity Impairment: Atopic Dermatitis (WPAI-AD)	The Work Productivity and Activity Impairment Questionnaire–Atopic Dermatitis (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	Question (Q)1	Single item, missing if missing.
		• Change in employment status	Employed at baseline and remained employed: Q1 = 1 at postbaseline visit and at baseline visit. Not employed at baseline and remain unemployed: Q1 = 0 at postbaseline visit and at baseline visit.	Missing if baseline or observed value is missing.
		• Percentage of absenteeism	Percent work time missed due to problem: (Q2 / [Q2 + Q4]) * 100	If Q2 or Q4 is missing, then missing.
		• Change from baseline in absenteeism	Change from baseline: observed absenteeism – baseline absenteeism	Missing if baseline or observed value is missing.
		• Percentage of presenteeism	Percent impairment (reduced productivity while at work) while working due to problem: (Q5 / 10) * 100	If Q5 is missing, then missing.
		• Change from baseline in presenteeism	Change from baseline: observed presenteeism – baseline absenteeism	Missing if baseline or observed value is missing.
		• Overall work impairment	Percent overall work impairment (combines absenteeism and presenteeism) due to problem: (Q2 / [Q2+ Q4] + [(1 – Q2 / (Q2 + Q4)) * (Q5 / 10)]) * 100	If Q2, Q4, or Q5 is missing, then missing.
		• Change from baseline in work impairment	Change from baseline: observed work impairment – baseline work impairment	Missing if baseline or observed value is missing.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
		• Percentage of impairment in activities	Percent activity impairment (performed outside of work) due to problem: (Q6 / 10) * 100	If Q6 is missing, then missing.
		• 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 0-100 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 1-5 have no arithmetic properties and should not be used as an ordinal score. The VAS records the respondent's self-rated health on a vertical VAS where the endpoints are labeled "best imaginable health state" and "worst	▪ 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 1-5 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 0-100. 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.

Measure	Description	Variable	Derivation / Comment	Imputation Approach if Missing Components
	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]).	• 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 score (health state index)	Derive EQ-5D-5L US Population-based index score according to the link by using the US algorithm to produce a patient-level index score between -0.11 and 1.0 (continuous variable).	N/A – partial assessments 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.

Abbreviations: AD = atopic dermatitis; eCOA = electronic clinical outcomes assessment; N/A = not applicable, QoL = quality of life; UK = United Kingdom; US = United States.

Table JAHN.6.7. Description of Health Outcomes and Quality-of-Life Measures Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Itch Numeric Rating Scale (NRS)	▪ Itch NRS score ▪ Change from baseline in Itch NRS score ▪ Percent change from baseline Itch score	mITT population from originating combination therapy study	Descriptive	Weeks 0-16	Exploratory analysis
		Nonresponder mITT population from originating combination therapy study	MMRM	Weeks 0-16	Exploratory analysis
		mITT population entering the downtitration and withdrawal study (Period 2)	▪ ANCOVA regression using mLOCF ▪ ANCOVA regression using mLOCF with Week 52 baseline	▪ Week 52 ▪ Week 56 ▪ Week 60 ▪ Week 64 ▪ Week 68	Exploratory analysis
		mITT population entering the downtitration and withdrawal study (Period 2) with Week 52 IGA of 0 or 1	▪ ANCOVA regression using mLOCF ▪ ANCOVA regression using mLOCF with Week 52 baseline	▪ Week 52 ▪ Week 56 ▪ Week 60 ▪ Week 64 ▪ Week 68	Exploratory analysis
	▪ Proportion of patients with a 4-point improvement from baseline of originating study in Itch NRS in subgroup of patients who had baseline Itch NRS ≥4	mITT population from originating combination therapy study	Descriptive	Weeks 0-16	▪ Secondary analysis ▪ Exploratory analysis ▪ Exploratory analysis
		Nonresponder mITT population from originating combination therapy study	Logistic regression	Weeks 0-16	▪ Secondary analysis ▪ Exploratory analysis ▪ Exploratory analysis

Description of Health Outcomes and Quality-of-Life Measures Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Itch Numeric Rating Scale (NRS)	▪ Proportion of patients with a 4-point improvement from baseline of originating study in Itch NRS in subgroup of patients who had baseline Itch NRS ≥4	mITT population entering the downtitration and withdrawal study (Period 2)	Logistic regression using NRI	▪ Week 52 ▪ Week 56 ▪ Week 60 ▪ Week 64 ▪ Week 68	Exploratory analysis
		mITT population entering the downtitration and withdrawal study (Period 2) with Week 52 IGA of 0 or 1	Logistic regression using NRI	▪ Week 52 ▪ Week 56 ▪ Week 60 ▪ Week 64 ▪ Week 68	Exploratory analysis
Skin Pain Numeric Rating Scale (NRS)	▪ Skin Pain NRS score ▪ Change from baseline in Skin Pain NRS score	mITT population from originating combination therapy study	Descriptive	Weeks 0-16	Exploratory analysis
		Nonresponder mITT population from originating combination therapy study	MMRM	Weeks 0-16	Exploratory analysis
		mITT population entering the downtitration and withdrawal study (Period 2)	▪ ANCOVA regression using mLOCF ▪ ANCOVA regression using mLOCF with Week 52 baseline	▪ Week 52 ▪ Week 56 ▪ Week 60 ▪ Week 64 ▪ Week 68	Exploratory analysis
		mITT population entering the downtitration and withdrawal study (Period 2) with Week 52 IGA of 0 or 1	▪ ANCOVA regression using mLOCF ▪ ANCOVA regression using mLOCF with Week 52 baseline	▪ Week 52 ▪ Week 56 ▪ Week 60 ▪ Week 64 ▪ Week 68	Exploratory analysis

Description of Health Outcomes and Quality-of-Life Measures Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Skin Pain Numeric Rating Scale (NRS)	▪ Proportion of patients with a 4-point improvement from baseline of originating study in Skin Pain in subgroup of patients who had baseline Skin Pain NRS ≥4	mITT population from originating combination therapy study	Descriptive	Weeks 0-16	Exploratory analysis
		Nonresponder mITT population from originating combination therapy study	Logistic regression	Weeks 0-16	Exploratory analysis
Atopic Dermatitis Sleep Scale (ADSS)	▪ ADSS item scores ▪ Change from baseline in ADSS item scores	mITT population from originating combination therapy study	Descriptive	Weeks 0-16	Exploratory analysis
		Nonresponder mITT population from originating combination therapy study	MMRM	Weeks 0-16	Exploratory analysis
	▪ Proportion of patients with a 1-point improvement from baseline of originating study in a subgroup of patients with baseline item score of ≥1 for the following ADSS item scores: - Item 1 score - Item 2 score - Item 3 score ▪ Proportion of patients with a 2-point improvement from baseline of originating study in item 2 score of ADSS in a subgroup of patients with baseline item 2 score of ≥2	mITT population from originating combination therapy study	Descriptive	Weeks 0-16	Exploratory analysis
		Nonresponder mITT population from originating combination therapy study	Logistic regression	Weeks 0-16	Exploratory analysis

Description of Health Outcomes and Quality-of-Life Measures Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Patient Global Impression of Severity–Atopic Dermatitis (PGI-S-AD)	▪ PGI-S-AD score ▪ Change from baseline in PGI-S-AD score	mITT population from originating combination therapy study	Descriptive	Weeks 0-16	Exploratory analysis
		Nonresponder mITT population from originating combination therapy study	MMRM	Weeks 0-16	Exploratory analysis
Patient-Oriented Eczema Measure (POEM)	▪ POEM score ▪ Change from baseline in POEM score	mITT population from originating studies	Descriptive, as observed	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
		mITT population from originating combination therapy study	Descriptive, as observed	Week 24	Exploratory analysis
		mITT population from originating studies	MMRM	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
	▪ Proportion of patients with a 4-point improvement in POEM score from baseline of originating study in a subgroup of patients with baseline POEM ≥4	mITT population from originating studies	Descriptive using NRI	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis

Description of Health Outcomes and Quality-of-Life Measures Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Patient-Oriented Eczema Measure (POEM)	▪ Proportion of patients with a 4-point improvement in POEM score from baseline of originating study in a subgroup of patients with baseline POEM ≥4	mITT population from originating combination therapy study	Descriptive using NRI	Week 24 (combination therapy study)	Exploratory analysis
		Nonresponder mITT population from originating studies	Logistic regression using NRI	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
Hospital Anxiety Depression Scale (HADS)	▪ Observed and change from baseline in HADS domain scores -Anxiety -Depression ▪ Change from baseline in HADS total score	mITT population from originating studies	Descriptive, as observed	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
		mITT population from originating combination therapy study	Descriptive, as observed	Week 24	Exploratory analysis

Description of Health Outcomes and Quality-of-Life Measures Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Hospital Anxiety Depression Scale (HADS)	▪ Observed and change from baseline in HADS domain scores -Anxiety -Depression ▪ Change from baseline in HADS total score	Nonresponder mITT population from originating studies	MMRM	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
	▪ Proportion of patients with HADS domain (anxiety and depression) score of <8 in a subgroup of patients with baseline HADS domain (anxiety and depression) score of ≥8 ▪ Proportion of patient with HADS anxiety or depression domain score of <8 in a subgroup of patients with baseline HADS anxiety or depression doman score of ≥8	mITT population from originating studies	Descriptive using NRI	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
		mITT population from originating combination therapy study	Descriptive using NRI	Week 24	Exploratory analysis
		Nonresponder mITT population from originating studies		▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis

Description of Health Outcomes and Quality-of-Life Measures Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Dermatology Life Quality Index (DLQI)	▪ DLQI total score ▪ Change from baseline in DLQI ▪ Observed and change from baseline in domain scores -Symptoms and feelings -Daily activities -Leisure -Work and school -Personal relationships -Treatment	mITT population from originating studies	Descriptive, as observed	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
		mITT population from originating combination therapy study	Descriptive, as observed	Week 24	Exploratory analysis
		Nonresponder mITT population from originating studies	MMRM	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
	▪ Proportion of patients with DLQI (0,1) ▪ Proportion of patients with a 4-point improvement from baseline in originating study in a subgroup of patients with baseline DLQI score ≥4	mITT population from originating studies	Descriptive using NRI	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
		mITT population from originating combination therapy study	Descriptive using NRI	Week 24	Exploratory analysis

Description of Health Outcomes and Quality-of-Life Measures Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Dermatology Life Quality Index (DLQI)	▪ Proportion of patients with DLQI (0,1) ▪ Proportion of patients with a 4-point improvement from baseline in originating study in a subgroup of patients with baseline DLQI score ≥4	Nonresponder mITT population from originating studies	Logistic regression using NRI	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
Work Productivity and Activity Impairment: Atopic Dermatitis (WPAI-AD)	Observed and change from baseline in employment status	mITT population from originating studies	Descriptive, as observed	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
		mITT population from originating combination therapy study	Descriptive, as observed	Week 24	Exploratory analysis
		Nonresponder mITT population from originating studies	MMRM	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
	Observed and change from baseline in: ▪ absenteeism ▪ presenteeism ▪ overall work impairment ▪ impairment in activities	mITT population from originating studies		▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis

Description of Health Outcomes and Quality-of-Life Measures Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
Work Productivity and Activity Impairment: Atopic Dermatitis (WPAI-AD)	Observed and change from baseline in: ▪ absenteeism ▪ presenteeism ▪ overall work impairment impairment in activities	mITT population from originating combination therapy study	Descriptive, as observed	Week 24	Exploratory analysis
		Nonresponder mITT population from originating studies	MMRM	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory 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	mITT population from originating studies	Descriptive, as observed	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
		mITT population from originating combination therapy study	Descriptive, as observed	Week 24	Exploratory analysis
		Nonresponder mITT population from originating studies	Logistic regression	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis

Description of Health Outcomes and Quality-of-Life Measures Analyses

Measure	Variable	Population[a] (Section 6.2.1)	Analysis Method[b] (Section 6.2.3)	Time Point	Analysis Type
European Quality of Life–5 Dimensions–5 Levels (EQ-5D-5L)	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	mITT population from originating studies	Descriptive, as observed	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis
		mITT population from originating combination therapy study	Descriptive, as observed	Week 24	Exploratory analysis
		Nonresponder mITT population from originating studies	MMRM	▪ Weeks 0-16 (combination therapy study) ▪ Week 0-52 (monotherapy studies)	Exploratory analysis

Abbreviations: ANCOVA = analysis of covariance; IGA = Investigator's Global Assessment; mITT = modified intent-to-treat; mLOCF = modified last observation carried forward; MMRM = mixed-model repeated measures; NRI = nonresponder imputation; VAS = visual analog scale.

[a] Populations from originating studies for reporting periods of Week 0-52 are on the mITT population; they are reported by responder status from originating study: responders, partial responders, responders + partial responders, and nonresponders.

[b] All categorical endpoints use NRI, unless otherwise stated.

6.13. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

Pharmacokinetic (PK), pharmacodynamic (PD), and biomarker analyses will not be conducted.

6.14. Safety Analyses

Detailed analyses and discussion of Study JAHN safety data are more thoroughly assessed in the context of combining the safety data from the originating Studies, JAHL, JAHM, and JAIY with the safety data from Study JAHN. The planned analyses are described in the compound level safety standards.

For the purpose of Study JAHN alone, the following are planned to be analyzed:

- duration of exposure by Treatment Periods 1 and 2
- overview of AEs by Treatment Periods 1 and 2
- treatment-emergent adverse events (TEAEs) by PT nested within System Organ Class (SOC) by Treatment Periods 1 and 2
- serious adverse events (SAEs) by PT nested within SOC by Treatment Periods 1 and 2, and listing of SAEs in Treatment Periods 1 and 2 combined (one listing)
- AEs leading to permanent study drug discontinuation by PT nested within SOC for Treatment Period 1, and listing of AEs leading to permanent study drug discontinuation and discontinuation from the study in Treatment Periods 1 and 2 combined (one listing)
- listing of deaths in Treatment Periods 1 and 2 combined (1 listing), if any
- clinical laboratory evaluation and vital signs in terms of box plots, descriptive and treatment-emergent summaries in Treatment Period 1
- clinical laboratory and vital signs in terms of box plots and descriptive summaries in Treatment Periods 1 and 2 combined

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

Safety analyses will be conducted according to the safety population defined in Section 6.2.1 and the treatment groups defined in Table JAHN.6.1 in Section 6.2.1.

Unless otherwise specified, by-visit summaries will include planned on-treatment visits. For tables that summarize events (such as AEs), on-treatment and follow-up data will be included up to 30 days after the last dose of treatment or up to dose or treatment change, where applicable.

6.14.1. Extent of Exposure

Duration of exposure (in days) for Treatment Periods 1 and 2 will be calculated as follows:

- duration of exposure to IP in Treatment Period 1: minimum (date of last dose of study drug, Week 52 date) – date of first dose of study drug in Study JAHN + 1.
- duration of exposure to IP in Treatment Period 2: minimum (date of last dose of study drug [>Week 52 date], >Week 52 date) – date of first dose of study drug in Study JAHN in Treatment Period 2 + 1

Last dose of study drug is calculated as last date on study drug or censored at dose or treatment change according to the treatment groups defined in Section 6.2.1. For patients discontinuing

treatment and the treatment stop date is missing, the duration of exposure will use date of last visit in exposure calculation. For patients whose treatment discontinuation is unknown because they were lost to follow-up, the date of the visit (scheduled or unscheduled) just prior to the last visit (if the treatment stop date is missing) is used for this calculation.

Descriptive statistics will be provided for patient-days of exposure. Total patient-years of exposure (PYE) will be reported for each treatment group by Treatment Periods 1 and 2 for duration of exposure.

Exposure ranges for Treatment Period 1 will be summarized as follows:

- >0 weeks, ≥4 weeks, ≥8 weeks, ≥12 weeks, ≥16 weeks, ≥24 weeks, ≥32 weeks and ≥52 weeks
- >0 to <4 weeks, ≥4 weeks to <8 weeks, ≥8 weeks to <12 weeks, ≥12 weeks to <16 weeks, ≥16 weeks to <24 weeks, ≥24 weeks to <32 weeks, ≥32 weeks to <52 weeks, and ≥52 weeks

Exposure ranges for Treatment Period 2 will be summarized as follows:

- > 52 weeks, ≥56 weeks, ≥60 weeks, ≥64 weeks, ≥68 weeks, ≥76 weeks, ≥84 weeks, and ≥104 weeks
- >52 to <56 weeks, ≥56 weeks to <60 weeks, ≥60 weeks to <64 weeks, ≥64 weeks to <68 weeks, ≥68 weeks to <76 weeks, ≥76 weeks to <84 weeks, ≥84 weeks to <104 weeks and ≥104 weeks

Overall exposure by Treatment Periods 1 and 2 will be summarized in total PYE, which is calculated according to the following formula:

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

6.14.2. Adverse Events

Adverse events are recorded in the eCRFs. Each AE will be coded to SOC and PT using the MedDRA version that is current at the time of DBL. Severity of AEs is recorded as mild, moderate, or severe. TEAEs will be analyzed by Treatment Periods 1 and 2.

A TEAE is defined as an event that first occurred or worsened in severity after the first dose of study treatment in Study JAHN 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 in Study JAHN will be used as baseline. If an event with missing severity is preexisting during the baseline period, and persists during the treatment period, then the baseline severity will be considered mild for determining treatment emergence (i.e, 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

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 in order to derive treatment-emergence. Should there be insufficient data for AE start date to make this comparison (e.g., 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 (i.e., 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.

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, or a severe TEAE will be summarized by treatment group.

The number and percentage of patients with TEAEs will be summarized by treatment group for both Treatment Periods 1 and 2 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.

6.14.2.1. Common Adverse Events

Common TEAEs will not be analyzed for Study JAHN data alone.

6.14.2.2. Serious Adverse Event Analyses

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

- death
- initial or prolonged inpatient hospitalization
- a life-threatening experience (i.e., 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. See examples in the ICH E2A guideline Section 3B

The number and percentage of patients who experienced any SAE will be summarized by treatment group for each Treatment Periods 1 and 2 using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency in the baricitinib 4-mg group within decreasing frequency in 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.

6.14.2.3. 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 by treatment group for Treatment Period 1 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 listing of all AEs leading to permanent discontinuation from the study drug or from the study will be provided for both Treatment Periods 1 and 2.

6.14.2.4. Criteria for Notable Patients

Patient narratives will be provided for all patients who experience certain "notable" events prior to data cutoff for the submission. See compound level safety standards for list of criteria.

6.14.3. 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 (eg, 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.

There is one 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 at first visit in Study JAHN) 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 nonmissing observation in the treatment period will be used as the last observation. Box plots will be provided for Treatment Period 1 and for Treatment Periods 1 and 2 combined. 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 (e.g., 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 during Treatment Period

1 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 comparison.

For laboratory analyte measurements collected qualitatively, a listing of abnormal findings will be provided for both treatment periods combined. The listing will include but will not be limited to patient ID, treatment group, laboratory collection date, analyte name, and analyte finding.

6.14.4. 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 (eg, 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, diastolic blood pressure, 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 6.14.3 will be used to analyze the vital signs and physical characteristics.

Table JAHN.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.

Table JAHN.6.8. 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 (BPM)	<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 (kg)	(Loss) decrease ≥7% from lowest value during baseline	(Gain) increase ≥7% from highest value during baseline

6.14.5. Special Safety Topics, including Adverse Events of Special Interest

Special safety topics will be analyzed in the context of combining the safety data from the originating tudies, JAHL, JAHM, and JAIY with the safety data from Study JAHN. The planned analyses are defined in the compound level safety standards.

6.15. Subgroup Analyses

Subgroup analyses with descriptive statistics will be provided for the responder and partial responder mITT population at Week 16. Subgroup analyses with comparisons for the nonresponder mITT population will be performed at Week 16. Analysis with censoring will be for the following endpoints:

- proportion of patients achieving IGA 0 or 1
- 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 using baseline of originating study where applicable:

- patient demographic and characteristics subgroups:
  - gender (male, female)
  - age group (<65, ≥65 years old)
  - age group (<65, ≥65 to <75, ≥75 to <85, ≥85 years old)
  - baseline weight (<60 kg, ≥60 to <100 kg, ≥100 kg)
  - baseline BMI (<25 kg/m^2, ≥25 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)
  - 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 Table JAHN.5.1)
  - 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)
  - cyclosporine contraindication [ineligible] (yes, no)
- baseline disease-related characteristics subgroup
  - baseline disease severity (IGA score)

Descriptive statistics will be provided for each treatment and stratum of a subgroup as outlined, regardless of sample size. If patient and even numbers allow, subgroup analyses for categorical outcomes will be performed using logistic regression using Firth's correction to accommodate (potential) sparse response rates. The model will include the categorical outcome as the dependent variable and baseline value from originating study (for EASI and itch), baseline severity from originating study, treatment, subgroup, and treatment-by-subgroup interaction as explanatory variables. Missing data will be imputed using NRI (Section 6.4.1). The treatment-by-subgroup interaction comparing treatment groups 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% 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 nonresponder 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.

6.16. 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.

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

The number and percentage of patients having IPD(s) will be summarized within category and subcategory of deviation by treatment group. The summary will be presented for the mITT population. Individual patient listings of IPDs will be provided.

6.17. Interim Analyses and Data Monitoring

A data monitoring committee (DMC) will oversee the conduct of this trial. The DMC will consist of members external to Lilly. This DMC will follow the rules defined in the DMC charter, focusing on potential and identified risks for this molecule and for this class of compounds. Data Monitoring Committee membership will include, at a minimum, specialists with expertise in dermatology, statistics, and other appropriate specialties.

The DMC will be authorized to review unblinded results of analyses by treatment group prior to DBL, including study discontinuation data, AEs including SAEs, clinical laboratory data, vital sign data, etc. 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 6.17.1.

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 DBL, to initiate the final population PK/PD model development processes or for preparation of regulatory documents. Information that may unblind the study during the analyses will not be reported to study sites or blinded study team until the study has been unblinded.

Unblinding details are given in Section 7.

In addition to the DMC analyses, 3 interim analyses will be conducted to support regulatory interactions and submission activities (at Week 16 and for the EMA, US and/or Japan submissions). Refer to the separate Blinding and Unblinding Plan for details.

6.17.1. Interim Analysis Plan

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

- patient disposition, demographics, and baseline characteristics
- exposure
- AEs, to include the following:
  - TEAEs
  - SAEs, including deaths
  - selected special safety topics
- clinical laboratory results
- vital signs
- Columbia-Suicide Severity Rating Scale

Summaries will include TEAEs, SAEs, special-topics AEs, and treatment-emergent high and low laboratory and vital signs in terms of counts and percentages where applicable. For continuous

analyses, box plots of laboratory analytes will be provided by time point and summaries will include descriptive statistics.

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.

6.18. Planned Exploratory Analyses

The planned exploratory analyses are described in Sections 6.11 and 6.12. Additional exploratory analyses may be conducted and will be documented in a supplemental SAP. Health Technology Assessment toolkit analyses, which may need to be produced, will also be documented in the supplemental SAP.

6.19. Annual Report Analyses

Annual report analyses, such as the Development Update Safety Report, will be documented in a separate document.

6.20. Clinical Trial Registry Analyses

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

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

- An AE is considered 'Serious' whether or not it is a TEAE.
- An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each SAE and 'Other' AE, for each term and treatment group, the following are provided:
  - the number of participants at risk of an event
  - the number of participants who experienced each event term
  - the number of events experienced.
- Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of patients/subjects in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- AE reporting is consistent with other document disclosures (e.g., the CSR, manuscripts, and so forth).

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

7. Unblinding Plan

Refer to a separate Blinding and Unblinding Plan document for details.

8. References

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

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

EuroQol 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.

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. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. E9(R1) Available at: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf. Accessed 20 January 2020.

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, Wyrwich K, Yosipovitch G. Psychometric properties of the Itch Numeric Rating Scale in patients with moderate-to-severe plaque psoriasis. *Br J Dermatol.* 2016;175(1):157-162.

Kunz B, Oranje A, Labrèze L, Stalder JF, Ring J, Taïeb A. Clinical Validation and Guidelines for the SCORAD Index: Consensus Report of the European Task Force Atopic Dermatitis. *Dermatology*. 1997;195(1):10-19.

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

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

Protocol I4V-MC-JAHN(a). A Phase 3 Multicenter, Double-Blind Study to Evaluate the Long-Term Safety and Efficacy of Baricitinib in Adult Patients with Atopic Dermatitis. Report on file, Eli Lilly and Company.

Protocol I4V-MC-JAHN(c). A Phase 3 Multicenter, Double-Blind Study to Evaluate the Long-Term Safety and Efficacy of Baricitinib in Adult Patients with Atopic Dermatitis. Report on file, Eli Lilly and Company.

Protocol I4V-MC-JAHN(7.1). A Phase 3 Multicenter, Double-Blind Study to Evaluate the Long-Term Safety and Efficacy of Baricitinib in Adult Patients with Atopic Dermatitis. 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 Quality 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.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRsearch.cfm?CFRPart=312. Accessed 21 July 2017.

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.

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

9. Appendices

Appendix 1. Statistical Analysis Plan Addendum

The statistical analysis plan described in this appendix is based on Treatment Period 1 of Protocol I4V-MC-JAHN(7.1) (JAHN[7.1]).

APP.1.1. Study Objectives

APP.1.1.1 Primary Objective

The **primary** objective of this study is to estimate the effect of starting baricitinib 2-mg on clinical measures and patient-reported outcomes as assessed by the proportion of patients with a response of Investigator's Global Assessment (IGA) 0, 1, or 2 at Week 16.

The associated estimand for this objective is to measure the effect of starting baricitinib 2-mg as assessed by the proportion of patients with a response of IGA 0, 1, or 2 at Week 16 assuming that treatment response disappears after patients are rescued or discontinue from study or treatment. See also Section APP.1.2.12.1 on how this estimand handles outcome after the occurrence of any intercurrent event through nonresponder imputation (NRI).

APP.1.1.2. Secondary Objectives

The secondary objectives of this study are:

Objectives	Endpoints
Weeks 0-52	
To estimate the effect of starting baricitinib 2-mg on clinical measures and patient-reported outcomes.	• Proportion of patients with a response of IGA 0, 1, or 2 assessed at Weeks 4, 24, and 52 • Proportion of patients with a response of IGA 0 or 1 assessed at Weeks 4, 16, 24, and 52 • Proportion of patients achieving response of EASI75 from baseline assessed at Weeks 4, 16, 24, and 52 • Proportion of patients with a 4-point improvement from baseline in Itch NRS at Week 16

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

APP.1.1.3. Exploratory Objectives

The exploratory objectives of this study are:

Objectives	Endpoints
Weeks 0-52	
To estimate the effect of starting baricitinib 2-mg on clinical measures and patient-reported outcomes.	• Proportion of patients with a 4-point improvement from baseline in Itch NRS at Weeks 1 and 4 • Proportion of patients achieving EASI90 at Weeks 4, 16, 24 and 52 • Percent change from baseline in EASI score at Weeks 4, 16, 24, and 52 • Proportion of patients achieving EASI50 at Weeks 4, 16, 24, and 52 • Proportion of patients achieving SCORAD75 at Weeks 4, 16, 24, and 52 • Mean change from baseline in SCORAD at Weeks 4, 16, 24, and 52 • Proportion of patients achieving SCORAD90 at Weeks 4, 16, 24, and 52 • Mean change from baseline in Skin Pain NRS at Weeks 1, 4, and 16 • Mean change from baseline in the score of the items of the ADSS at Weeks 1, 4, and 16 • Mean change from baseline in BSA affected at Weeks 4, 16, 24, and 52 • Percent change from baseline in Itch NRS at Weeks 1, 4 and 16 • Mean change from baseline in the total score of the POEM at Weeks 4, 16, 24, 52 • Mean change from baseline in the PGI-S-AD scores at Weeks 1, 4, and 16 • Mean change from baseline in the HADS at Weeks 4, 16, 24, and 52 • Mean change from baseline in the DLQI scores at Weeks 4, 16, 24, and 52 • Mean change from baseline in the WPAI-AD components (absenteeism, presenteeism, work productivity loss, and activity impairment) at Weeks 4, 16, 24, and 52 • Mean change from baseline in the EQ-5D-5L scores at Weeks 4, 16, 24, and 52

Abbreviations: ADSS = Atopic Dermatitis Sleep Scale; BSA = body surface area; DLQI = Dermatology Life Quality Index; EASI = Eczema Area and Severity Index; EQ-5D-5L = European Quality of Life-5 Dimensions-5 Levels; HADS = Hospital Anxiety Depression Scale; NRS = Numeric Rating Scale; PGI-S-AD = Patient Global Impression of Severity–Atopic Dermatitis; POEM = Patient-Oriented Eczema Measure; SCORAD = SCORing Atopic Dermatitis; WPAI-AD = Work Productivity and Activity Impairment-Atopic Dermatitis.

APP.1.2. Study Design

APP.1.2.1 Summary of Study Design

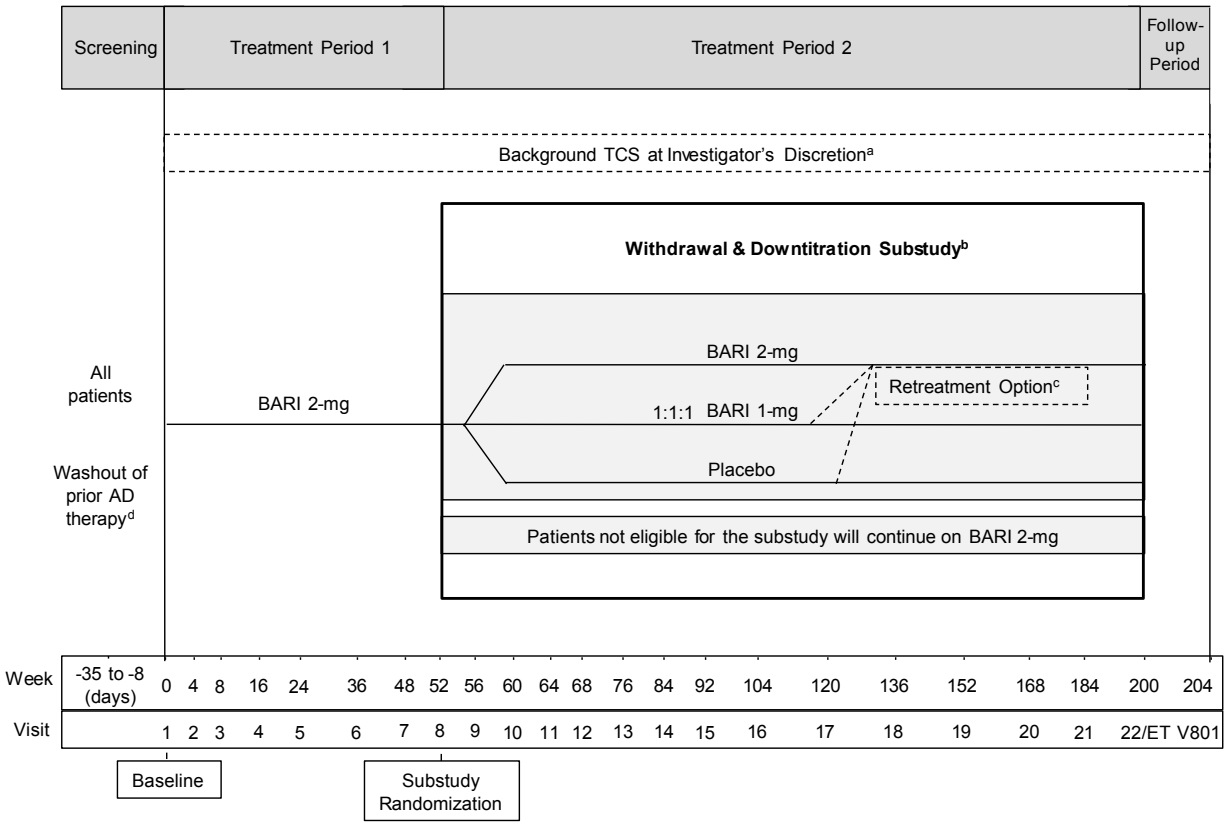
Study JAHN(7.1) is a Phase 3, multicenter, open-label study to evaluate the effects of starting baricitinib 2-mg in patients with atopic dermatitis (AD). The study will consist of 4 study periods and 1 substudy: randomized treatment withdrawal and downtitration. Unlike the main protocol, patients do not have to complete an originating study.

Figure APP.1.1. illustrates the study design. The full visit schedule is outlined in the Study Schedule of Activities in Protocol JAHN(7.1). After the screening period, all safety and efficacy assessments will be collected at identical time points as in the main Study JAHN protocol.

Patients can participate in Study JAHN(7.1) if all the criteria are met during the screening period (up to 5 weeks in duration). All patients enrolled in Study JAHN(7.1) will receive open-label baricitinib 2-mg for 52 weeks (Treatment Period 1). During Treatment Period 2 (Weeks 52 to 200), patients may participate in the randomized withdrawal and downtitration substudy if all eligibility criteria are met, otherwise patients will continue to be treated with open-label baricitinib 2-mg. Patients will use emollients daily throughout the study.

Study JAHN(7) will consist of 4 periods:

- Screening Period: Duration is between 8 and 35 days prior to Visit 1 (Week 0).
- Treatment Period 1: Treatment period from Week 0 to Week 52. Patients who meet all study eligibility criteria at Visit 1 (Week 0, baseline) will proceed to enrollment and begin the 52-week open-label treatment period.
- Treatment Period 2: Treatment period from Week 52 to 200 and will contain a randomized withdrawal and downtitration substudy. At Week 52, all patients will be evaluated for substudy eligibility and the criteria are given in the Protocol. Patients not entered in the substudy will continue on their current treatment.
- Posttreatment follow-up period: Period from Week 200 (Visit 22) or Early Termination Visit (ETV) to approximately 28 days after the last dose of investigational product.



Abbreviations: AD = atopic dermatitis; Bari = baricitinib; ET = early termination; IGA = Investigator's Global Assessment; TCS = topical corticosteroids; V = visit; W = week.

a Background TCS may be initiated or reinitiated at any time during the study, however, we recommend waiting at least 1 week after initial treatment assignment since everyone will be on active, open-label treatment. For those eligible for the substudy, TCS will also be provided as part of retreatment any time a patient's IGA score becomes ≥3. Please refer to the main protocol.

b Eligible patients will be randomized in the withdrawal and downtitration substudy as described in the main protocol. Patients who do not enroll in the substudy will remain on their treatment as described in the main protocol.

c Patients enrolled in the substudy will automatically be retreated if their IGA score becomes ≥3 as described in the main protocol.

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

Figure APP.1.1. Illustration of study design for Clinical Protocol I4V-MC-JAHN.

APP.1.2.2. Method of Assignment to Treatment

At entry into Study JAHN(7.1), patients who meet all criteria for enrollment will be assigned treatment at Visit 1 (Week 0). The interactive web response system will assign open-label bottles starting at Visit 1 (Week 0), and at each visit up to and including Visit 7 (Week 48) for all patients.

At Week 52, patients eligible for the withdrawal and downtitration substudy will be assigned to treatment as described in the main protocol.

Patients who do not qualify for the substudy will continue to be assigned open-label bottles from Visit 8 (Week 52) until Visit 15 (Week 92).

This study will be conducted internationally at multiple sites. Table APP.1.1 describes how regions were defined.

Table APP.1.1. Geographic Regions

Region	Country
Europe	Austria, Czech Republic, Denmark, Germany, Hungary, Italy, Spain, Switzerland
Rest of World	Argentina, Australia, Israel, Mexico, Russia

APP.1.2.3. A Priori Statistical Methods

APP.1.2.3.1. Determination of Sample Size

Approximately 250 patients may be enrolled in Study JAHN(7.1). Given that this addendum was planned, these patients are included in the anticipated number of participants enrolling into the randomized withdrawal and downtitration substudy as discussed in Section 6.1.

This study is intended to estimate the effect of starting baricitinib 2-mg and thus, the analyses will be descriptive.

Patients at Week 52 will be stratified by responder status (IGA 0 or 1 versus IGA 2) when entering the randomized withdrawal and downtitration substudy. It is estimated that there will be approximately 60 patients entering into the randomized withdrawal and downtitration substudy from JAHN(7.1). The substudy is meant to evaluate the change in clinical response after treatment withdrawal or downtitration and does not account for whether the sample size is sufficient to detect a difference between baricitinib and placebo. Maintenance of treatment benefit is defined as response of IGA 0, 1, or 2.

APP.1.2.3.2. General Considerations

Refer to Section 6.2.

APP.1.2.3.2.1. Analysis Populations

Intent-to-treat (ITT) population: The ITT population analysis set is defined as all enrolled patients in Study JAHN(7.1).

Unless otherwise specified, the efficacy and health outcome analyses will be conducted on the ITT population (Gillings and Koch 1991).

Safety population: The Safety population for this addendum is defined the same as the safety population in Section 6.2.1.

Safety analyses for this addendum will be performed on the safety population and data will be pooled with the data from Study JAHN.

For the database lock that supports the US submission, the population for this addendum, Week 24 efficacy and health outcome analyses will be conducted on the **Week 24 efficacy evaluable set**. This subset of the ITT Population is anchored on the database cut-off date for the US DBL (13 December 2019). Specifically, a patient will be included in the Week *24* Efficacy Evaluable set if their Week *24* visit has occurred or their expected Week *24* visit date plus a 15-day buffer is on or prior to the database cutoff date. The expected Week *24* visit date will be calculated as follows: Date of first dose date + (*24* weeks * 7 days) + 15 days.

APP.1.2.3.2.2. Definition of Baseline and Postbaseline Measures

The baseline value for the efficacy and health outcome analyses is defined similarly to Section 6.2.2 except it uses the date of first study drug administration (expected at Week 0, Visit 1) in place of the first study drug administration in the originating study. The baseline value for the daily diary assessments (Itch Numeric Rating Scale (NRS), Atopic Dermatitis Sleep Scale, Skin Pain NRS, Patient Global Impression of Severity-AD) is also defined similarly (using the date of first study administration in Study JAHN).

For the purpose of Study JAHN addendum, the safety analyses will be conducted as described in Section 6.2.2.

Postbaseline measurements are described in Section 6.2.2.

APP.1.2.3.2.3. Analysis Methods

All discrete efficacy and health outcome variables will be summarized using frequencies and percentages. The 95% confidence intervals will also be provided.

APP.1.2.3.2.4. Derived Data

Refer to Section 6.2.4, using the Study JAHN(7.1) baseline where applicable.

APP.1.2.4. Adjustments for Covariates

Not applicable.

APP.1.2.5. Handling of Dropouts or Missing Data

Refer to Section 6.4.

The censoring rule will be applied to all efficacy and health outcome endpoints conducted for the ITT population.

Sections 6.4.1 and 6.4.2 summarize the imputation methods for the various efficacy and health outcome endpoints. Table APP.1.2 summarizes the various imputation techniques being used for the efficacy and health outcomes analyses.

Table APP.1.2. Imputation Techniques for Various Variables

Efficacy and Health Outcome Endpoints[a]	Imputation Method[b]
IGA(0,1), IGA(0,1,2), EASI75, 4-point Itch NRS improvement	NRI
All remaining categorical endpoints	NRI

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

[a] Refer to Table JAHN.6.4 and Table JAHN.6.6.

[b] Analysis utilizing the censoring rule.

APP.1.2.6. Multicenter Studies

This study will be conducted by multiple investigators at multiple sites internationally. The countries will be categorized into geographic regions, as described in Section APP.1.2.2.

APP.1.2.7. Multiple Comparisons/Multiplicity

As this study is designed to assess the effect of starting baricitinib 2-mg in patients with AD, no adjustments for multiple comparisons will be utilized in the statistical analyses for this study.

APP.1.2.8. Patient Disposition

Patient disposition will be described as discussed in Section 6.7 for the Study JAHN(7.1) ITT population only. A listing of the ITT population with their treatment assignments will not be produced as all patients are on open-label baricitinib 2-mg.

APP.1.2.9. Patient Characteristics

Patient characteristics including demographics, baseline characteristics, and pre-existing conditions will be summarized descriptively.

APP.1.2.9.1. Demographics

Patient demographics will be summarized as described above. Refer to Section 6.8.1 for the list of demographics. A listing of patient demographics will also be provided for the ITT population.

APP.1.2.9.2. Baseline Disease Characteristics

Refer to Section 6.8.2 for the baseline AD clinical characteristics, baseline health outcome measures, and other baseline demographic and disease characteristics which will be presented using the Study JAHN(7.1) baseline where applicable. In addition, the following will be summarized:

- 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)
- Prior use of Cyclosporine (yes, no)
- Reasons Cyclosporine was not used
- Prior use of topical calcineurin inhibitor (TCNI) (yes, no)
- Reasons TCNI was not used

APP.1.2.9.3. 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 or similar pre-defined lists of preferred terms of interest.

Pre-existing conditions are defined in Section 6.8.3 and will be summarized using the ITT population.

APP.1.2.10. Treatment Compliance

Patient compliance with study medication will be assessed from Week 0 (Visit 2) to Week 104 (Visit 16) or ETV using the ITT population.

All patients are expected to take 1 tablet daily from a bottle for Treatment Period 1 as described in APP.1.2.2. Each bottle contains 36 tablets.

Refer to Section 6.9 for further details on compliance. Analyses will be presented for the ITT population.

APP.1.2.10.1. Background Therapy

Analyses will be presented as described in Section 6.9.1 using the ITT population.

APP.1.2.11. Previous and Concomitant Therapy

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

At screening, previous and current AD treatments are recorded for each patient. A summary of previous medications used for AD, including zoster immunization and tuberculosis vaccine and medications that are discontinued after screening and before the first dose of study drug, will be prepared using frequency counts and percentages by preferred medication name, with preferred medication names sorted by frequency. Concomitant therapy will be recorded at each visit and will be classified similarly. An additional summary for previous medications used for AD will be created containing the reason of discontinuation.

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 as concomitant for the treatment period.

Summaries of previous medications will be provided for the following categories:

- Previous AD therapies
- Previous AD therapies including reason for discontinuation

A summary of concomitant therapy including concomitant therapies of special interest will be provided for the ITT population.

APP.1.2.12. Efficacy Analyses

The general methods used to summarize efficacy data, including the definition of baseline value for assessments are described in Section APP.1.2.3.2.

Efficacy analyses will generally be analyzed according to the following formats:

- Week 0 to Week 52

Table JAHN.6.4 includes the descriptions and derivations of the primary, secondary, and exploratory efficacy outcomes. Table APP.1.3 provides the detailed analyses including analysis type, method and imputation, population, and time point for efficacy analyses.

Table APP.1.3. Description of Primary, Secondary and Exploratory Efficacy Analyses

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Time Point	Analysis Type
Validated Investigator's Global Assessment for AD (IGA)	▪ Proportion of patients with a response of IGA [0,1,2]	Descriptive statistics (Observed and NRI)	ITT	Weeks 0 to 16	Secondary analysis
			ITT Week 24 EE set	Week 24	Exploratory analysis
	▪ Proportion of patients with a response of IGA [0,1]	Descriptive statistics (Observed and NRI)	ITT	Weeks 0 to 16	Primary analysis
			ITT Week 24 EE set	Week 24	Exploratory analysis
Eczema Area and Severity Index (EASI)	▪ EASI score ▪ Change from baseline in EASI score ▪ Percent change from baseline in EASI score	Descriptive statistics (Observed)	ITT	Weeks 0 to 24	Exploratory analysis
	▪ Proportion of patients achieving EASI75	Descriptive statistics (Observed and NRI)	ITT	Weeks 0 to 16	Secondary analysis
			ITT Week 24 EE set	Week 24	Exploratory analysis
	▪ Proportion of patients achieving EASI90 ▪ Proportion of patients achieving EASI50	Descriptive statistics (Observed and NRI)	ITT	Weeks 0 to 16	Exploratory analysis
			ITT Week 24 EE set	Week 24	Exploratory analysis
Body Surface Area (BSA) Affected by AD	▪ BSA score ▪ Change from baseline in BSA score	Descriptive statistics (Observed)	ITT	Weeks 0 to 24	Exploratory analysis
SCORing Atopic Dermatitis (SCORAD)	▪ SCORAD score ▪ Change from baseline in SCORAD score ▪ Percent change from baseline in SCORAD score	Descriptive statistics (Observed)	ITT	Weeks 0 to 24	Exploratory analysis
	▪ Proportion of patients achieving SCORAD75 ▪ Proportion of patients achieving SCORAD90	Descriptive statistics (Observed and NRI)	ITT	Weeks 0 to 16	Exploratory analysis
			ITT Week 24 EE set	Week 24	Exploratory analysis

Description of Primary, Secondary and Exploratory Efficacy Analyses

Abbreviations: AD = atopic dermatitis; EE = efficacy evaluable; ITT = intent-to-treat; MMRM = mixed-model repeated measures; NRI = nonresponder imputation.

APP.1.2.12.1. Primary Outcome and Methodology

The validated IGA for AD uses the clinical characteristics of erythema, papulation/induration, oozing/crusting and lichenification to produce a single-item score ranging from 0 to 4. The primary analysis of the study is to estimate the effect of starting baricitinib 2-mg on clinical measures and patient-reported outcomes as assessed by the proportion of patients with a response of IGA 0, 1, or 2 assessed at Week 16 using the ITT population assuming that treatment response disappears after the patient is rescued or discontinued study or study treatment. This will serve as the primary estimand. In this estimand, missing data due to the application of the censoring rule and the occurrence of other noncensor intercurrent events will be imputed using the NRI method described in Section 6.4.1.

APP.1.2.12.2. Secondary and Exploratory Efficacy Analyses

There will be no adjustment for multiple comparisons for any analyses. The secondary and exploratory efficacy analyses are detailed in Section 6.2.3 and Table JAHN.6.4. Health outcomes analyses are described in Section APP.1.2.13.

Note that the exploratory efficacy analyses on a continuous outcome that is measured over time uses a slightly different estimand which is to estimate the effect of starting baricitinib 2-mg on a specific endpoint evaluated at specified time points using the ITT population assuming all patients remained in their treatment throughout the defined period of the study.

APP.1.2.13. Health Outcomes/Quality-of-Life Analyses

The general methods used to summarize health outcomes and quality-of-life measures, including the definition of baseline value for assessments are described in Section APP.1.2.3.2.2.

Health outcomes and quality-of-life measures will generally be analyzed according to the formats discussed in Section APP.1.2.12.

Table JAHN.6.6 includes the descriptions and derivations of the health outcomes and quality-of-life measures. Table APP.1.4 provides the detailed analyses including analysis type, method and imputation, population, and time point for health outcomes and quality-of-life measures.

Table APP.1.4. Description of Health Outcomes and Quality-of-Life Measures Analyses

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Time Point	Analysis Type
Itch Numeric Rating Scale (NRS)	▪ Itch NRS score ▪ Change from baseline in Itch NRS score ▪ Percent change from baseline Itch score	Descriptive statistics	ITT	Weeks 0 to 16	Exploratory analysis
	▪ Proportion of patients with a 4-point improvement from baseline of originating study in Itch NRS in subgroup of patients who had baseline Itch NRS ≥4	Descriptive statistics (Observed and NRI)	ITT	Weeks 0 to 16	Secondary analysis
Skin Pain Numeric Rating Scale (NRS)	▪ Skin Pain NRS score ▪ Change from baseline in Skin Pain NRS score	Descriptive statistics (Observed)	ITT	Weeks 0 to 16	Exploratory analysis
	▪ Proportion of patients with a 4-point improvement from baseline of originating study in Skin Pain NRS in subgroup of patients who had baseline Skin Pain NRS ≥4	Descriptive statistics (Observed and NRI)	ITT	Weeks 0 to 16	Exploratory analysis
Atopic Dermatitis Sleep Scale (ADSS)	▪ ADSS item scores ▪ Change from baseline in ADSS item scores	Descriptive statistics (Observed)	ITT	Weeks 0 to 16	Exploratory analysis
	▪ Proportion of patients with a 1.5 point improvement from baseline of originating study in ADSS2 score in subgroup of patients who had baseline ADSS2 score ≥ 1.5 ▪ Proportion of patients with a 2 point improvement from baseline of originating study in ADSS2 score in subgroup of patients who had baseline ADSS2 score ≥ 2	Descriptive statistics (Observed and NRI)	ITT	Weeks 0 to 16	Exploratory analysis

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Time Point	Analysis Type
Patient Global Impression of Severity–Atopic Dermatitis (PGI-S-AD)	▪ PGI-S-AD score ▪ Change from baseline in PGI-S-AD score	Descriptive statistics (Observed)	ITT	Weeks 0 to 16	Exploratory analysis
Patient-Oriented Eczema Measure (POEM)	▪ POEM score ▪ Change from baseline in POEM score	Descriptive statistics (Observed)	ITT	Weeks 0 to 24	Exploratory analysis
	▪ Proportion of patients with a 4 point improvement from baseline of originating study in POEM score in subgroup of patients who had baseline POEM score ≥4	Descriptive statistics (Observed and NRI)	ITT	Weeks 0 to 16	Exploratory analysis
			ITT Week 24 EE set	Week 24	Exploratory analysis
Hospital Anxiety Depression Scale (HADS)	▪ HADS domain scores and total scores ▪ Change from baseline in HADS domain and total scores	Descriptive statistics (Observed)	ITT	Weeks 0 to 24	Exploratory analysis
Dermatology Life Quality Index (DLQI)	▪ DLQI total score ▪ Change from baseline in DLQI ▪ Observed and change from baseline in domain scores -Symptoms and feelings -Daily activities -Leisure -Work and school -Personal relationships -Treatment	Descriptive statistics (Observed)	ITT	Weeks 0 to 24	Exploratory analysis
	▪ Proportion of patients with a DLQI (0, 1) response ▪ Proportion of patients with a 4 point improvement from baseline of originating study in DLQI in subgroup of patients who had baseline DLQI ≥4	Descriptive statistics (Observed and NRI)	ITT	Weeks 0 to 16	Exploratory analysis

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Time Point	Analysis Type
			ITT Week 24 EE set	Weeks 24	Exploratory analysis
Work Productivity and Activity Impairment: Atopic Dermatitis (WPAI-AD)	Observed and Change from baseline in employment status	Descriptive statistics (observed)	ITT	Weeks 0 to 24	Exploratory analysis
	Observed and Change from baseline in: ▪ absenteeism ▪ presenteeism ▪ overall work impairment ▪ impairment in activities	Descriptive statistics (Observed)	ITT	Weeks 0 to 24	Exploratory 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	Descriptive statistics (Observed)	ITT	Weeks 0 to 24	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	Descriptive statistics (Observed)	ITT	Weeks 0 to 24	Exploratory analysis

Abbreviations: AD = atopic dermatitis; EE = efficacy evaluable; ITT = intent-to-treat; NRI = nonresponder imputation.

APP.1.2.14. Bioanalytical and Pharmacokinetic / Pharmacodynamic Methods

Pharmacokinetic, pharmacodynamic, and biomarker analyses will not be conducted.

APP.1.2.15. Safety Analyses

The safety data from this addendum will be pooled with the bari 2-mg treatment group from Study JAHN and will follow the same planned analyses given in Section 6.14. Refer to Section 6.14 for the planned analyses.

APP.1.2.15.1. Extent of Exposure

Refer to Section 6.14.1 for the planned analyses.

APP.1.2.15.2. Adverse Events

Refer to Section 6.14.2 for the planned analyses.

APP.1.2.15.2.1. Common Adverse Events

Not applicable.

APP.1.2.15.2.2. Serious Adverse Event Analyses

Refer to Section 6.14.2.2 for the planned analyses.

APP.1.2.15.2.3. Other Significant Adverse Events

Refer to Section 6.14.2.3 for the planned analyses.

APP.1.2.15.3. Clinical Laboratory Evaluation

Refer to Section 6.14.3 for the planned analyses.

APP.1.2.15.4. Vital Signs and Other Physical Findings

Refer to Section 6.14.3 for the planned analyses.

APP.1.2.15.5. Special Safety Topics, including Adverse Events of Special Interest

Refer to Section 6.14.5 for the planned analyses.

APP.1.2.16. Subgroup Analyses

Subgroup analyses will not be performed for Study JAHN(7.1).

APP.1.2.17. Protocol Deviations

Protocol deviations will be analyzed as described in Section 6.16 using the ITT population.

APP.1.2.18. Interim Analyses and Data Monitoring

Refer to Section 6.17 for the planned analyses.

APP.1.2.19. Planned Exploratory Analyses

The planned exploratory analyses are described in Section 6.18.

APP.1.2.20. Annual Report Analyses

Refer to Section 6.19.

APP.1.2.21. Clinical Trial Registry Analyses

Refer to Section 6.20.

APP.1.3. Unblinding Plan

Refer to Section 7.

APP.1.4. References

Gillings D, Koch G. The Application of the Principle of Intention–to–Treat to the Analysis of Clinical Trials. *Drug Inf J*. 1991;25:411-424.

Program Safety Analysis Plan: Baricitinib (LY3009104) Version 65. Report on file, Eli Lilly and Company.

Protocol I4V-MC-JAHN(7.1). A Phase 3 Multicenter, Double-Blind Study to Evaluate the Long-Term Safety and Efficacy of Baricitinib in Adult Patients with Atopic Dermatitis. Report on file, Eli Lilly and Company.

Appendix 2. Statistical Analysis for 4-Month Safety Update

This appendix describes the efficacy analyses to be performed for the 4-month safety update (20 May 2020).

APP.2.1. *4-Month Safety Update Lock*

For the DBL that supports the 4-month safety update, the population originating from combination therapy Study JAIY has a definition for Week 36 efficacy evaluable set, similar to Week 24 efficacy evaluable set defined in Section 6.2.1. Week 36 efficacy and health outcome analyses will be conducted on the **Week 36 efficacy evaluable set**. The following efficacy summaries/analyses will be provided:

- Proportion of patients achieving IGA (0,1) and IGA (0,1,2) assessed at Weeks 0 to 24 for nonresponders (NR). Analysis to be repeated for responders and partial responders (RPR).
- Proportion of patients achieving IGA (0,1) and IGA (0,1,2) assessed at Week 36 efficacy evaluable set for NR. Analysis to be repeated for RPR.
- Proportion of patients achieving EASI50, EASI75, and EASI90 response assessed at Weeks 0 to 24 for NR. Analysis to be repeated for RPR.
- EASI change from baseline and percent change from baseline mixed-model repeated measures assessed at Weeks 0 to 24 for NR.
- Proportion of patients achieving EASI50, EASI75, and EASI90 response assessed at Week 36 efficacy evaluable set for NR. Analysis to be repeated for PRP.
- EASI change and percent change from baseline assessed at Week 36 efficacy evaluable set for NR. Analyses to be repeated for RPR.
- Summary of sponsor provided topical corticosteroids (TCS) background therapy weight (g) by potency for weeks 0 to 52.

For the 4-month safety update database lock, the open label addendum has a definition for **Week 36 and 48 efficacy evaluable sets**, similar to Week 24 efficacy evaluable set defined in App.1.2.3.2.1. Week 36 and 48 efficacy and health outcome analyses will be conducted on the **Week 36 and 48 efficacy evaluable se**ts. The following efficacy summaries/analyses will be provided:

- Proportion of patients achieving IGA (0,1) and IGA (0,1,2) assessed at Weeks 0 to 24.
- Proportion of patients achieving IGA (0,1) and IGA (0,1,2) assessed at Week 36 and Week 48 efficacy evaluable sets, respectively.
- Proportion of patients achieving EASI50, EASI75, and EASI90 response assessed at Weeks 0 to 24.
- Proportion of patients achieving EASI50, EASI75, and EASI90 response rate assessed at Week 36 and Week 48 efficacy evaluable sets, respectively.

- EASI change and percent change from baseline assessed at Weeks 0 to 48 based on observed data.
- Summary of sponsor provided TCS background therapy weight (g) by potency for weeks 0 to 52.

For the substudy the following analyses/summaries will be provided:

- Proportion of patients achieving IGA (0,1) and IGA (0,1,2) assessed at Weeks 52 to 68.
- Proportion of patients achieving EASI50, EASI75, and EASI90 assessed at Weeks 52 to 68.
- EASI change from baseline and percent change from baseline (mLOCF) assessed at Weeks 52 to 68; Week 0 used as baseline.
- EASI change from baseline and percent change from baseline (mLOCF) assessed at Weeks 52 to 68; Week 52 used as baseline.
- Proportions of patients achieving a ≥4-point improvement assessed at Weeks 52 to 68.
- Itch NRS change from baseline and percent change from baseline (mLOCF) assessed at Weeks 52 to 68; Week 0 used as baseline.
- Itch NRS change from baseline and percent change from baseline (mLOCF) assessed at Weeks 52 to 68; Week 52 used as baseline.

Leo Document ID = 5f2a73e3-402c-4211-9b9a-88496b13b33b

Approver: PPD

Approval Date & Time: 19-May-2020 13:08:32 GMT

Signature meaning: Approved