I4V-MC-JAIR Statistical Analysis Plan v3

A Multicenter, Randomized, Double-Blind, Placebo Controlled, Phase 3 Study to Evaluate the Efficacy and Safety of Baricitinib in Adult Patients with Severe or Very Severe Alopecia Areata

NCT03899259

Approval Date: 22-Feb-2021

1. Statistical Analysis Plan: I4V-MC-JAIR: A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Phase 3 Study to Evaluate the Efficacy and Safety of Baricitinib in Adult Patients with Severe or Very Severe Alopecia Areata

BRAVE-AA2

Confidential Information

Baricitinib (LY3009104) Alopecia Areata

Study I4V-MC-JAIR (JAIR) is a Phase 3, multicenter, randomized, double-blind, placebo-controlled, parallel-group, outpatient study to evaluate the efficacy and safety of baricitinib 4-mg and 2-mg in adult patients with severe or very severe scalp Alopecia Areata.

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

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

Approval Date: 22-Feb-2021 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

3. Revision History

Statistical analysis plan (SAP) Version 1 was based on Protocol I4V-MC-JAIR(b) (JAIR) and was approved prior to the production transfer for the first data monitoring committee (DMC). Statistical analysis plan Version 2 was based on Protocol I4V-MC-JAIR(d) and was approved prior to the Week 36 interim data base lock and includes the following changes:

- Modified objectives in Section 4 to align with protocol I4V-MC-JAIR(d). The exploratory objectives were also updated to address efficacy assessments beyond Week 104.
- In Section 5.1.1, a 96-week bridging extension was added and Figure JAIR.5.1 was updated to align with protocol I4V-MC-JAIR(d). Exceptions to the posttreatment follow-up period were also clarified.
- Updated Section 5.2 to clarify stratification by duration of current episode at baseline.
- Modified Section 6.2.1 to remove the Follow-up Population. Added language to clarify analyses performed at the Week 36 interim data base lock and removed language regarding all baricitinib exposure analyses after the final database lock as this will be done at the integrated level.
- Clarified definition of baseline in Section 6.2.2 and referred to protocol I4V-MC-JAIR(d) for definition of visit windows.
- Added language in Section 6.2.3 to address the definition of remotely collected data and clarified that Kaplan-Meier curves will be produced for time-to-event analysis. Language was also added regarding presentation of relative risk for the primary analysis.
- Modified Section 6.2.4 to include further details on age, weight, and BMI groups, onset age, and duration of AA at baseline.
- Removed MMRM as an analysis method.
- Added language in Section 6.4 to explain the application of censoring rules to remotely collected data.
- Added Hybrid Imputation Section 6.4.3 to address the handling of missing data and missing data due to the coronavirus disease 2019 (COVID-19) pandemic. Removed placebo multiple imputation as an imputation method and updated Table JAIR.6.4.
- Clarified in Section 6.6 that multiplicity adjustments will be applied to the FAS population and updated the graphical testing figure and explanation of graphical testing procedure.
- Added language in Section 6.7 to specify that treatment disposition will be summarized using the FAS population and removed language specific to the Randomized Downtitration Population as these details will be supplied in a later version of the SAP.
- Section 6.8.1 was updated to include additional age, weight, and BMI group categorizations as well as a not reported category for ethnicity.
- Updated Section 6.8.1 to include an additional category for current episode of AA, removed a duplicate row for SALT category, and added more details for prior therapy.
- The definition of preexisting condition was updated in Section 6.8.3.

- Clarified the definition of treatment compliance assessed by treatment period in Section 6.9.
- Added language in Section 6.10 to specify that previous and concomitant therapy will be summarized by treatment period.
- Updated Table JAIR.6.5 in Section 6.11 to align with the updated objectives in protocol I4V-MC-JAIR(d). In addition, the exploratory analyses for the proportion of patients achieving a PRO of zero and the proportion of patients achieving a ClinRO of zero at weeks 24 and 36 were removed and are to be included in the supplemental SAP. Table JAIR.6.6 was updated to clarify the supplementary analyses and additional sensitivity analyses and dosing evaluation analyses.
- Section 6.11.1 was updated to remove a duplicate definition of the primary endpoint and language regarding a supplemental estimand. Language was added to address the impact of the COVID-19 pandemic. The title of Section 6.11.2 was updated to Secondary and Exploratory Outcome Analyses.
- Sections 6.11.3, 6.11.4, and 6.11.5 were added to include details on supplementary analyses, dosing evaluation analyses, and analyses beyond the Week 36 Placebo-controlled period.
- In Section 6.12, language for the SF-36 description was clarified and the incorrect reference to stage 2 was removed from the Skindex-16 description. The HADS description was also updated so that the anxiety domain is presented separately from the depression domain. Details were also added for the US and UK versions of EQ-5D-5L.
- In addition, for Section 6.12, language regarding SF-36 components was added to the table describing health outcomes analyses and all time points were updated to those analyzed at the time of the Week 36 primary data base lock. The exploratory analyses for the HADS and SF-36 components were updated to logistic regression. The exploratory analyses for EQ-5D-5L were updated to ANCOVA.
- In Section 6.13, clarifying language was added to the definitions of the analysis periods.
- Section 6.13.1 was updated to include duration of exposure in weeks instead of days. The duration of exposure calculation was clarified as excluding exposure post treatment change. Language regarding exposure in patient years was also updated.
- The analysis period for TEAEs was clarified in Section 6.13.2 and language was added to summarize TEAEs by maximum severity by treatment.
- In Section 6.13.3, duplicate information regarding analysis periods was removed and a reference to Section 6.13 was added in Sections 6.13.3 and 6.13.4 for the detailed analysis period definition.
- Section 6.13.5.6 was updated to remove association between infection and neutropenia/lymphopenia.
- The subgroup analyses in Section 6.14 were edited to match the updated demographics and baseline characteristics categories. The subgroup analysis for previous treatment was removed. Language was added to clarify the covariates and censoring rule for subgroup analyses.
- Section 6.15 was added to describe analysis for the Japan submission.
- Section 6.17.1 was updated to include more details on DMC analyses.

- Sections 6.17.2 and 6.17.3 were added to provide further information on interim analyses and adjudication of MACE.
- Section 6.20 was updated to address the requirements for the European Clinical Trials Database.

Statistical analysis plan Version 3 was based on Protocol I4V-MC-JAIR(d) and was approved post Week 36 interim data base lock but prior to unblinding the lead statistician and includes the following change:

- Modified the graphical testing of Figure JAIR.3.1 and Figure JAIR.3.2 in Section 6.6 by correcting the testing time point of $SALT_{50}$ from Week 16 to Week 12, and inserting the missing key secondary endpoint, SLAT≤10 at Week 36, respectively.

4. Study Objectives

4.1. Primary Objective

The primary objective of this study is to test the hypothesis that baricitinib 4-mg once daily (QD) or baricitinib 2-mg QD is superior to placebo in the treatment of patients with severe or very severe alopecia areata (AA), as assessed by the proportion of patients achieving Severity of Alopecia Tool (SALT) ≤20 at Week 36.

In particular, the associated estimand for this objective is to measure the effect of baricitinib 4-mg or baricitinib 2-mg vs placebo on patients with severe or very severe AA as assessed by the proportion of patients achieving SALT ≤20 at Week 36, assuming that treatment response disappears at the visits conducted remotely as a consequence of the COVID-19 pandemic or after patients discontinue from study or treatment. See also Section 6.4.1 and Section 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 are listed in Table JAIR.4.1.

Table JAIR.4.1. Secondary Objectives

Key Secondary (double-blind, placebo-controlled treatment period) *These are prespecified objectives that will be adjusted for multiplicity*	
Objectives	**Endpoints**
To compare the efficacy of baricitinib 4-mg dose or 2-mg dose to placebo in AA during the double-blind, placebo-controlled treatment period as measured by **physician-assessed** signs and symptoms of AA	• Proportion of patients achieving SALT ≤20 Weeks 16 and 24 • Percent change from baseline in SALT score at Week 36 • Proportion of patients achieving a $SALT_{50}$ at Week 12 • Proportion of patients achieving $SALT_{90}$ at Week 36 • Proportion of patients achieving an absolute SALT ≤10 at Weeks 24 and 36 • Proportion of patients achieving ClinRO Measure for EB Hair Loss 0 or 1 with ≥2-point improvement from baseline at Week 36 (among patients with ClinRO Measure for EB Hair Loss ≥2 at baseline). • Proportion of patients achieving ClinRO Measure for EL Hair Loss 0 or 1 with ≥2-point improvement from baseline at Week 36 (among patients with ClinRO Measure for EL Hair Loss ≥2 at baseline).

To compare the efficacy of baricitinib 4-mg dose or 2-mg dose to placebo in AA during the double-blind, placebo-controlled treatment period as assessed by a **PRO measure**	• Proportion of patients with PRO for Scalp Hair Assessment score of 0 or 1 with ≥2-point improvement from baseline at Week 36 among patients with a score of ≥3 at baseline
Other Secondary (double-blind, placebo-controlled treatment period) *These are prespecified objectives that will NOT be adjusted for multiplicity*	
To compare the efficacy of baricitinib 4-mg dose or 2-mg dose to placebo in AA during the double-blind, placebo-controlled treatment period as measured by **physician-assessed** signs and symptoms of AA	• Proportion of patients achieving $SALT_{50}$ at Weeks 16, 24, and 36 • Proportion of patients achieving $SALT_{75}$ at Weeks 24 and 36 • Proportion of patients achieving a $SALT_{90}$ at Week 24 • Change from baseline in SALT score at Weeks 12, 16, 24, and 36 • Percent change from baseline in SALT score at Weeks 12, 16, and 24 • Time to achieve SALT ≤20 • Proportion of patients achieving $SALT_{100}$ at Weeks 24 and 36 • Proportion of patients achieving ClinRO Measure for EB Hair Loss 0 or 1 with ≥2-point improvement from baseline at Weeks 16 and 24 (among patients with ClinRO Measure for EB Hair Loss ≥2 at baseline) • Proportion of patients achieving ClinRO Measure for EL Hair Loss 0 or 1 with ≥2-point improvement from baseline at Weeks 16 and 24 (among patients with ClinRO Measure for EL Hair Loss ≥2 at baseline)
To compare the efficacy of baricitinib 4-mg dose or 2-mg dose to placebo in AA during the double-blind, placebo-controlled treatment period as assessed by **PRO measures** and quality of life tools	• Proportion of patients with PRO for Scalp Hair Assessment score of 0 or 1 with a ≥2-point improvement from baseline at Weeks 12 and 24 among patients with a score of ≥3 at baseline • Proportion of patients achieving PRO Measure for EB 0 or 1 with ≥2-point improvement from baseline at Weeks 16, 24, and 36 (among patients with PRO Measure for EB ≥2 at baseline) • Proportion of patients achieving PRO Measure for EL 0 or 1 with ≥2-point improvement from baseline at Weeks 16, 24, and 36 (among patients with PRO Measure for EL ≥2 at baseline) • Mean change from baseline at Weeks 24 and 36 in Skindex-16 AA domain scores (Symptoms, Emotions, Functioning) • Mean change from baseline in HADS-A and HADS-D total scores at Weeks 24 and 36

Other Secondary (patients entering randomized downtitration) *These are prespecified objectives that will NOT be adjusted for multiplicity*	
To compare the maintenance of efficacy for patients randomized to remain on baricitinib 4 mg, compared with patients randomized to baricitinib 2 mg at Week 52 of the long-term extension period, as measured by **physician-assessed** signs of AA	• Proportion of patients maintaining SALT ≤20 at Weeks 64, 76, 88, 104, 120, 136, 152, 168, 184, and 200 • Proportion of patients experiencing a loss of treatment benefit (>20-point absolute worsening in SALT score) at Weeks 64, 76, 88, 104, 120, 136, 152, 168, 184, and 200 • Time to loss of treatment benefit (>20-point absolute worsening in SALT score)
For patients experiencing loss of treatment benefit after randomization to baricitinib 2 mg at Week 52: • To evaluate the recapture of efficacy for patients who were retreated after experiencing a loss of treatment benefit during the long-term maintenance period as measured by **physician-assessed** signs of AA • To evaluate the recapture of efficacy for patients who were retreated after experiencing a loss of treatment benefit during the long-term maintenance period as assessed by **PRO** and quality of life tools	• Proportion of patients that achieve a SALT score ≤20 at 12, 16, 24, and 36 weeks of retreatment with baricitinib 4-mg • Percent change in SALT score at 12, 16, 24, and 36 weeks of retreatment with baricitinib 4-mg • Proportion of patients with a PRO for Scalp Hair Assessment score of 0 or 1 at 12, 16, 24, and 36 weeks of retreatment with baricitinib 4-mg

Abbreviations: AA = alopecia areata; ClinRO = clinician reported outcome; EB = eyebrow; EL = eyelash; HADS = Hospital Anxiety and Depression Scale; PRO = patient-reported outcome; SALT = Severity of Alopecia Tool; $SALT_{50/75/90/100}$ = at least 50%/75%/90%/100% improvement from baseline in SALT score; Skindex-16 AA = Skindex-16 Adapted for Alopecia Areata.

4.3. Exploratory Objectives

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: SALT; at least 30% improvement from baseline in SALT score ($SALT_{30}$); clinician-reported outcomes (ClinROs) for nail appearance, eyebrows, and/or eyelash hair loss; PROs for Scalp Hair Assessment, eyebrows and eyelashes, nail appearance, and eye irritation; Skindex-16 adapted for alopecia areata (Skindex-16 AA); Short-Form 36-Item Health Survey acute version 2 (SF-36); European Quality of Life – 5 Dimensions – 5 Level (EQ-5D-5L); Hospital Anxiety and Depression Scale (HADS). Assessments of efficacy may be performed beyond Week 104 up to Week 200.

5. Study Design

5.1. Summary of Study Design

Study JAIR is a Phase 3, multicenter, randomized, double-blind, placebo-controlled, parallel-group, outpatient study designed to evaluate the efficacy and safety of baricitinib 4-mg or baricitinib 2-mg in adult patients with severe (SALT score of 50% to 94%) or very severe (SALT score of 95% to 100%) scalp AA. Approximately 476 adult patients will be enrolled into Study JAIR.

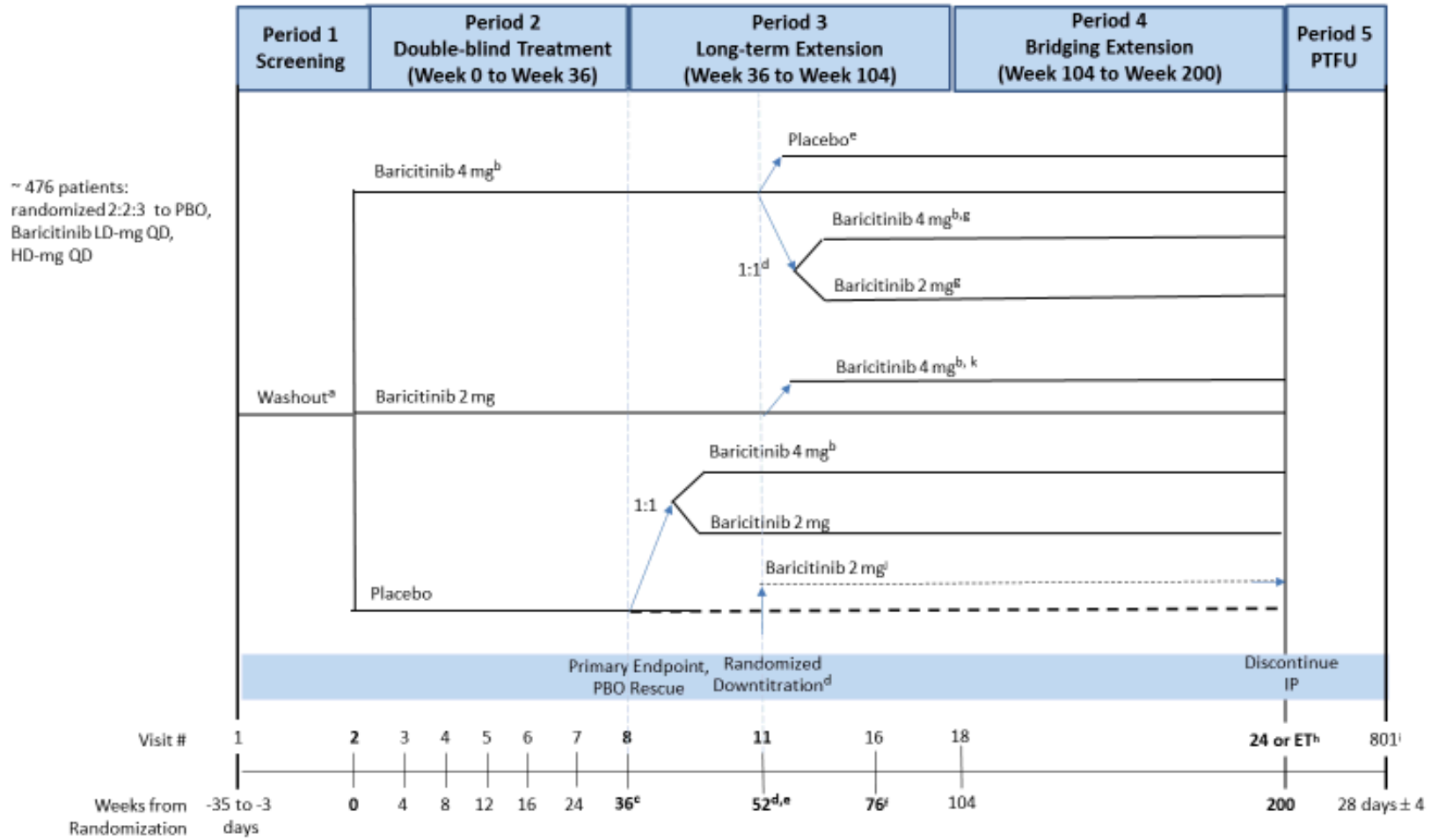
Patients must have a current AA episode of more than 6 months duration prior to screening (Visit 1), with at least 50% scalp involvement at screening AND baseline (Visits 1 and 2) with no spontaneous improvement (no more than a 10-point reduction in SALT) over the past 6 months. Patients with a current episode of severe or very severe AA of more than 8 years will not be eligible for inclusion in the study unless episodes of regrowth, spontaneous or under treatment, have been observed on the affected areas over the past 8 years.

5.1.1. Study Design

The study design includes 5 periods: a 5-week screening period; a 36-week double-blind placebo-controlled treatment period; a 68-week long-term extension period; a 96-week bridging extension; and a posttreatment follow-up period.

- **Period 1:** screening period (Visit 1) is between 3 and 35 days prior to Visit 2 (Week 0).
- **Period 2:** 36-week double-blind, placebo-controlled treatment period is from Week 0 (baseline; Visit 2) to Week 36 (Visit 8).
- **Period 3:** 68-week, long-term extension period with randomized downtitration (for responders) is from Week 36 (Visit 8) to Week 104 (Visit 18).
- **Period 4:** 96-week bridging extension period is from Week 104 (Visit 18) and up to Week 200 (Visit 24).
- **Period 5:** posttreatment follow-up period; the posttreatment follow-up visit should occur approximately 4 weeks after the last dose of investigational product (IP) Patients who have completed Week 200 and who will continue on marketed product beyond Week 200 do not need to complete Period 5 (Visit 801).

Note: Patients who have discontinued IP and remain in the study for more than 28 days without IP will have an Early Termination Visit (ETV); however, a separate follow-up visit (Visit 801) is not required. Figure JAIR.5.1 illustrates the study design. The full visit schedule is outlined in the protocol.



Abbreviations: ClinRO = clinician reported outcome; EC = exclusion criterion; eGFR = estimated glomerular filtration rate; ET = early termination; HD = high dose; IP = investigational product; LD = low dose; PBO = placebo; PTFU = post treatment follow-up; QD = once daily administration; SALT = Severity of Alopecia Tool.

a Applicable to all patients at time of screening. See EC [9] in the protocol for treatments that will require washout.

b The maximal baricitinib dose for patients with renal impairment (defined as eGFR <60 mL/min/1.73 m2) will be 2 mg QD (see Section 5.5.1 in the protocol).

[c] At Week 36, patients in the placebo treatment arm who have not achieved SALT ≤20 will be rescued and rerandomized in a 1:1 ratio to baricitinib 2-mg dose or baricitinib 4-mg dose. Patients in the baricitinib treatment arms will continue in their current treatment arm regardless of treatment response at Week 36. Patients in the placebo arm who have achieved a SALT ≤20 will remain on placebo at Week 36.

[d] At Week 52, responders (SALT ≤20) in the baricitinib 4-mg treatment group who are eligible (i.e., stayed on the same dose of baricitinib from initial randomization at Visit 2) will be randomized in a 1:1 ratio to either stay on baricitinib 4 mg or transition to baricitinib 2 mg (randomized downtitration).

[e] Patients who have been in the baricitinib 4-mg dose treatment group from baseline and who have never achieved a SALT ≤20 by Week 52 AND do not have a ≥2-point improvement in ClinRO measure for eyebrow or eyelash hair loss (nonresponders) at Week 52 will be automatically transitioned to placebo. See footnote "f" for discontinuation criteria at Week 76.

[f] Patients who are nonresponders (SALT ≥20) at Weeks 52 AND 76 will be automatically discontinued from the study at Week 76, unless they have a ≥2-point improvement from baseline in ClinRO measure for eyebrow or eyelash hair loss. See Protocol for more details.

[g] Responders who experience a loss of treatment benefit after Week 52 (>20-point absolute worsening in SALT score) who were randomized to baricitinib 2-mg at Week 52 (randomized downtitration) will be retreated with baricitinib 4-mg, as randomized at baseline (Visit 2). Patients who were randomized to remain on baricitinib 4 mg (randomized downtitration) will continue to receive the same dose of baricitinib.

[h] ET visit is required for patients that terminate IP early. Patients who remain in the study for more than 28 days after discontinuation of IP do not need a separate follow-up visit (Visit 801).

[i] Visit 801 occurs approximately 28 days after the last dose of IP. Patients who have completed Week 200 and will continue on marketed product beyond Week 200 do not need to complete Period 5.

[j] Patients who were randomized to placebo at baseline and were not eligible for rescue to baricitinib at Week 36 (due to spontaneous remission) will be rescued to baricitinib 2 mg if they are nonresponders at Week 52 (SALT ≥20) or experience a loss of treatment benefit after Week 52 (>20-point absolute worsening in SALT score).

[k] Week 52 nonresponders who have been in the baricitinib 2-mg-dose treatment group from baseline will be rescued to baricitinib 4-mg dose. Responders who have been in the baricitinib 2-mg-dose treatment group from baseline who experience a loss of treatment benefit after Week 52 (>20-point absolute worsening in SALT score) will be rescued to baricitinib 4-mg.

Figure JAIR.5.1. Illustration of study design for Clinical Protocol I4V-MV-JAIR(d).

5.2. Method of Assignment to Treatment

Patients who meet all criteria for enrollment will be randomized in a 2:2:3 ratio to receive placebo QD, baricitinib 2-mg QD, or baricitinib 4-mg QD double-blind treatment at Visit 2 (Week 0). Baseline randomization will be stratified by geographic region (North America, Asia, and Rest of World), and duration of current episode at baseline (less than 4 years versus at least 4 years) for the whole study. Randomization for the randomized downtitration period will not be stratified. Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive web-response system (IWRS). The IWRS will be used to assign bottles, each containing double-blind IP tablets, to each patient, starting at Visit 2 (Week 0), and at each visit up to and including Visit 23 (Week 184). Site personnel will confirm that they have located the correct bottles by entering a confirmation number found on the bottle into the IWRS.

This study will be conducted internationally in multiple sites. Table JAIR.5.1 describes how regions will be defined for stratification. Regions may be combined for statistical analyses in the case when 1 of the region strata fails to meet the required minimum number of 30 patients. The 2 region strata with the least number of patients will then be pooled.

Table JAIR.5.1. Geographic Regions for Stratification

Region	Country
North America	United States
Asia	Japan, Korea, Taiwan, China
Rest of World	Australia, Argentina, Brazil, Israel

6. Priori Statistical Methods

6.1. Determination of Sample Size

Study JAIR will screen approximately 678 patients in order to enroll approximately 476 patients. The enrolled patients will be randomized in a 2:2:3 ratio for placebo QD (136 subjects), baricitinib 2-mg QD (136 subjects), or baricitinib 4-mg QD (204 subjects). This sample size will provide approximately 90% power to test the superiority of the baricitinib 4-mg dose to placebo or the superiority of the baricitinib 2-mg dose to placebo in the primary endpoint (the proportion of patients with SALT ≤20 at Week 36) based on a 2-sided Fisher's exact test within the graphical testing scheme, at an initial significance level of 0.04 for 4-mg dose and 0.01 for 2-mg dose. The assumptions used for the power calculation are as follows: 30% response rate for the baricitinib 4-mg dose, 20% response rate for the baricitinib 2-mg dose, and 5% response rate for placebo (Kennedy Crispin et al. 2016; Mackay-Wiggan et al. 2016). The initial α allocation may be adjusted in a later version of the SAP when newer information is obtained on the endpoints that are being tested and will be finalized prior to the primary database lock.

Patients who achieve a SALT ≤20 at Week 52 (responders) AND who have remained on the 4-mg dose of baricitinib from randomization (Visit 2) to Week 52, will enter the randomized downtitration, which is meant to evaluate the change in clinical response after treatment downtitration, and does not account for whether the sample size is sufficient to detect statistical difference between baricitinib 4-mg dose and 2-mg dose. It is expected that there would be approximately 60 patients eligible for the randomized downtitration.

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 will 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 (for example, graphical multiple testing strategy in Section 6.6).

Data collected at ETVs will be mapped to the closest scheduled visit number for that patient if it falls within the visit window as discussed in Section 6.2.2. For by-visit summaries, only visits in which a measure was scheduled to be collected will be summarized. Any unscheduled visit data will be included at the patient-level listings. However, the data may still be used in other analyses, including but not limited to shift analyses for safety analyses, change from baseline

using modified last observation carried forward (mLOCF) endpoint analyses, and other categorical analyses including safety.

6.2.1. Analysis Populations

Table JAIR.6.1. Analysis Populations

Population	Description
Full Analysis Set (FAS)	All patients randomized in Study JAIR will be included in the FAS. Patients will be analyzed according to the IP to which they were randomized at baseline (Visit 2). Of note, FAS is essentially the ITT population.
Modified Full Analysis Set (mFAS) Population	All patients randomized in Study JAIR that received at least 1 dose of IP will be included in the mFAS. It excludes patients with female pattern baldness and male patients with diffuse AGA[a] (Grade IV and above) (Norwood 1975) identified at Week 36. Patients will be analyzed according to the IP to which they were randomized at baseline (Visit 2).
Per-Protocol Set (PPS)	The PPS will include all mFAS patients who are not deemed noncompliant with treatment, who do not have any of the important protocol deviations that exclude patients from the PPS, and whose investigator site does not have significant GCP deviations that require a report to regulatory agencies. The important protocol deviations, including the subset that result in exclusion from the PPS, will be determined while the study team remains blinded, prior to the primary outcome database lock.
Randomized Downtitration Population	All patients who enter the randomized downtitration will be included in the Randomized Downtitration Population. They will be analyzed according to the IP to which they were randomized at Week 52.
Retreated Population	All patients who will be retreated after experiencing loss of treatment benefit on baricitinib 2-mg in the randomized downtitration will be included in the Retreated Population.
Safety Population	The safety population is defined as all patients who were randomized in Study JAIR, received at least 1 dose of IP, and did not discontinue from the study for the reason "Lost to Follow-up" at the first postbaseline visit. Patients will be analyzed according to the IP to which they were actually assigned.

Abbreviations: AGA = androgenetic alopecia; GCP = good clinical practice; IP = investigational product; ITT = intent-to-treat; JAIR = I4V-MC-JAIR.

[a] Some male patients with Grade IV AGA and female patients with patterned baldness may only be identified after hair regrowth.

The efficacy analysis of the primary and key secondary endpoints will be conducted in the full analysis set (FAS) population. All other efficacy or health outcome analyses will be conducted in the FAS population or other populations that are dependent on the objective. Efficacy analyses using the randomized downtitration population or the retreated population will not be performed at the Week 36 primary outcome database lock (PO-DBL). Additional exploratory analyses will be conducted on the FAS population unless otherwise stated.

Safety analyses will be conducted using the safety population. Safety data will be analyzed by treatment cohort. The treatment cohorts include "as randomized" treatment groups and may include "rescued or switched" to baricitinib 4-mg or 2-mg dose, as appropriate. Data from patients randomized to the different treatment groups and followed up to treatment or dose change or data cut (if no treatment or dose change) will be analyzed.

At the PO-DBL, the safety data through Week 36 will be analyzed by treatment groups including placebo, baricitinib 2-mg, or baricitinib 4-mg.

In the rare situation where a patient is lost to follow-up at the first postbaseline visit, but some safety data exists (for example, unscheduled laboratory assessments) after first dose of study drug, a listing of the data or a patient profile will be provided, if requested.

6.2.2. Definition of Baseline and Postbaseline Measures

The baseline utilized in the efficacy analyses depends on the analysis being performed. The baseline value for the efficacy and health outcome analyses for all populations except for randomized downtitration and retreated populations is defined as the last nonmissing measurement on or prior to the date of first study drug administration (expected at Week 0, Visit 2) unless otherwise stated. If a patient is randomized but does not receive study drug, then the date of randomization is used instead of the first dose date. The efficacy and health outcome analyses for the randomized downtitration population will use the measurement on or prior to the date of Visit 11 (Week 52) as a baseline unless otherwise stated. The efficacy and health outcome analyses for the retreated population will use the measurement on or prior to the date when patients got retreated.

Baseline for the safety analyses 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 all nonmissing measurements on or prior to the date of first study drug administration for all other analyses.

Postbaseline measurements are collected after study drug administration through Week 200 (Visit 24) or early discontinuation visit. For data collected in the electronic Clinical Outcomes Assessment (eCOA) tablet (including PROs and ClinROs) and related to efficacy assessments, unscheduled postbaseline visits that fall within the visit windows defined by Lilly will be summarized in the by-visit analyses if there is no scheduled visit available. Refer to clinical protocol I4V-MC-JAIR(d) for detail of the visit windows. If there is more than 1 unscheduled visit within the defined visit window and no scheduled visit is available, the unscheduled visit closest to the scheduled visit date will be used. If 2 unscheduled visits of equal distance are available, then the latter of the 2 will be used.

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

6.2.3. Analysis Methods

Unless otherwise stated, the primary analysis of discrete efficacy and health outcomes variables will use a logistic regression analysis with geographic region, duration of current episode at baseline (<4 years versus ≥4 years), baseline value, and treatment group in the model, except for outcomes related to SF-36 and HADS where the baseline value will not be included . Firth's correction will be used in order to accommodate (potential) sparse response data. The p-value

and 95% CI for the odds ratio from the logistic regression model are used for primary statistical inference, unless Firth's correction still results in quasi-separation. In that case, Fisher's exact test will be used for statistical inference. The difference in percentages and 95% CI of the difference in percentages using the Newcombe-Wilson method without continuity correction are used for descriptive purposes unless otherwise specified. The relative risk and associated 95% CI using the normal approximation method may also be presented. Missing data will generally be imputed using NRI, as described in Section 6.4.1.

The primary analyses for the continuous efficacy and health outcome variables will use analysis of covariance (ANCOVA) with geographic region, duration of current episode at baseline (<4 years versus ≥4 years), treatment group, and baseline value in the model unless otherwise stated. Type III tests for least squares means (LSM) will be used for statistical comparison between treatment groups. The LSM difference, standard error, p-value, and 95% CI will also be reported. The method used to handle missing data will be mLOCF, which will use the most recent nonmissing postbaseline assessment. The specific modification to the LOCF is data after an intercurrent event will not be carried forward to replace the missing data. Additional details of the intercurrent event and mLOCF method are described in Section 6.4 and Section 6.4.2.

Time-to-event analysis will be performed and analyzed using log-rank test. Kaplan–Meier curves will also be produced. A Cox proportional hazards model may be used with treatment and other stratification variables in the model unless otherwise stated. Hazard ratio with CIs may be reported. Diagnostic tests for checking the validity of the proportional hazards assumption may be performed. If the assumption of proportional hazards is not justified, nonproportionality may be modeled by stratification.

Note that for analysis conducted on the randomized downtitration population or retreated population, the geographic region and duration of current episode at baseline may not be used as covariates in the statistical analysis models.

Fisher's exact test will be used to test for differences between the baricitinib and placebo groups for adverse events (AEs), discontinuations, and other categorical safety data. Continuous vital signs, body weight, and other continuous safety variables, including laboratory variables, will be analyzed using an ANCOVA with treatment 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 0; 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 will also be provided. In addition to the LSM for each group, the within-group p-value for the change from baseline will be displayed.

Due to the COVID-19 pandemic, some visits may have been conducted remotely. In order to evaluate the impact of remote visits on the clinical trial, sites were required to record the visit method (e.g., onsite visit, virtual visit, etc.) for visits beginning March 1, 2020. For data collected at the unscheduled postbaseline visit that falls within the visit window, the visit method should be considered the same as recorded for the scheduled visit for this window. If the visit

method is a telephone interview or a virtual visit, the visit is considered remote. However, if the visit method is missing for the scheduled visit, but central lab was collected and/or vital assessments are available, then it will be considered an onsite visit, otherwise it will be considered a remote visit.

6.2.4. Derived Data

- age (year)
- age group (<40, ≥40 years old; <60, ≥60 years old; <65, ≥65 years old)
- weight group (<60 kg, ≥60 to <100 kg, ≥100kg)
- body mass index (BMI) (kg/m^2) = weight (kg)/([height (cm)/100]^2)
- BMI groups (<25 kg/m^2, ≥25 to <30 kg/m^2, ≥30 kg/m^2)
- the duration from onset of AA (year) = [(date of informed consent – date of AA onset)+1]/365.25
  If year of onset is missing, duration of AA will be set as missing. Otherwise, unknown month will be taken as January, and unknown day will be taken as 01. The duration of AA will be rounded to 1 decimal place before deriving any duration categories.
- the duration from onset of AA (years) category (<5; ≥5 to <10; ≥10 to <15; ≥15 years)
- the AA onset age: derived using AA onset date as the reference start date and July 1 of birth year and truncated to a whole-integer age
- the AA onset age category (<18; ≥18 years old)
- duration of the current episode of AA (year) at baseline = [(Date of first dose–Date of current episode of AA onset)+1]/365.25. If a patient is randomized but does not receive study drug, then the date of randomization is used instead of the first dose date. The duration of current episode of AA will be rounded to 1 decimal place before deriving any duration categories.
- duration of the current episode of AA at baseline category (≥0.5 to <1; ≥1 to <2; ≥2 to <4; ≥4 to <8; ≥8 years)
- duration of the current episode of AA at baseline category (≥0.5 to <4; ≥4 to <8; ≥8 years)
- duration of the current episode of AA at baseline category (<4; ≥4 years)
- change from baseline = postbaseline measurement at Visit x–baseline measurement
  If a baseline value is missing, it will not be imputed and the change from baseline will not be calculated.
- percent change from baseline at Visit x:
  ([Postbaseline measurement at Visit x–baseline measurement]/baseline measurement)*100
  If a baseline value is missing, it will not be imputed and the percent change from baseline will not be calculated.
- weight (kg) = weight (lbs)*0.454
- height (cm) = height (in)*2.54

6.3. Adjustments for Covariates

The randomization to treatment groups at Week 0 (Visit 2) is stratified by duration of current episode at baseline and geographic region. Unless otherwise specified, the statistical analysis models will adjust for duration of current episode at baseline and geographical region. The covariates used in the logistic model for categorical data will additionally include the parameter value at baseline except for endpoints related to SF-36 and HADS. The covariates used in the ANCOVA model for continuous data generally will include the parameter value at baseline. Inclusion of baseline in the ANCOVA model ensures treatment LSM are estimated at the same baseline value.

6.4. Handling of Dropouts or Missing Data

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 but not limited to the following:

- application of 1 of the censoring rules (including after permanent study drug discontinuation, after rescue therapy, or retreatment)
- discontinuation of inadvertently enrolled patients
- discontinuation from the study due to enrollment in other trials, medical, safety, regulatory reasons, investigator decision, and patient decision
- missing an intermediate visit prior to discontinuation, rescue, or retreatment
- lost to follow-up

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

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

The primary censoring rule will censor efficacy and health outcome results after permanent study drug discontinuation or results that were collected during remote visits due to the COVID-19 pandemic. Therefore, the data collected remotely will be considered “missing”. This censoring rule will generally be applied to all efficacy and health outcome endpoints and conducted for all defined efficacy analysis populations except for the randomized downtitration population (defined in Section 6.2.1).

A secondary censoring rule will censor efficacy and health outcome results after permanent study drug discontinuation. This censoring rule will not exclude the data collected during remote visits due to the COVID-19 pandemic and will be applied to selected efficacy and health outcome endpoints conducted for the FAS population (defined in Section 6.2.1).

A tertiary censoring rule will censor efficacy and health outcome results after permanent study drug discontinuation or after retreatment. This censoring rule will be applied to the randomized downtitration population (defined in Section 6.2.1).

Table JAIR.6.4 describes the planned imputation methods for selected endpoints, including but not limited to, primary and key secondary efficacy and health outcome endpoints with associated censoring rules. Sections 6.4.1 through 6.4.4 summarize the imputation methods for the various efficacy and health outcome endpoints.

6.4.1. Nonresponder Imputation

For the analyses of categorical efficacy and health outcomes variables such as SALT ≤20 and PRO for Scalp Hair Assessment score of 0 or 1 with a ≥2-point improvement from baseline, the primary imputation method when an intercurrent event occurs will be NRI, which can be justified based on the composite strategy (ICH 2019) for handling intercurrent events. This imputation procedure assumes that the effects of treatments disappear after the occurrence of the intercurrent event. For analyses that utilize any of the censoring methods, randomized patients without at least 1 postbaseline observation will also be defined as nonresponders for all visits. As well, patients who are missing a value prior to discontinuation, rescue, or retreatment (if censoring on rescue or retreatment), i.e., the patient is missing an intermediate visit, will be imputed as nonresponders on that visit only.

6.4.2. Modified Last Observation Carried Forward

For continuous efficacy and health outcome variables, such as SALT percent change from baseline, a mLOCF imputation technique replaces missing data with the most recent nonmissing postbaseline assessment. The specific modification to the LOCF is data after an intercurrent 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 remains 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 indication as very few patients experienced waxing and waning in scalp hair coverage during the course of treatment from the Phase 2 portion of the I4V-MC-JAHO trial. The persistence in treatment effect is also demonstrated in the clinical response seen in other AA studies (Mackay-Wiggan et al. 2016).

6.4.3. Hybrid Imputation (Multiple Imputation and Nonresponder Imputation for Categorical Variables; Multiple Imputation and Modified Last Observation Carried Forward for Continuous Variables)

To determine the effect of missing data due to the COVID-19 pandemic on the clinical trial, a sensitivity analysis will be conducted using hybrid imputation method. The missing data due to the COVID-19 pandemic includes the data collected remotely but considered as “missing” or

data which were not collected due to the COVID-19 pandemic (i.e., some efficacy assessments are not to be collected at the remote visits or the whole visit was missing due to pandemic).

For the binary endpoints, the hybrid method will impute the missing data due to COVID-19 by multiple imputation (MI) whereas other missing data not due to COVID-19 by NRI. This imputation procedure addresses the hybrid estimand assuming that the effects of treatments will be the same had patients not experienced any intercurrent event related to COVID-19 (e.g., either remote visits or missed visits due to COVID-19, etc.) or the effect will disappear after any intercurrent event not related to COVID-19. Specifically, the algorithm is as follows:

1. Identify all missing data (including the missing data due to COVID-19 and not due to COVID-19).
2. Implement the MI to impute all missing data and generate *m* imputed complete data sets.
3. Identify the missing data due to COVID-19 and not due to COVID-19 in the original data set.
4. For each of these *m* imputed complete data sets from Step 2, the imputed data for missing data not due to COVID-19 will be replaced by NRI and all other data including imputed or observed will be used to derive the binary outcome.

For the continuous endpoints, the hybrid method will impute the missing data due to COVID-19 by MI whereas other missing data not due to COVID-19 by mLOCF. This imputation procedure addresses the hybrid estimand assuming that the effects of treatments will be the same had patients not experienced any intercurrent event related to COVID-19 (e.g., either remote visits or missed visits due to COVID-19, etc.) or will remain the same after the event that caused missing data not due to COVID-19 as it was just prior to the missing data event. Specifically, the algorithm is as follows:

1. Identify all missing data (including the missing data due to COVID-19 and not due to COVID-19).
2. Implement the MI to impute all missing data and generate *m* imputed complete data sets.
3. Identify the missing data due to COVID-19 and not due to COVID-19 in the original data set.
4. For each of these *m* imputed complete data sets from Step 2, the imputed data for missing data not due to COVID-19 will be set as missing again and imputed by mLOCF.

The sensitivity analysis aforementioned will be performed on the primary and key secondary endpoints. The number of imputed data sets will be m=100 and a 6-digit seed value will be prespecified for each analysis. Within the program, the seed will be used to generate the *m* seeds needed for imputation. The initial seed values are given below:

Table JAIR.6.2. Seed Values for Multiple Imputation

Analysis	Seed value
Proportion of patients achieving SALT≤ 20 at Weeks 16, 24, and 36	123450
Proportion of patients achieving a PRO for Scalp Hair Assessment 0 or 1 with a ≥2-point improvement from baseline at Week 36	123451
Proportion of patients achieving an absolute SALT score ≤10 at Weeks 24 and 36	123450
Proportion of patients achieving $SALT_{90}$ at Week 36	123450
Proportion of patients achieving $SALT_{50}$ at Week 12	123450
Percent change from baseline in SALT score at Week 36	123450
Proportion of patients achieving ClinRO Measure for EB Hair Loss score 0 or 1 with ≥2-point improvement from baseline at Week 36	123452
Proportion of patients achieving ClinRO Measure for EL Hair Loss score 0 or 1 with ≥2-point improvement from baseline at Week 36	123453

Abbreviations: ClinRO = clinician-reported outcome; EB = eyebrow; EL = eyelash; PRO = patient-reported outcome; SALT = Severity of Alopecia Tool; $SALT_{50}$ = at least 50% improvement from baseline in SALT score; $SALT_{90}$ = at least 90% improvement from baseline in SALT score.

Analysis: A logistic regression or ANCOVA will be applied, as appropriate, on each imputed data set. Details about logistic and ANCOVA models can be found in Section 6.2.3. The final inference on treatment difference is conducted from the multiple data sets using Rubin's combining rules, as implemented in SAS PROC MIANALYZE.

6.4.4. Tipping Point Analyses

To investigate the missing data mechanism, an additional analysis using MI under the missing not-at-random assumption will be provided for the primary objective, which compares the proportion of patients achieving SALT ≤20 of each of the Baricitinib 4-mg and 2-mg doses and placebo at Week 36. The tipping point analysis may also be used as an additional analysis for some key secondary objectives.

All patients in the FAS population are included. Data after the occurrence of intercurrent events (including application of any of the censoring rules) will be set to missing. Within each analysis, the most extreme case will be considered, in which all missing data for patients randomized to baricitinib 2-mg or 4-mg will be imputed using the worst possible result, and all missing data for patients randomized to placebo will be imputed with the best possible result. Treatment differences will be analyzed using logistic regression or ANCOVA, as appropriate.

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

1. To handle intermittent missing visit data, a Markov chain Monte Carlo (MCMC) method (SAS Proc MI with MCMC option) will be used to create a monotone missing pattern.
2. A set of Bayesian regressions (using SAS Proc MI with MONOTONE option) will be used for the imputation of monotone dropouts. Starting from the first visit with at least

1 missing value, the regression models will be fit sequentially with treatment as a fixed effect and values from the previous visits as covariates.

3. A delta score is added to all imputed scores at the time point where the analysis is conducted for patients in the baricitinib treatment groups, thus, worsening the imputed value. The delta score is capped for patients based on the range of the outcome measure being analyzed.
4. Treatment differences between baricitinib and placebo are analyzed for each imputed data set using ANCOVA. Results across the imputed data sets are aggregated using SAS Proc MIANALYZE in order to compute a p-value for the treatment comparisons for the given delta value.
5. Steps 3 and 4 are repeated, and the delta value added to the imputed baricitinib scores is gradually increased. The tipping point is identified as the delta value which leads to a loss of statistical significance (aggregated p-value >0.05) when evaluating baricitinib relative to the placebo group.

As a reference, for each delta value used in steps 3 through 5, a fixed selection of delta values (ranging from slightly negative to slightly positive) will be added to imputed values in the placebo group, and step 4 will be performed for the combination. This will result in a 2-d table for each time point of interest, with the columns representing the delta values added to the imputed placebo responses, and the rows representing the delta values added to the imputed baricitinib responses. Separate 2-d tables will compare each baricitinib dose group to placebo.

A similar process will be used for the categorical variables:

1. Missing responses in the baricitinib groups will be imputed with a range of low response probabilities, including probabilities of 0, 0.1, and 0.2.
2. For missing responses in the placebo group, a range of response probabilities will be used to impute the missing values. Multiple imputed data sets will be generated for each response probability.
3. Treatment differences between baricitinib and placebo are analyzed for each imputed data set using logistic regression. Results across the imputed data sets are aggregated using SAS Proc MIANALYZE in order to compute a p-value for the treatment comparisons for the given response probability. If the probability values do not allow for any variation between the multiple imputed data sets (for example, all missing responses in the placebo and baricitinib groups are imputed as responders and nonresponders, respectively), then the p-value from the single imputed data set will be used.

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

For tipping point analyses the number of imputed data sets will be m=100 and the seed value to start the pseudorandom number generator of SAS Proc MI (same values for MCMC option and for MONOTONE option) will be as specified in Table JAIR.6.3.

Table JAIR.6.3. Seed Values for Tipping Point Analyses

Analysis	Seed value
Proportion of patients achieving a SALT ≤20 at Week 36	123461

Abbreviations: SALT = Severity of Alopecia Tool

Table JAIR.6.4. Imputation Techniques for Various Variables

Analysis population	Endpoints	Imputation
FAS	SALT ≤20	NRI[a,b], MI[a]+NRI[a], Tipping Point[a]
	PRO for Scalp Hair Assessment score of 0 or 1 with a ≥2-point improvement from baseline	NRI[a,b], MI[a]+NRI[a]
	$SALT_{50}$, $SALT_{90}$, absolute SALT score ≤10	NRI[a,b], MI[a]+NRI[a]
	ClinRO Measure for EB Hair Loss 0 or 1 with a ≥2-point improvement from baseline	NRI[a,b], MI[a]+NRI[a]
	ClinRO Measure for EL Hair Loss 0 or 1 with a ≥2-point improvement from baseline	NRI[a,b], MI[a]+NRI[a]
	SALT PCFB	mLOCF[a,b], MI[a]+mLOCF[a]
For all other categorical and continuous efficacy or health outcome analyses, details of censoring rule or imputation implementation will be found in Table JAIR.6.6.		

Abbreviations: ClinRO = clinician-reported outcome; EB = eyebrow; EL = eyelash; FAS = full analysis set; mLOCF = modified last observation carried forward; MI = multiple imputation; NRI = nonresponder imputation; PCFB = percent change from baseline; PRO = patient reported outcome; SALT = Severity of Alopecia Tool; $SALT_{50}$ = at least 50% improvement from baseline in SALT score; $SALT_{90}$ = at least 90% improvement from baseline in SALT score.

[a] Analyses utilizing the primary censoring rule.

[b] Analyses utilizing the secondary censoring rule.

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.

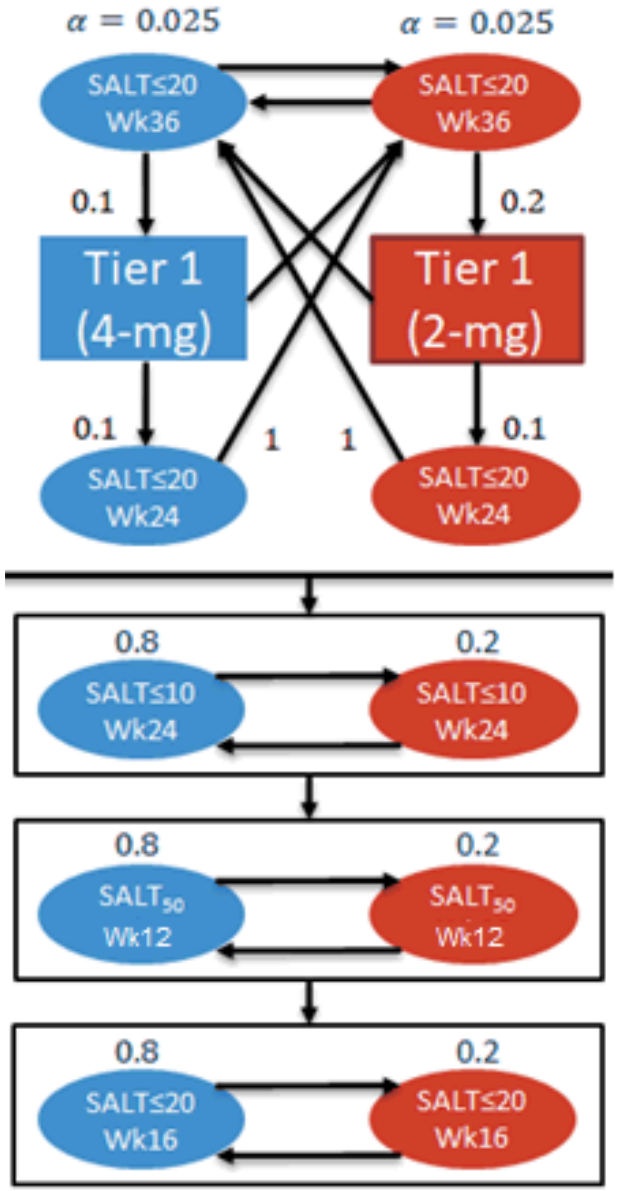
6.6. Multiple Comparisons/Multiplicity

Multiplicity adjusted analyses will be performed on the primary and key secondary objectives (See Sections 4.1 and 4.2) using the FAS population to control the overall familywise Type I error rate at a 2-sided α level of 0.05. The graphical multiple testing procedure described in Bretz et al. (2011) will be used. The graphical approach is a closed testing procedure; hence, it strongly controls the familywise error rate across all endpoints (Alosh et al. 2014). Figure JAIR.6.1 illustrates the graphical testing procedures that will be used. The secondary endpoints tested at Week 36 are grouped together and labelled as Tier 1 group of endpoints. Figure JAIR.6.2 illustrates the testing scheme used within the Tier 1 group. The testing steps are outlined below:

Step 1: The primary endpoint SALT ≤20 will be first tested at a 2-sided α=0.025 for both 2-mg and 4-mg doses. If neither of the null hypotheses is rejected, no further testing is conducted, as the α for that test is considered "spent" and cannot be passed to other endpoints. If at least 1 of null hypotheses is rejected, the testing process continues to Step 2, with the remaining α propagated according to the weights on the corresponding edges displayed in

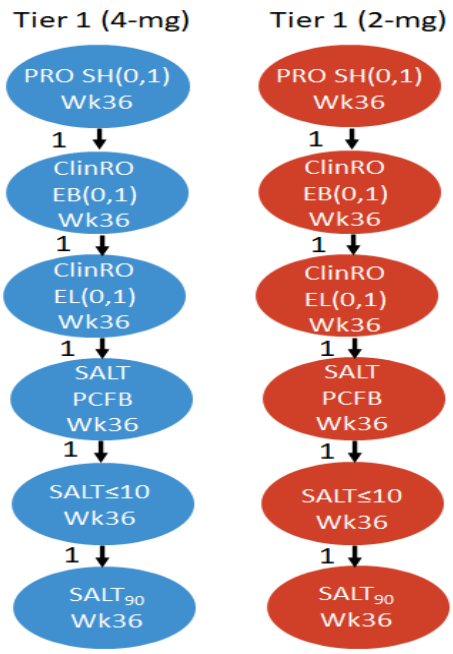
Step 2: The testing process continues as long as there is at least 1 hypothesis in the scheme that can be rejected at its allocated α level at that point. Each time a hypothesis is rejected, the graph is updated to reflect the reallocation of α, which is considered "recycled" by Alosh et al. (2014). This iterative process of updating the graph and reallocating α is repeated until all hypotheses have been tested or when no remaining hypotheses can be rejected at their corresponding α levels.

Step 3: If any of the endpoint in either dose up to SALT ≤20 at Week 24 is not rejected, the testing procedure will stop. Otherwise, the remainder 3 endpoints, SALT ≤10 at Week 24, at least 50% improvement from baseline in SALT score ($SALT_{50}$) at Week 12, and SALT ≤20 at Week 16, will be tested in a sequential manner. But each time an endpoint is tested, the significance level α will be allocated to 2-mg and 4-mg according to the proportion shown in Figure JAIR.6.1.



Abbreviations: SALT = Severity of Alopecia Tool; Wk = week.

Figure JAIR.6.1. Overview of the graphical testing procedure for I4V-MC-JAIR.



Abbreviations: ClinRO = clinician-reported outcome; EB = eyebrow; EL = eyelash; PCFB = percent change from baseline; PRO = patient-reported outcome; SALT = Severity of Alopecia Tool; $SALT_{90}$ = at least 90% improvement from baseline in SALT score; Wk = week.

Figure JAIR.6.2. Graphical testing procedure within Tier 1 group of endpoints

6.7. Patient Disposition

An overview of patient populations will be summarized by treatment group. Frequency counts and percentages of patients excluded prior to randomization by primary reason for exclusion will be provided for patients who failed to meet study entry requirements during screening.

Patient study disposition will be summarized using the FAS 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 and the reason for study discontinuation. Treatment disposition will also be summarized using the FAS population. Frequency counts and percentages of patients who complete the treatment through a certain period of time or discontinue treatment early will also be summarized separately by treatment group and the reason for treatment discontinuation. A listing of patient disposition will be provided for the FAS population, with the extent of their participation in the study and the reason for discontinuation. A listing of all patients in the FAS population with their treatment assignment will also be provided.

6.8. Patient Characteristics

Patient characteristics including demographics and baseline characteristics will be summarized descriptively by treatment group. Analyses will be presented for the FAS population. Historical illnesses and preexisting conditions will be summarized descriptively by treatment group for the FAS population. No formal statistical comparisons will be made among treatment groups unless otherwise stated.

6.8.1. Demographics

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

- Age
- Age group (<40 vs ≥40 years old)
- Age group (<60 vs ≥60 years old)
- Age group (<65 vs ≥65 years old)
- Genetic gender (female, male)
- Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Multiple)
- Ethnicity (US patients only: Hispanic or Latino, Non-Hispanic, and non-Latino, Not reported)
- Region (as defined in Table JAIR.5.1)
- Country
- Weight (kg)
- Weight group (<60 kg, ≥60 to <100 kg, ≥100 kg)
- Height (cm)
- BMI (kg/m^2)
- BMI groups (<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 FAS population.

6.8.2. Baseline Disease Characteristics

The following baseline disease information (but not limited to only these) will be categorized and presented for baseline AA clinical characteristics, baseline health outcome measures, and other baseline demographic and disease characteristics as described above:

- duration from onset of AA (years)
- duration from onset of AA category (<5; ≥5 to <10; ≥10 to <15; ≥15 years)
- age at onset of AA (years)
- age at onset of AA category (<18 vs ≥18 years old)
- duration of the current episode of AA
- duration of the current episode of AA category (≥0.5 to <1; ≥1 to <2; ≥2 to <4; ≥4 to <8; ≥8 years)
- duration of the current episode of AA category (≥0.5 to <4; ≥4 to <8; ≥8 years)
- duration of the current episode of AA category (<4 vs ≥4 years)

- habits (alcohol: never, current, former; tobacco: never, current, former)
- with atopic background versus no atopic background
  - Atopic background is defined as "medical history of, or on-going Atopic Dermatitis, or allergic rhinitis, or allergic conjunctivitis, or allergic asthma."
- SALT Score
- SALT category: Severe (SALT score of 50% to 94%) vs very severe (SALT score of 95% to 100%)
- classified as ophiasis
- classified as universalis
- Hamilton-Norwood Scale (applies only to male patients) (Norwood 1975)
- PRO for Scalp Hair Assessment
- PRO measure for eyebrows
- PRO measure for eyelashes
- PRO measure for eye irritation
- PRO measure for nail appearance
- ClinRO measure for eyebrow hair loss
- ClinRO measure for eyelash hair loss
- ClinRO measure for nail appearance
- Skindex-16 AA
- HADS (anxiety and depression domain total scores will be presented separately)
- prior therapy (naïve, Systemic [All Immunosuppressants/Immunomodulators], Systemic Agents [Corticosteroids]*, Systemic Agents [Janus kinases (JAK) inhibitor]*, Systemic Agents [others]*, Other Systemic [Nonimmunosuppressant], Intralesional Therapy, Topical Therapy excluding Immunotherapy, Topical Immunotherapy, Procedures, Phototherapy)
- screening period renal function status: impaired (estimated glomerular filtration rate [eGFR] <60 mL/min/1.73 m^2) or not impaired (eGFR ≥60 mL/min/1.73 m^2)
- immunoglobulin E (IgE): <200 kU/I or ≥200 kU/I

*These 3 categories are subcategories of Systemic [All Immunosuppressants/Immunomodulators]

6.8.3. Historical Illness and Preexisting Conditions

Historical illnesses are defined as those conditions recorded in the Preexisting Conditions and Medical History electronic case report form (eCRF) or 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 FAS population. Historical diagnoses will be categorized using the Medical Dictionary for Regulatory Activities (MedDRA, most current available version) algorithmic standardized MedDRA queries (SMQs) or similar predefined lists of preferred terms (PTs) of interest.

Preexisting conditions are defined as those conditions with a start date prior to the informed consent date and an end date after the informed consent date or have no stop date (ie, are ongoing). In addition, AEs that occur prior to the first dose are also included. For events

occurring on the day of the first dose of study treatment, the date and time of the onset of the event will both be used to determine if the event was 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 MedDRA SMQs or similar predefined lists of 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 FAS population.

6.9. Treatment Compliance

Patient compliance with study medication by counting returned tablets will be assessed at each scheduled visit by treatment period

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. Similarly, a patient will be considered significantly noncompliant if he/she is judged by the investigator to have intentionally or repeatedly taken more than the prescribed amount of medication during the study (ie, compliance ≥120%). 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 last dose in the period of interest–date of first dose in the period of interest + 1)–number of days of temporary drug interruption]*number of tablets taken per day

Patients who are significantly noncompliant through Week 36 will be excluded from the per-protocol set (PPS) population.

Descriptive statistics for percent compliance and noncompliance rate will be summarized for the FAS population by treatment group for Weeks 0 through 36, with data up to permanent treatment discontinuation. Subintervals of interest, such as compliance between visits, may also be presented. The number of expected doses, tablets dispensed, tablets returned, and percent compliance will be listed by patient for Weeks 0 to 36, with data up to permanent treatment discontinuation.

6.10. Previous and Concomitant Therapy

Summaries of previous AA therapies will be based on the FAS population. Concomitant medications will be summarized by treatment period.

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

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

- previous AA therapies
- previous AA therapies including reason for discontinuation

Summaries of concomitant medications will be provided as well.

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 following formats and patients will be analyzed according to the IP to which they were randomized at baseline.

Table JAIR.6.5 includes the descriptions and derivations of the primary, secondary, and exploratory efficacy outcomes.

Table JAIR.6.6 provides the detailed analyses including analysis type, method and imputation, population, time point, and comparisons for efficacy analyses.

Table JAIR.6.5. Description and Derivation of Primary, Secondary and Exploratory Efficacy Outcomes

Description and Derivation of Primary, Secondary and Exploratory Efficacy Outcomes

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
Severity of Alopecia Tool (SALT)	The SALT uses a visual aid showing the division of the scalp hair into 4 areas with the top of the head constituting 40% of total surface, the posterior/back of head 24%, right side and left side of head 18% each. The percentage of hair loss in each area is determined and is multiplied by the percentage of scalp covered by that area. The total sum of the 4 products of each area will give the SALT score, as developed by the National Alopecia Areata Foundation Working Committee (Olsen et al. 2004). Only terminal hair is included in the SALT; vellus hair or any fine downy hair is not taken into account in the SALT scoring process (Olsen et al. 1999, 2004). The SALT score will range from 0% to 100%.	SALT score	Derive the SALT score as follows: SALT=percentage of hair loss on the top of scalp*40% + percentage of hair loss on the posterior/back of scalp*24% + percentage of hair loss on the left side of scalp*18% + percentage of hair loss on the right side of scalp*18%. SALT will be rounded to a whole number before deriving any subsequent variables.	N/A – partial assessments cannot be saved
		Change from baseline in SALT score; Percent change from baseline in SALT score	Change from baseline: observed SALT score – baseline SALT score. % change from baseline: $100 \times \frac{Observed\ score - Baseline}{Baseline}$	Missing if baseline or observed value is missing
		$SALT_{30}$	Improvement in baseline ≥30% [% change from baseline ≤-30]	Missing if baseline or observed value is missing
		$SALT_{50}$	Improvement in baseline ≥50% [% change from baseline ≤ -50]	Missing if baseline or observed value is missing
		$SALT_{75}$	Improvement in baseline ≥75% [% change from baseline ≤-75]	Missing if baseline or observed value is missing

Description and Derivation of Primary, Secondary and Exploratory Efficacy Outcomes

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
SALT		$SALT_{90}$	Improvement in baseline ≥90% [% change from baseline ≤-90]	Missing if baseline or observed value is missing
		$SALT_{100}$	Improvement in baseline = 100% [% change from baseline = -100]	Missing if baseline or observed value is missing
		SALT score ≤20	Observed SALT score ≤20	Missing if observed value is missing
		Absolute SALT score ≤10	Observed SALT score ≤10	Missing if observed value is missing
		Time to achieve SALT ≤20	Date of visit for first time achieving SALT ≤20–randomization date at Visit 2	Censored at the last SALT collection date, scheduled visit date or ETV date during the blind treatment period, whichever is the latest and applicable
Patient-Reported Outcomes (PROs) for Scalp Hair Assessment	It's a novel PRO assessment of the patient's current extent of scalp involvement. It is comprised of 5 category response options: 0 = No missing hair (0% of my scalp is missing hair; I have a full head of hair); 1 = A limited area (1% to 20% of my scalp is missing hair); 2 = A moderate area (21% to 49% of my scalp is missing hair); 3 = A large area (50% to 94% of my scalp is missing hair); and 4 = Nearly all or all (95% to 100% of my scalp is missing hair).	PRO for Scalp Hair Assessment score	Single item. Range: 0 to 4	Single items, missing if missing
		PRO for Scalp Hair Assessment score of 0 or 1 with a ≥2-point improvement from baseline	Observed score of 0 or 1 and change from baseline ≤-2	Missing if baseline or observed value is missing
		PRO for Scalp Hair Assessment score of 0 or 1	Observed score of 0 or 1	Single items, missing if missing

Description and Derivation of Primary, Secondary and Exploratory Efficacy Outcomes

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
PRO for Appearance of Eyebrows (EB)	It's a novel PRO assessment of the patient's current appearance of eyebrows. It is comprised of 4 category response options: 0 = I have full EB on each eye; 1= I have a minimal gap(s) or a minimal amount of thinning in at least 1 of my EBs; 2 = I have a large gap(s) or a large amount of thinning in at least 1 of my EBs; and 3 = I have no or barely any EB hairs.	PRO measure for EB	Single item. Range: 0 to 3	Single items, missing if missing
		PRO Measure for EB 0 or 1 with ≥2-point improvement from baseline	Observed score of 0 or 1 and change from baseline ≤-2	Missing if baseline or observed value is missing
PRO for Appearance of Eyelashes (EL)	It's a novel PRO assessment of the patient's current appearance of EL. It is comprised of 4 category response options: 0 = I have full EL on each eyelid; 1 = I have a minimal gap or minimal gaps along the eyelids; 2 = I have a large gap or large gaps along the eyelids; and 3 = I have no or barely any EL hair.	PRO measure for EL	Single item. Range: 0 to 3	Single items, missing if missing
		PRO Measure for EL 0 or 1 with ≥2-point improvement from baseline	Observed score of 0 or 1 and change from baseline ≤-2	Missing if baseline or observed value is missing
PRO for Eye Irritation (EI)	It's a novel PRO assessment of the patient's extent of EI. It is comprised of 4 category response options: 0 = My eyes have not been irritated; 1 = My eyes have been a little irritated; 2 = My eyes have been moderately irritated; and 3 = My eyes have been severely irritated.	PRO Measure for EI	Single item. Range: 0 to 3	Single items, missing if missing
		PRO Measure for EI 0 or 1 with ≥2-point improvement from baseline	Observed score of 0 or 1 and change from baseline ≤-2	Missing if baseline or observed value is missing

Description and Derivation of Primary, Secondary and Exploratory Efficacy Outcomes

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
PRO for Nail Appearance	It's a novel PRO assessment of the patient's current nail appearance. It is comprised of 4 category response options: 0 = Nails are not at all damaged (e.g., pitted, rough, brittle, split); 1 = At least 1 nail is a little damaged (e.g., pitted, rough, brittle, split); 2 = At least 1 nail is moderately damaged (e.g., pitted, rough, brittle, split); 3 = At least 1 nail is very damaged (e.g., pitted, rough, brittle, split) or you have lost at least 1 nail.	PRO Measure for Nail Appearance	Single item. Range: 0 to 3	Single items, missing if missing
		PRO Measure for Nail Appearance 0 or 1 with ≥2-point improvement from baseline (among patients with PRO Measure for Nail Appearance ≥2 at baseline)	Observed score of 0 or 1 and change from baseline ≤-2	Missing if baseline or observed value is missing
Clinician-Reported Outcomes (ClinRO) for EB Hair Loss	It's a novel ClinRO assessment measuring patient's EB hair loss. It is comprised of 4 category response options: 0 = The EB have full coverage and no areas of hair loss; 1 = There are minimal gaps in EB hair and distribution is even; 2 = There are significant gaps in EB hair or distribution is not even; 3 = No notable EB.	ClinRO measure for EB Hair Loss	Single item. Range: 0 to 3	Single items, missing if missing
		ClinRO Measure for EB Hair Loss 0 or 1 with a ≥2-point improvement from baseline (among patients with ClinRO Measure for EB Hair Loss ≥2 at baseline)	Observed score of 0 or 1 and change from baseline ≤-2.	Missing if baseline or observed value is missing

Description and Derivation of Primary, Secondary and Exploratory Efficacy Outcomes

Measure	Description	Variable	Derivation/Comment	Imputation Approach if with Missing Components
ClinRO for EL Hair Loss	It's a ClinRO assessment measuring patient's EL hair loss. It is comprised of 4 category response options: 0 = The EL form a continuous line along the eyelids on both eyes; 1 = There are minimal gaps and the EL are evenly spaced along the eyelids on both eyes; 2 = There are significant gaps along the eyelids or the EL are not evenly spaced along the eyelids; 3 = No notable EL.	ClinRO measure for EL Hair Loss	Single item. Range: 0 to 3	Single items, missing if missing
		ClinRO Measure for EL Hair Loss 0 or 1 with a ≥2-point improvement from baseline (among patients with ClinRO Measure for EL Hair Loss ≥2 at baseline)	Observed score of 0 or 1 and change from baseline ≤-2.	Missing if baseline or observed value is missing
ClinRO for Nail Appearance	It's a novel clinician-reported outcome ClinRO assessment measuring patient's nail appearance. It is comprised of 4 category response options: 0 = Nails are not at all damaged (e.g., pitted, rough, brittle, split); 1 = At least 1 nail is a little damaged (e.g., pitted, rough, brittle, split); 2 = At least 1 nail is moderately damaged (e.g., pitted, rough, brittle, split); 3 = At least 1 nail is very damaged (e.g. pitted, rough, brittle, split) or subject has lost at least 1 nail.	ClinRO Measure for Nail Appearance	Single item. Range: 0 to 3	Single items, missing if missing
		ClinRO Measure for Nail Appearance 0 or 1 with a ≥2-point improvement from baseline (among patients with ClinRO Measure for Nail Appearance ≥2 at baseline)	Observed score of 0 or 1 and change from baseline ≤-2.	Missing if baseline or observed value is missing

Abbreviations: ETV = early termination visit; N/A = not applicable; $SALT_{30,50,75,90,100}$ = at least 30/50/75/90/100% improvement from baseline in SALT score.

Table JAIR.6.6. Description of Primary, Secondary and Exploratory Efficacy Analyses

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
Severity of Alopecia Tool (SALT)	Proportion of patients achieving SALT ≤20	Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Primary analysis
		Logistic regression using NRI[a]	mFAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic regression using NRI[a]	PPS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic regression using MI[a]+NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic Regression using NRI[b]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Sensitivity analysis
		Tipping point analysis[a] with logistic regression	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 16 and 24	Key secondary analysis
		Logistic regression using NRI[a]	mFAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 16 and 24	Supplementary analysis
		Logistic regression using MI[a]+NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 16 and 24	Supplementary analysis
		Logistic regression using NRI[b]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 16 and 24	Sensitivity analysis
		Logistic regression using NRI[a]	FAS (Severe SALT subgroup[c]); FAS (Very severe SALT subgroup[c])	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Dosing evaluation analysis

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
		Logistic regression using NRI[a]	FAS (Duration of current AA episode <4 years subgroup[d]); FAS (Duration of current AA episode >4 years subgroup[d])	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Dosing evaluation analysis

Description of Primary, Secondary and Exploratory Efficacy Analyses

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
	Proportion of patients achieving $SALT_{100}$	Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Other secondary analysis
	Time to achieve SALT ≤20	Time-to-event analysis[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO up to Week 36	Other secondary analysis
	Proportion of patients achieving an absolute SALT ≤10	Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Key secondary analysis
		Logistic regression using NRI[a]	mFAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Supplementary analysis
		Logistic regression using MI[a]+NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Supplementary analysis
		Logistic regression using NRI[b]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Sensitivity analysis
	Proportion of patients achieving a $SALT_{90}$	Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Key secondary analysis
		Logistic regression using NRI[a]	mFAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic regression using MI[a]+NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic regression using NRI[b]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36;	Sensitivity analysis
		Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 24;	Other secondary analysis

Description of Primary, Secondary and Exploratory Efficacy Analyses

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
SALT	Proportion of patients achieving a $SALT_{50}$	Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 12	Key secondary analysis
		Logistic regression using NRI[a]	mFAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 12	Supplementary analysis
		Logistic regression using MI[a]+NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 12	Supplementary analysis
		Logistic regression using NRI[b]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 12	Sensitivity analysis
		Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 16, 24, and 36	Other secondary analysis
	Proportion of patients achieving a $SALT_{75}$	Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Other secondary analysis
	Proportion of patients achieving a $SALT_{30}$	Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis
	▪ SALT score ▪ Percent change from baseline in SALT score	ANCOVA using mLOCF[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Key secondary analysis
		ANCOVA using mLOCF[a]	mFAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		ANCOVA using mLOCF[a]+MI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		ANCOVA using mLOCF[b]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Sensitivity analysis
		ANCOVA using mLOCF[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 12, 16, and 24	Other secondary analysis
		ANCOVA using mLOCF[a]	FAS (Severe SALT subgroup[c]); FAS (Very severe SALT subgroup[c])	Bari 4-mg dose or Bari 2-mg dose vs PBO through Week 36	Dosing evaluation analysis

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
		ANCOVA using mLOCF[a]	FAS (Duration of current AA episode <4 years subgroup[d]); FAS (Duration of current AA episode >4 years subgroup[d])	Bari 4-mg dose or Bari 2-mg dose vs PBO through Week 36	Dosing evaluation analysis
	Change from baseline in SALT score	ANCOVA using mLOCF[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 12, 16, 24 and 36	Other secondary analysis
		ANCOVA using mLOCF[a]	FAS (Severe SALT subgroup[c]); FAS (Very severe SALT subgroup[c])	Bari 4-mg dose or Bari 2-mg dose vs PBO through Week 36	Dosing evaluation analysis
		ANCOVA using mLOCF[a]	FAS (Duration of current AA episode <4 years subgroup[d]); FAS (Duration of current AA episode >4 years subgroup[d])	Bari 4-mg dose or Bari 2-mg dose vs PBO through Week 36	Dosing evaluation analysis

Description of Primary, Secondary and Exploratory Efficacy Analyses

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
Patient-Reported Outcomes (PRO) for Scalp Hair Assessment	Proportion of patients with PRO for Scalp Hair Assessment score of 0 or 1 with a ≥2-point improvement from baseline	Logistic regression using NRI[a]	FAS (among patients with a PRO for Scalp Hair Assessment score of ≥3 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Key secondary analysis
		Logistic regression using NRI[a]	mFAS (among patients with a PRO for Scalp Hair Assessment score of ≥3 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic regression using MI[a]+NRI[a]	FAS (among patients with a PRO for Scalp Hair Assessment score of ≥3 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic regression using NRI[b]	FAS (among patients with a PRO for Scalp Hair Assessment score of ≥3 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Sensitivity analysis
		Logistic regression using NRI[a]	FAS (among patients with a PRO for Scalp Hair Assessment score of ≥3 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 12 and 24	Other secondary analysis

Description of Primary, Secondary and Exploratory Efficacy Analyses

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
PRO for Appearance of Eyebrows (EB)	Proportion of patients achieving PRO Measure for EB 0 or 1with ≥2-point improvement from baseline	Logistic regression using NRI[a]	FAS (among patients with PRO Measure for EB ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 16, 24, and 36	Other secondary analysis
PRO for Appearance of Eyelashes (EL)	Proportion of patients achieving PRO Measure for EL 0 or 1 with ≥2-point improvement from baseline	Logistic regression using NRI[a]	FAS (among patients with PRO Measure for EL ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 16, 24, and 36	Other secondary analysis
PRO for Eye Irritation (EI)	Proportion of patients achieving PRO Measure for EI 0 or 1 with ≥2-point improvement from baseline	Logistic regression using NRI[a]	FAS (among patients with PRO Measure for EI ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis

Description of Primary, Secondary and Exploratory Efficacy Analyses

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
PRO for Nail Appearance	Proportion of patients achieving PRO Measure for Nail Appearance 0 or 1 with ≥2-point improvement from baseline	Logistic regression using NRI[a]	FAS (among patients with PRO Measure for Nail Appearance ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis
Clinician-Reported Outcomes (ClinRO) for EB Hair Loss	Proportion of patients achieving ClinRO Measure for EB Hair Loss 0 or 1 with a ≥ 2-point improvement from baseline	Logistic regression using NRI[a]	FAS (among patients with ClinRO Measure for EB Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Key secondary analysis
		Logistic Regression using NRI[a]	mFAS (among patients with ClinRO Measure for EB Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic Regression using MI[a]+NRI[a]	FAS (among patients with ClinRO Measure for EB Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic Regression using NRI[b]	FAS (among patients with ClinRO Measure for EB Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Sensitivity analysis

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
		Logistic Regression using NRI[a]	FAS (among patients with ClinRO Measure for EB Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 16 and 24	Other secondary analysis
		Logistic Regression using NRI[a]	FAS (Severe/very severe SALT subgroups[c] among patients with ClinRO Measure for EB Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Dosing evaluation analysis
		Logistic Regression using NRI[a]	FAS (Duration of current AA episode <4 years/≥4 years subgroups[d] among patients with ClinRO Measure for EB Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Dosing evaluation analysis

Description of Primary, Secondary and Exploratory Efficacy Analyses

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
ClinRO for EL Hair Loss	Proportion of patients achieving ClinRO Measure for EL Hair Loss 0 or 1 with a ≥2-point improvement from baseline	Logistic regression using NRI[a]	FAS (among patients with ClinRO Measure for EL Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Key secondary analysis
		Logistic Regression using NRI[a]	mFAS (among patients with ClinRO Measure for EL Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic Regression using MI[a]+NRI[a]	FAS (among patients with ClinRO Measure for EL Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Supplementary analysis
		Logistic Regression using NRI[b]	FAS (among patients with ClinRO Measure for EL Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Sensitivity analysis
		Logistic Regression using NRI[a]	FAS (among patients with ClinRO Measure for EL Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 16 and 24	Other secondary analysis

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
		Logistic Regression using NRI[a]	FAS (Severe/very severe SALT subgroups[c] among patients with ClinRO Measure for EL Hair Loss ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Dosing evaluation analysis
		Logistic Regression using NRI[a]	FAS (Duration of current AA episode <4 years/≥4 years subgroups[d] among patients with ClinRO Measure for EL Hair Loss ⩾2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Week 36	Dosing evaluation analysis
ClinRO for Nail Appearance	Proportion of patients achieving ClinRO measure for Nail Appearance 0 or 1 with a ≥2-point improvement from baseline	Logistic regression using NRI[a]	FAS (among patients with ClinRO measure for Nail Appearance ≥2 at baseline)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis

Description of Primary, Secondary and Exploratory Efficacy Analyses

Abbreviations: AA = alopecia areata; ANCOVA = analysis of covariance; Bari = baricitinib; FAS = full analysis set; mFAS = modified full analysis set; MI = multiple imputation; mLOCF = modified last observation carried forward; NRI = nonresponder imputation; PBO = placebo; PPS = per-protocol set; $SALT_{30,50,75,90,100}$ = at least 30/50/75/90/100% improvement from baseline in SALT score.

[a] Primary censoring rule.

[b] Secondary censoring rule.

[c] Severe SALT subgroup = patients with SALT score of 50% to 94% at baseline; Very severe SALT subgroup = patients with SALT score of 95% to 100% at baseline.

[d] Duration of current AA episode <4 years subgroup = patients with duration of current AA episode at baseline < 4 years; Duration of current AA episode ≥4 years subgroup = patients with duration of current AA episode at baseline ≥ 4 years.

6.11.1. Primary Outcome and Methodology

The primary analysis of this study is to test the hypothesis that the 4-mg dose or 2-mg dose of baricitinib is superior to placebo in the treatment of patients with severe or very severe AA, as assessed by the proportion of patients achieving SALT ≤20 at Week 36 using the FAS population, assuming that treatment response disappears at the visits conducted remotely due to COVID-19 or after the patient discontinued study or study treatment. This will serve as the primary estimand. In this estimand, missing data due to the application of the primary censoring rule and the occurrence of other noncensor intercurrent events will be imputed using the NRI method described in Section 6.4.1.

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. In the case when Firth's correction still results in quasi-separation, Fisher's exact test will be used for the primary analysis.

6.11.2. Secondary and Exploratory Outcome Analyses

Multiplicity controlled analyses will be performed on the primary and key secondary (see Sections 4.1 and 4.2) objectives to control the overall family-wise Type I error rate at a 2-sided α level of 0.05. A graphical multiple testing procedure described in Bretz et al. (2011) will be used to perform the multiplicity controlled analyses as described in Section 6.6.

There will be no adjustment for multiple comparisons for any other analyses. The secondary and exploratory efficacy analyses are detailed in Table JAIR.6.6. Health outcomes/health-related quality-of-life analyses are described in Section 6.12.

6.11.3. Supplementary Analyses

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

- Analyses of key endpoints using the modified full analysis set (mFAS) (Section 6.2.1)
- Analyses of the primary endpoint using the PPS (Section 6.2.1)
- Hybrid imputation approach with NRI and MI for categorical variables, and mLOCF and MI for continuous variables (Section 6.4.3)
- Tipping point analysis (Section 6.4.4)

6.11.4. Dosing Evaluation Analyses

Additional analyses will be conducted within the following subgroups of the FAS population for the treatment dosing evaluation:

- SALT baseline severity subgroups: severe (SALT score of 50% to 94%) and very severe (SALT score of 95% to 100%)
- Duration of current AA episode at baseline subgroups: <4 years and ≥4 years.

The dosing analyses will be evaluated on the following endpoints:

- SALT ≤20 at Week 36;
- ClinRO Measure for EB Hair Loss score of 0 or 1 with ≥2-point improvement from baseline at Week 36 (among patients with ClinRO Measure for EB Hair Loss ≥2 at baseline);
- ClinRO Measure for EL Hair Loss score of 0 or 1 with ≥2-point improvement from baseline at Week 36 (among patients with ClinRO Measure for EL Hair Loss ≥2 at baseline);
- SALT change and percent change from baseline through Week 36.

The statistical analyses will follow the analysis methods specified in Section 6.2.3. For the categorical endpoints, the odds ratio with CI and corresponding p-value from the logistic regression model, percentages, difference in percentages, and CIs of the difference in percentages using the Newcombe-Wilson method without continuity correction will be reported. For the continuous endpoints, ANCOVA will be used. For the analyses performed on the subgroups defined by the duration of current AA episode at baseline (<4 years or ≥4 years), the covariate of duration of current episode at baseline will not be included in the model.

6.11.5. Analysis Beyond Week 36 Placebo-controlled Period

Statistical analysis beyond the Week-36 Placebo-controlled period will be used to support the long-term efficacy and safety assessment of the treatment. Since the long-term extension and bridging extension periods are not placebo-controlled, only descriptive statistics will be provided unless otherwise stated. Table JAIR.6.7 summarizes the analyses planned beyond week 36. Further details will be specified in a future version of the SAP.

Table JAIR.6.7. Description of Analysis Beyond Week 36 Placebo-controlled Period

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
Severity of Alopecia Tool (SALT)	Proportion of patients maintaining SALT ≤20	Descriptive	Randomized Downtitration Population	Summary statistics will be provided at each post-baseline visit during the Long-Term Extension and Bridging Long-Term Extension Period	Other Secondary
	Proportion of patients with >20-point absolute worsening in SALT score	Descriptive	Randomized Downtitration Population	Summary statistics will be provided at each post-baseline visit during the Long-Term Extension and Bridging Long-Term Extension Period	Other Secondary

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
	Time to >20-point absolute worsening in SALT score	Kaplan-Meier Plot	Randomized Downtitration Population	Summary statistics will be provided at each post-baseline visit during the Long-Term extension and Bridging Long-Term Extension Period.	Other secondary
	Proportion of patients achieving SALT ≤20	Descriptive	Retreated Population	Summary statistics will be provided at Weeks 12, 16, 24, and 36 of retreatment with baricitinib 4-mg	Other secondary
	Percent change in SALT score	Descriptive	Retreated Population	Summary statistics will be provided at Weeks 12, 16, 24, and 36 of retreatment with baricitinib 4-mg	Other secondary
PRO for Scalp Hair Assessment	Proportion of patients with a PRO for Scalp Hair Assessment score of 0 or 1	Descriptive	Retreated Population	Summary statistics will be provided at Weeks 12, 16, 24, and 36 weeks of retreatment with baricitinib 4-mg	Other secondary

Abbreviations: AA = alopecia areata.

6.12. Health Outcome/Health-related Quality-of-Life Analyses

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

Health outcomes and health-related quality-of-life measures will generally be analyzed according to the formats discussed in Section 6.11.

Table JAIR.6.8 includes the descriptions and derivations of the health outcomes and health-related quality-of-life measures.

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

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

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

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
Skindex-16 Adapted for Alopecia Areata (Skindex-16 AA)	Skindex-16 has been used to assess the health-related quality of life in patients with skin diseases. The Skindex-16 items' wordings were adapted for use among adults with AA. It examines the degree to which the subject is bothered by alopecia (hair loss) and associated symptoms. It is composed of 16 items grouped under 3 domains: Symptoms (4 items), Emotions (7 items), and Functioning (5 items). The score of each item ranges from 0 (never bothered) to 6 (always bothered).	▪ Skindex-16 AA score for symptoms, emotions, and functioning domains	Symptoms domain score is sum of 4 items, range 0 to 24; Emotions domain score is sum of 7 items, range 0 to 42; Functioning score is sum of 5 items, range 0 to 30.	N/A – partial assessments cannot be saved.
		▪ Change from baseline in Skindex-16 AA domain	Change from baseline: observed Skindex-16 AA domain score – baseline Skindex-16 AA domain score	Missing if baseline or observed value is missing
Medical Outcomes Study 36-Item Short-Form (SF-36) Health Survey Version 2 Acute	The SF-36 is a 36-item, patient-completed measure designed to be a short, multipurpose assessment of health (The SF Community – SF-36 Health Survey Update). The summary scores range from 0 to 100, with higher scores indicating better levels of function and/or better health. Items are answered on Likert scales of varying lengths. The SF-36 comprises 8 domain scores and 2 overarching component scores. SF-36 domain scores are: (1) Physical Functioning; (2) Role-Physical; (3) Role-Emotional; (4) Bodily Pain; (5) Vitality; (6) Social Functioning; (7) Mental	8 associated domain scores: • Physical Functioning • Role-Physical • Bodily Pain • General Health • Vitality • Social Functioning • Role-Emotional • Mental Health 2 component scores: • MCS score • PCS score	Per copyright owner, the QualityMetric Health Outcomes™ Scoring Software will be used to derive SF-36 domain and component scores. After data quality-controls, the SF-36 software will recalibrate the item-level responses for calculation of the domain and component scores. These raw scores will be transformed into the domain scores (t-scores) using the 1-week recall period. No missing imputation method will be used. Both, raw and domain scores without	Missing item-level data handling offered by SF-36. No missing-imputation

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
	Health; and (8) General Health. The component scores are: (1) Physical Component Summary (PCS); and (2) Mental Component Summary (MCS). The SF-36 acute version will be used, which has a 1-week recall period. Responder definitions were determined in the user's manual (Maruish 2011)		missing-data imputation will be recorded in the SDTM dataset; however, only the domain scores will be used for analyses specified in the SAP.	
		▪ Change from baseline in domain and component scores	Change from baseline: observed SF-36 score – baseline SF-36 score	Missing if baseline or observed value is missing
		▪ SF-36 Domain score Responder Definition	Domain score increase (change from baseline) (1) Physical Functioning >4.3; (2) Role-Physical >4.0; (3) Role-Emotional >4.6; (4) Bodily Pain >5.5; (5) Vitality >6.7; (6) Social Functioning >6.2; (7) Mental Health >6.7; (8) General Health >7.0	Missing if baseline or observed value is missing
		▪ SF-36 PCS Responder Definition	PCS component score increase (change from baseline) >3.8	Missing if baseline or observed value is missing
		▪ SF-36 MCS Responder Definition	MCS component score increase (change from baseline) >4.6	Missing if baseline or observed value is missing

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
Hospital Anxiety and 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 (for example, 0 to 3) for each question and is intended for ages 12 to 65 years (Zigmond and Snaith 1983; White et al. 1999). Scores for each domain (anxiety and depression) can range from 0 to 21, with higher scores indicating greater anxiety or depression (Zigmond and Snaith 1983; Snaith 2003).	▪ HADS score for anxiety and depression domains	Anxiety domain score is sum of the 7 anxiety questions, range 0 to 21. Depression domain score is sum of the 7 depression questions, range 0 to 21.	N/A – partial assessments cannot be saved
		▪ Change from baseline in HADS Anxiety and Depression domains	Change from baseline: observed HADS domain score – baseline HADS domain score	Missing if baseline or observed value is missing
		▪ Anxiety Domain Responder Definition	Anxiety domain score <8	Missing if observed value is missing
		▪ Depression Domain Responder Definition	Depression domain score <8	Missing if observed value is missing
European Quality of Life–5 Dimensions–5 Levels (EQ-5D-5L)	The EQ-5D-5L is a standardized measure of health status that provides a simple, generic measure of health for clinical and economic appraisal. The EQ-5D-5L consists of 2 components: a descriptive system of the respondent's health and a rating of his or her current health state using a 0 to 100-mm VAS. The descriptive system comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort,	▪ EQ-5D mobility ▪ EQ-5D self-care ▪ EQ-5D usual activities ▪ EQ-5D pain/discomfort ▪ EQ-5D anxiety/depression	5 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 through 5 have no arithmetic properties and should not be used as a primary score.	Each dimension is a single item, missing if missing

Measure	Description	Variable	Derivation / Comment	Imputation Approach if with Missing Components
	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 through 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 2019).	▪ EQ-5D VAS	Single item. Range 0 to 100. 0 represents "worst health you can imagine" 100 represents "best health you can imagine"	Single item, missing if missing
		▪ Change from baseline in EQ-5D VAS	Change from baseline: observed EQ-5D VAS score – baseline EQ-5D VAS score	Missing if baseline or observed value is missing
		▪ EQ-5D-5L 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
		▪ 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 US algorithm to produce a patient-level index score between -0.59 and 1.0 (continuous variable)	N/A-partial assessments cannot be saved on the eCOA tablet
		▪ Change from baseline in EQ-5D-5L UK population-based index score	Change from baseline: observed EQ-5D-5L UK score – baseline EQ-5D-5L UK score	Missing if baseline or observed value is missing

Abbreviations: AA = alopecia areata; eCOA = electronic Clinical Outcomes Assessment; N/A = not applicable; SDTM = study data tabulation model; VAS = visual analog scale.

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

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
Skindex-16 Adapted for Alopecia Areata (Skindex-16 AA)	▪ Skindex-16 Adapted for AA score for symptoms domain ▪ Change from baseline in Skindex-16 Adapted for AA score for symptoms domain	ANCOVA using mLOCF[a]	FAS (among patients with baseline assessment)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Other secondary/ exploratory analysis
	▪ Skindex-16 Adapted for AA score for emotions domain ▪ Change from baseline in Skindex-16 Adapted for AA score for emotions domain	ANCOVA using mLOCF[a]	FAS (among patients with baseline assessment)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Other secondary/ exploratory analysis
	▪ Skindex-16 Adapted for AA score for functioning domain ▪ Change from baseline in Skindex-16 Adapted for AA score for functioning domain	ANCOVA using mLOCF[a]	FAS (among patients with baseline assessment)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Other secondary/ exploratory analysis
Medical Outcomes Study 36-Item Short-Form (SF-36) Health Survey Version 2 Acute	▪ SF-36 score for 8 health domains, physical component score (PCS), and mental component score (MCS) ▪ Change from baseline in SF-36 score for 8 health domains ▪ Change from baseline in SF-36 score for 2 component scores	ANCOVA using mLOCF[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis
	▪ Proportion of patients achieving minimum clinically important difference (MCID) at each of 8 domain scores	Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
	▪ Proportion of patients achieving minimum clinically important difference (MCID) at each of 2 component scores	Logistic regression using NRI[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis

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

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
Hospital Anxiety and Depression Scale (HADS)	▪ HADS score for 2 domains ▪ Change from baseline in HADS score for anxiety domain ▪ Change from baseline in HADS score for depression domain	ANCOVA using mLOCF[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Other secondary analysis
	▪ Proportion of patients achieving HADS score for depression domain <8	Logistic regression using NRI[a]	FAS (Among patients with baseline HADS depression total score ≥8)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis
	▪ Proportion of patients achieving HADS score for anxiety domain <8	Logistic regression using NRI[a]	FAS (Among patients with baseline HADS anxiety total score ≥8)	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis
European Quality of Life–5 Dimensions–5 Levels (EQ-5D-5L)	▪ EQ-5D VAS ▪ Change from baseline in EQ-5D VAS	ANCOVA using mLOCF[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis
	▪ EQ-5D-5L US Population-based index score (Health state index) ▪ Change from baseline in EQ-5D-5L US Population-based index score	ANCOVA using mLOCF[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis

Measure	Variable	Analysis Method (Section 6.2.3)	Population (Section 6.2.1)	Comparison/Time Point	Analysis Type
	▪ EQ-5D-5L UK Population-based index score (Health state index) ▪ Change from baseline in EQ-5D-5L UK Population-based index score	ANCOVA using mLOCF[a]	FAS	Bari 4-mg dose or Bari 2-mg dose vs PBO at Weeks 24 and 36	Exploratory analysis

Abbreviations: AA = alopecia areata; ANCOVA = analysis of covariance; Bari = baricitinib; FAS = full analysis set; mLOCF = modified last observation carried forward; NRI = non responder imputation; PBO = placebo; VAS = visual analog scale..

[a] Primary Censoring Rule.

6.13. Safety Analyses

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

Safety analyses will include data from first dose of the study treatment including follow-up data, where applicable. Patients will be analyzed according to the IP to which they were randomized at Week 0 (Visit 2), unless otherwise stated. Safety analyses will take place using the safety population defined in Section 6.2.1.

Safety topics that will be addressed include the following: AEs (includes treatment-emergent adverse event [TEAEs] and serious adverse events [SAEs]), clinical laboratory evaluations, vital signs and physical characteristics, Columbia Suicide Severity Rating Scale (C-SSRS), the Self-Harm Supplement Form, safety in special groups and circumstances including adverse events of special interest (AESI) (see Section 6.13.5), and IP interruptions.

Unless otherwise specified, by-visit summaries will include planned on-treatment visits. For tables that summarize events (such as AEs, categorical treatment-emergent lab abnormalities, shift to minimum or maximum value in labs, and vital signs), post last dose follow-up data will be included. Follow-up data is defined as all data occurring up to a minimum of 30 days (planned maximum follow-up time) or data cut date, whichever occurs first after last dose of treatment, where applicable.

For the interim lock(s), all safety data from ongoing patients at time of the interim lock (all data up to the data cut date) will be included in the safety analysis censored at treatment change (including rescue to a higher dose), unless otherwise stated. Safety data from patients who permanently discontinued the study treatment prior to an interim lock will be included in the interim lock safety analysis up to 30 days post last dose, censored at treatment change, unless otherwise stated.

For the Weeks 0 to 36 tables, figures, and listings (TFLs) summarizing events in a nonvisit-specific manner, including:

- AEs
- C-SSRS
- shift in laboratory testing
- treatment-emergent abnormal laboratory testing
- treatment-emergent abnormal vital signs,

the analysis period is defined as first dose date up to min (last dose date+30 days, Week 36 visit date, study disposition date).

For the Weeks 0 to 36 TFLs summarizing events in a by-visit manner, including:

- Observed and change in laboratory testing at scheduled visit
- Observed and change in vital signs at scheduled visit,

the analysis period is defined as first dose date up to min (last dose date, Week 36 visit date, study disposition date). The Week 36 visit date will be imputed if it is missing.

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

Refer to the compound level safety standards for more details.

6.13.1. Extent of Exposure

Duration of exposure (in weeks) to study drug will be summarized for the safety population by treatment group using descriptive statistics. Cumulative exposure and duration of exposure will be summarized in terms of frequency counts and percentages by category and treatment group.

Duration of exposure will be calculated as follows, unless otherwise stated:

- duration of exposure to IP excluding exposure post treatment change or rescue to baricitinib: *date of last dose of study drug–date of first dose of study drug +1*.

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

Total patient-years (PY) of exposure will be reported for each treatment group for overall duration of exposure. Descriptive statistics will be provided for patient-weeks of exposure and the frequency of patients falling into different exposure ranges will be summarized. Exposure ranges will generally be reported in weeks using the following as a general guide and may be adjusted based on exposure time at the interim locks:

- ≥4 weeks, ≥8 weeks, ≥12 weeks, ≥16 weeks, ≥24 weeks, ≥36 weeks, ≥52 weeks, ≥76 weeks, and ≥104 weeks
- >0 to <4 weeks, ≥4 weeks to <8 weeks, ≥8 weeks to <12 weeks, ≥12 to <16 weeks, ≥16 to 24 weeks, ≥24 to <36 weeks, ≥36 to <52 weeks, ≥52 to <76 weeks, ≥76 to <104 weeks, and ≥104 weeks

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

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

6.13.2. Adverse Events

Adverse events are recorded in the eCRFs. Each AE will be coded to system organ class (SOC) and PT using the MedDRA version that is current at the time of database lock. Severity of AEs is recorded as mild, moderate, or severe.

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

time. For the Weeks 0 to 36 TFLs, the analysis period is defined as first dose date up to min (last dose date + 30 days, Week 36 visit date, study disposition date). The Week 36 visit date will be imputed if it is missing.

Refer to the compound level safety standards for more details including data imputations.

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

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 in 3 formats:

- by MedDRA PT nested within SOC with decreasing frequency in SOC, and events ordered within each SOC by decreasing frequency in the baricitinib 4-mg dose group.
- by MedDRA PT with events ordered by decreasing frequency in the baricitinib 4-mg dose group.
- by maximum severity by treatment using MedDRA PT ordered by decreasing frequency in the baricitinib 4-mg dose group. For each patient and TEAE, the maximum severity for the MedDRA level being displayed is the maximum postbaseline severity observed from all associated lowest level terms (LLTs) mapping to that MedDRA PT.

6.13.2.1. Common Adverse Events

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

6.13.2.2. 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), a SAE is any AE that results in any 1 of the following outcomes:

- death
- initial or prolonged inpatient hospitalization
- life-threating experience (that is, immediate risk of dying)
- persistent or significant disability/incapacity
- congenital anomaly/birth defect
- important medical events that may not be immediately life-threatening or result in death or hospitalization but may jeopardize the patient or may require intervention to prevent 1

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 using MedDRA PT nested within SOC. Events will be ordered by decreasing frequency in the baricitinib 4-mg dose group within decreasing frequency in SOC. The SAEs will also be summarized by treatment using MedDRA PT without SOC.

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

6.13.2.3. Other Significant Adverse Events

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

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

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

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

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

6.13.2.4. Criteria for Notable Patients

Patient narratives will be provided for all patients who experience “notable” events. See the compound level safety standards for list of criteria.

6.13.3. Clinical Laboratory Evaluation

For the categorical laboratory analyses (shift and treatment emergent), the analysis period is defined as the treatment period plus up to 30 days off-drug including follow-up time. The analysis period for the continuous laboratory and visit-specific analyses is defined as the treatment period excluding off-drug follow-up time. See Section 6.13 for a detailed definition of analysis period.

Refer to the compound level safety standards for the details pertaining to box plots and treatment-emergent low and high abnormalities.

6.13.4. Vital Signs and Other Physical Findings

For the treatment-emergent categorical analyses (shift and treatment-emergent), the analysis period is defined as the treatment period plus up to 30 days off-drug including follow-up time. The analysis period for the continuous analyses (for example, change from baseline by time point) is defined as the treatment period excluding off-drug including follow-up time. For the Weeks 0 to 36 TFLs, the analysis period is defined in the same way as Section 6.13.

Refer to the compound level safety standards for the details.

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

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

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

6.13.5.1. Abnormal Hepatic Tests

Analyses for abnormal hepatic tests will involve 4 laboratory analytes: alanine aminotransferase, aspartate transaminase, total bilirubin, and alkaline phosphatase. Refer to the compound level safety standards for more details.

6.13.5.2. Hematologic Changes

Hematologic changes will be defined based on clinical laboratory assessments. Refer to the compound level safety standards for the details.

6.13.5.3. Lipids Effects

Lipids effects will be assessed through analysis of elevated total cholesterol, elevated low-density lipoproteins cholesterol, decreased and increased high-density lipoproteins cholesterol, elevated triglycerides, and with TEAEs potentially related to hyperlipidemia. Refer to the compound level safety standards for the details.

6.13.5.4. Renal Function Effects

Effects on renal function will be assessed through analysis of elevated creatinine using common terminology criteria for AEs (CTCAE). Refer to the compound level safety standards for the details.

6.13.5.5. Elevations in Creatine Phosphokinase (CPK)

Elevations in CPK will be addressed using CTCAE criteria and TEAEs potentially related to muscle symptoms will be analyzed, based on reported AEs. Refer to the compound level safety standards for the details.

6.13.5.6. Infections

Refer to the compound level safety standards.

Potential opportunistic infection

Refer to the compound level safety standards.

Herpes zoster

Refer to the compound level safety standards.

Herpes simplex

Refer to the compound level safety standards.

Hepatitis B Virus DNA

Refer to the compound level safety standards.

6.13.5.7. Major Adverse Cardiovascular Events and Other Cardiovascular Events

Refer to the compound level safety standards.

6.13.5.8. Venous Thromboembolic Events

Refer to the compound level safety standards.

6.13.5.9. Arterial Thromboembolic Events

Refer to the compound level safety standards.

6.13.5.10. Malignancies

Refer to the compound level safety standards.

6.13.5.11. Allergic Reactions/Hypersensitivities

Refer to the compound level safety standards.

6.13.5.12. Gastrointestinal Perforations

Refer to the compound level safety standards.

6.13.5.13. Columbia Suicide Severity Rating Scale

Refer to the compound level safety standards.

6.13.5.13.1. Self-Harm Supplement Form and Self-Harm Follow-up Form

The Self-Harm Supplement Form is a single question to enter the number of suicidal behavior events, possible suicide behaviors, or nonsuicidal self-injurious behaviors. If the number of behavioral events is greater than 0, it will lead to the completion of the Self-Harm Follow-Up Form. The Self-Harm Follow-Up Form is a series of questions that provides a more detailed description of the behavior cases. A listing of the responses given on the Self-Harm Follow-Up Form will be provided.

6.14. Subgroup Analyses

Subgroup analyses comparing each dose of baricitinib to placebo will be performed on the FAS population at Week 36 for the following:

- proportion of patients achieving SALT ≤20.

The following subgroups (but may not be limited to only these) will be categorized into disease-related characteristics and demographic characteristics and will be evaluated:

- patient demographic and characteristics subgroups:
  - Genetic gender (male versus female)
  - Geographic region (North America, Asia, and Rest of World)
  - Age group (<40 versus ≥40 years old)
  - Age group (<65 versus ≥65 years old)
  - Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Multiple)
  - Weight group (<60 kg, ≥60 to <100 kg, ≥100 kg)

  - BMI group (<25 kg/m^2, ≥25 to <30 kg/m^2, ≥30 kg/m^2)
  - Screening period renal function status: impaired (eGFR <60 mL/min/1.73 m^2) or not impaired (eGFR ≥60 mL/min/1.73 m^2)
- Baseline Disease-Related Characteristics Subgroups:
  - SALT baseline severity category (Severe (SALT score of 50% to 94%) vs very severe (SALT score of 95% to 100%))
  - Duration of current episode of AA category (< 4 years vs ≥4 years)

Descriptive statistics will be provided for each treatment and stratum of a subgroup as outlined, regardless of sample size. The 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, stratification variables, treatment, subgroup, and treatment-by-subgroup interaction as explanatory variables. Note that, when the subgroup variable is SALT baseline severity category, the SALT baseline value will not be included as a covariate in the model. Missing data will be imputed using NRI using the primary censoring rule (Section 6.4.1). The treatment-by-subgroup interaction will be tested at the 0.1 significance level. The p-value from the logistic regression model will be reported for the interaction test and the subgroup test, unless the model did not converge. Response counts and percentages will be summarized by treatment for each subgroup category. The difference in percentages and 100(1-α)% CI of the difference in percentages using the Newcombe-Wilson method without continuity correction will be reported. The p-value from the Fisher's exact test will also be produced.

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

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

6.15. Analysis for Japan Submission

A subset of the planned efficacy, health outcomes ,and safety analyses will be reproduced based on patients from Japan sites, in support of the regulatory submission in Japan. The list of tables, listings, and figures for the patients from Japan sites (Japanese population) will be in a separate document.

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% or ≥120% of assigned doses taken, failure to take study medication, and taking incorrect study medication), patients incorrectly

enrolled in the study, and patients whose data are questionable due to significant site quality or compliance issues. Refer to a separate document for the important protocol deviations.

The Trial Issue Management Plan includes the categories and subcategories of important protocol deviations and whether or not these deviations will result in the exclusion of patients from the PPS.

The number and percentage of patients having important protocol deviations will be summarized within category and subcategory of deviation by treatment group. The summary will be presented for the FAS population. Individual patient listings of important protocol deviations will be provided. A summary of reasons patients were excluded from the PPS will be provided by treatment group.

6.17. Interim Analyses and Data Monitoring

6.17.1. 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. 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 final database lock (F-DBL), including study discontinuation data, AEs/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.

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
- C-SSRS

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

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.

The DMC is authorized to evaluate unblinded interim efficacy and safety analyses during the study. Further details of the DMC will be documented in the DMC charter. Study sites will receive information about interim results if they need to know about a dose change or the safety of their patients. Unblinding details will be specified in a separate unblinding plan document.

6.17.2. Other Interim Analyses

6.17.2.1. Week 36 Primary Outcome Analysis and other regulatory submission activities

- After all randomized patients complete the primary efficacy assessment at Week 36 (Visit 8) or discontinue early, the database will be locked and data will be unblinded to a limited number of preidentified individuals to initiate work for submission. Although it is called an interim analysis, the PO-DBL interim analysis is the only and final analysis for the primary endpoint. Therefore, no α adjustment for this interim analysis is planned. 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.
- Another interim analysis will occur for the 4-month safety update database lock.
- Additional efficacy or safety interim analyses prior to the F-DBL may occur to support regulatory submissions and scientific disclosures.

If an unplanned interim analysis is deemed necessary, the appropriate Lilly medical director or designee will be consulted to determine whether it is necessary to amend the protocol.

6.17.3. Adjudication Committee

A blinded Clinical Event Committee will adjudicate potential major adverse cardiovascular events (MACEs; cardiovascular death, myocardial infarction, stroke), other cardiovascular events (such as hospitalization for unstable angina, hospitalization for heart failure, serious arrhythmia, resuscitated sudden death, cardiogenic shock, coronary revascularization [e.g., coronary artery bypass graft or percutaneous coronary intervention]), venous and arterial thrombotic events, and noncardiovascular deaths. Details of membership, operations, recommendations from the Committee, and the communication plan will be documented in the 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 be produced, will also be documented in the supplemental SAP.

6.19. Annual Report Analyses

Annual report analyses, such as the Development Safety Update 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 and by MedDRA PT:

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

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

7. Unblinding Plan

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

8. References

Alosh M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med.* 2014;33(4):693-713. https://doi.org/10.1002/sim.5974

Bretz F, Posch M, Glimm E, et al. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J.* 2011;53(6):894-913. https://doi.org/10.1002/bimj.20100239

EuroQol Group. EQ-5D-5L user guide. Version 3.0. Available at: https://euroqol.org/wp-content/uploads/2019/09/EQ-5D-5L-English-User-Guide_version-3.0-Sept-2019-secured.pdf. Published September 2019. Accessed February 2020.

Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-1736. https://doi.org/10.1007/s11136-011-9903-x

[ICH]. Harmonised tripartite guideline: clinical safety data management: definitions and standards for expedited reporting: E2A: current step 4 version. Available at: https://database.ich.org/sites/default/files/E2A_Guideline.pdf. Published 27 October 1994. Accessed February 2020.

[ICH]. 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. Published 20 November 2019. Accessed February 15, 2020.

Kennedy Crispin M, Ko JM, Craiglow BG, et al. Safety and efficacy of the JAK inhibitor tofacitinib citrate in patients with alopecia areata. *JCI Insight*. 2016;1(15):e89776. https://doi.org/10.1172/jci.insight.89776

Mackay-Wiggan J, Jabbari A, Nguyen N, et al. Oral ruxolitinib induces hair regrowth in patients with moderate-to-severe alopecia areata. *JCI Insight*. 2016;1(15):e89790. https://doi.org/10.1172/jci.insight.89790

Olsen E, Hordinsky M, McDonald-Hull S, et al. Alopecia areata investigational assessment guidelines. National Alopecia Areata Foundation. *J Am Acad Dermatol.* 1999;40(2 Pt 1):242-246. https://doi.org/10.1016/s0190-9622(99)70195-7

Olsen EA, Hordinsky MK, Price VH, et al. Alopecia areata investigational assessment guidelines--part II. National Alopecia Areata Foundation. *J Am Acad Dermatol.* 2004;51(3):440-447. https://doi.org/10.1016/j.jaad.2003.09.032

Snaith RP. The hospital anxiety and depression scale. *Health Qual Life Outcomes*. 2003;1:29. https://doi.org/10.1186/1477-7525-1-29

US National Archives and Records Administration. Code of Federal Regulations (CFR). investigational new drug application (IND) safety reporting: title 21: section 312.32. Available

at: https://www.ecfr.gov/cgi-bin/text-idx?SID=907b43ab06a92594142732f57207daf5&mc=true&node=se21.5.312_132&rgn=div8 . Published 29 September 2010. Accessed July 21, 2017.

White D, Leach C, Sims R, et al. Validation of the Hospital Anxiety and Depression Scale for use with adolescents. *Br J Psychiatry*. 1999;175:452-454. https://doi.org/10.1192/bjp.175.5.452

Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67(6):361-370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x