

**Statistical Analysis Plan**

BRAVE-AA1 and BRAVE-AA2 Phase 3 Trials: Baricitinib for Severe Alopecia Areata

Jason Dagher (JD4141), Adeena Moghni (am6592), Shehzrin Shah (ss6985), Puyuan Zhang,

Longyi Zhao (lz2949), Yixin Zheng (yz4993)

Columbia University Mailman School of Public Health

P8144 Pharmaceutical Statistics

Instructor: Naitee Ting

May 6, 2025

## TABLE OF CONTENTS

<b>TABLE OF CONTENTS</b>	<b>2</b>
<b>1. INTRODUCTION</b>	<b>3</b>
1.1. Background	3
1.2. Study Objective	3
Primary Objective	3
Secondary Objectives	3
Safety Objective	3
1.3. Eligibility Criteria	3
Inclusion Criteria	3
Exclusion Criteria	4
1.4. Study Design	5
1.4.1. Sample Size Estimation	5
<b>2. ENDPOINTS AND COVARIATES</b>	<b>6</b>
2.1. Efficacy Endpoints	6
2.2. Safety Endpoints	7
2.3. Covariates	7
<b>3. HYPOTHESIS AND DECISION RULES</b>	<b>7</b>
3.1. Statistical Hypothesis	7
3.2. Statistical Decision Rules	8
<b>4. ANALYSIS SETS</b>	<b>8</b>
4.1. Full Analysis Sets	8
4.2. Safety Analysis Sets	8
<b>5. HANDLING OF MISSING VALUES</b>	<b>8</b>
<b>6. STATISTICAL METHODOLOGY AND ANALYSES</b>	<b>8</b>
6.1. Statistical Methods	8
6.2. Statistical Analysis	9
6.2.1. Primary Analysis	9
6.2.2. Secondary Analysis	9
6.2.3. Safety Analysis	10
<b>7. REFERENCES</b>	<b>10</b>
<b>8. APPENDICES</b>	<b>11</b>
APPENDICES A: Team Member Contribution	11

## 1. INTRODUCTION

### 1.1. Background

Alopecia areata (AA) is a chronic, immune-mediated disorder characterized by non-scarring hair loss, most commonly affecting the scalp (Lepe & Zito, 2024). With an estimated prevalence of approximately 0.27% in China alone, AA affects millions of people globally and is associated with considerable psychosocial burden, particularly among adolescents and young adults with severe disease (Wang et al., 2023).

The pathogenesis of AA involves autoimmune targeting of hair follicles, with cytokine signaling, through the Janus kinase (JAK) pathways JAK1 and JAK2 particularly. Here, JAK pathways play a central role in disease progression (Yamaoka et al., 2004).

Baricitinib is an oral selective JAK1/JAK2 inhibitor. It has demonstrated efficacy in 2 large, randomized Phase 3 trials—BRAVE-AA1 and BRAVE-AA2—by significantly increasing the proportion of patients achieving meaningful scalp hair regrowth (SALT  $\leq 20$ ) after 36 weeks of treatment (King et al., 2022).

Since its approval for AA treatment in several countries including China, baricitinib has represented a shift in the treatment paradigm, especially for patients who previously relied on systemic corticosteroids, which cause long-term safety concerns (Wang et al., 2023). The emergence of JAK inhibitors offers a great alternative due to their targeted mechanism, improved tolerability, and increasing physician preference.

### 1.2. Study Objective

#### *Primary Objective*

To investigate the efficacy of once-daily oral baricitinib (4 mg and 2 mg) compared to placebo in achieving clinically meaningful scalp hair regrowth in adults with severe alopecia areata after 36 weeks.

#### *Secondary Objectives*

To compare baricitinib and placebo regarding the effect on:

- Patient-reported improvement in scalp hair appearance
- Clinician-rated regrowth of eyebrow and eyelash hair
- Changes in scalp hair loss severity, measured by SALT score percentage change from baseline

#### *Safety Objective*

To evaluate the safety and tolerability of baricitinib by comparing the incidence and severity of adverse events, discontinuations, and relevant lab/vital sign changes over 36 weeks of treatment

### 1.3. Eligibility Criteria

#### ***Inclusion Criteria***

Participants must be adults: males aged 18 to 60 years and females up to 70 years. Eligible individuals must have severe or very severe alopecia areata, defined as:

- A current episode lasting more than 6 months,
- $\geq 50\%$  scalp hair loss, measured by the SALT score at both screening visits,
- No significant spontaneous improvement (defined as  $\leq 10$ -point SALT score reduction) in the past 6 months.

The current episode must be less than 8 years in duration. However, individuals with disease lasting  $\geq 8$  years may still be eligible if they have shown episodes of regrowth, either spontaneously or with treatment, within that period.

Participants must agree not to use any treatments for alopecia areata during the study period. Prohibited treatments include:

- Systemic therapies (e.g., corticosteroids, methotrexate, cyclosporine, JAK inhibitors, apremilast)
- Biologics (e.g., monoclonal antibodies)
- Intralesional corticosteroids
- Topical irritants or immunotherapies (e.g., DPCP)
- Phototherapy (including laser)
- Platelet-rich plasma injections
- Cryotherapy
- statins used for alopecia

Exceptions include:

- Bimatoprost for eyelashes, if on a stable dose for  $\geq 8$  weeks prior to randomization,
- Finasteride or minoxidil (oral or topical), if on a stable dose for  $\geq 12$  months and expected to continue through Week 36.

#### ***Exclusion Criteria***

Participants will be excluded if they have:

- Diffuse-type alopecia areata (with generalized shedding)
- Other hair loss conditions such as androgenetic alopecia, telogen effluvium, trichotillomania, chemotherapy-induced alopecia, or scalp conditions (e.g., psoriasis, tinea capitis) that could interfere with assessment.

Additionally, individuals with unstable medical conditions requiring frequent hospitalization or ongoing systemic immunosuppressants may be excluded at the investigator's discretion.

Participants are also excluded if they received the following treatments within the specified time windows before randomization:

- Topical corticosteroids (scalp or eyebrows): within 1 week
- Systemic or intralesional corticosteroids: within 8 weeks
- JAK inhibitors (oral or topical): oral within 8 weeks, topical within 4 weeks
- Topical immunotherapies: within 4 weeks

## Statistical Analysis Plan

- Monoclonal antibodies: within 5 half-lives
- Platelet-rich plasma, phototherapy, cryotherapy, or statins for AA: within 4–8 weeks
- Finasteride/minoxidil: unless stable for  $\geq 12$  months
- Other immunosuppressants or targeted therapies: within 4–8 weeks, depending on the agent.

### 1.4. Study Design

This plan outlines the proposed analyses for a randomized, double-blind, placebo-controlled, parallel-group Phase 3 trial. Patients are to be randomized in a 3:2:2 ratio to receive either 4 mg baricitinib, 2 mg baricitinib, or placebo once daily for 36 weeks.

Stratified block randomization will be implemented using an Interactive Web Response System (IWRS) (King et al., 2022). Randomization will be stratified by:

- Geographic region (North America vs. Rest of World), and
- Duration of current alopecia areata episode at baseline:
  - $< 4$  years (with SALT score 50–94%), vs.
  - $\geq 4$  years (or very severe AA with SALT score 95–100%).

Participants and study personnel will remain blinded to treatment allocation throughout the study. Eligibility criteria will include patients aged 18–70 years with a clinical diagnosis of alopecia areata, a baseline SALT score of  $\geq 50$ , and no prior use of JAK inhibitors. Study medication will be administered daily, and compliance will be monitored through pill counts and patient self-report logs. All statistical analyses will be performed after the study is completed and the database has been locked to preserve the integrity of the blinded trial.

#### 1.4.1. Sample Size Estimation

According to the original JAHO clinical study protocol, ~1035 patients were screened to enroll ~725 patients across Stage 1 and Stage 2 (King et al., 2022). Stage 1 was a dose-finding phase that has up to 300 patients randomized: the first ~100 in a 1:1:1:1 ratio (placebo, 1 mg, 2 mg, 4 mg), and the next ~200 in a 2:2:3 ratio (2 mg, 4 mg, placebo). Stage 2 began after the interim analysis, with ~425 patients randomized in a 2:2:3 ratio to baricitinib 2 mg, 4 mg, or placebo (King et al., 2022). So in total, approximately 625 patients (Stage 2 + non-1 mg arms from Stage 1) are included in the primary efficacy analysis.

Both BRAVE-AA1 and BRAVE-AA2 trials were planned to achieve 90% power to test the superiority of each baricitinib dose compared to placebo for the primary endpoint, the proportion of patients achieving a SALT score  $\leq 20$  at Week 36, under the assumptions of response rates of 30% for 4 mg, 20% for 2 mg, and 5% for placebo.

The required per-group sample size was calculated using the standard formula for comparing two proportions via a Chi-square test without continuity correction:

$$n = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 \times [p_1(1-p_1) + p_0(1-p_0)]}{(p_1 - p_0)^2}$$

## Statistical Analysis Plan

Where:

- Placebo response rate ( $p_0 = 0.05$ )
- Expected response rate for baricitinib 4 mg ( $p_1 = 0.30$ )
- Expected response rate for baricitinib 2 mg ( $p_1 = 0.20$ )
- Standard normal value for  $\alpha = 0.05$ :  $Z_{1-\alpha/2} = 1.96$
- Standard normal value for  $\beta = 0.10$ :  $Z_{1-\beta} = 1.28$

Using  $p_1 = 0.30$  for the 4 mg vs. placebo comparison:

$$n = \frac{(1.96 + 1.28)^2 \times [0.21 + 0.0475]}{(0.25)^2} = \frac{10.5 \times 0.2575}{0.0625} \approx 43$$

To maintain the 2:2:3 randomization ratio (2 mg : 4 mg : placebo), the total sample size per comparison set becomes:

$$43 \times \frac{7}{2} \approx 150 \text{ patients}$$

Since the trial is designed to test two independent comparisons (2 mg vs. placebo and 4 mg vs. placebo), the total required sample size is:

$$150 \times 2 = 300 \text{ patients}$$

To account for attrition, noncompliance, and missing data (especially in the context of the COVID-19 pandemic), a 25% inflation factor is applied:

$$300 \div 0.75 = 400 \text{ patients}$$

Adding ~200 eligible patients from Stage 1, the total planned sample size for BRAVE-AA1 becomes approximately 600. BRAVE-AA2 followed the same assumptions and power requirements.

Sample size calculations were conducted using EAST v6.4 in protocol guideline, so minor deviations occur when performing manual calculations; therefore, for consistency and transparency, we follow the final reported sample sizes in the published paper: 625 for BRAVE-AA1 and 476 for BRAVE-AA2 (King et al., 2022).

## 2. ENDPOINTS AND COVARIATES

### 2.1. Efficacy Endpoints

The primary efficacy endpoint of this trial is the proportion of patients with a Severity of Alopecia Tool (SALT) score of 20 or less at week 36. A SALT score  $\leq 20$  corresponds to 20% or less scalp hair loss, which has been recognized as clinically meaningful threshold and used on prior clinical research (Atanaskova et al., 2024).

The secondary efficacy endpoints include the following:

(1) A Scalp Hair Assessment Patient-Reported Outcome ( $PRO^{TM}$ ) score of 0 or 1 with a decrease of at least 2 points from baseline at week 36;

(2) A clinician-reported (*ClinRO*<sup>TM</sup>) Measure for Eyebrow Hair Loss score of 0 or 1 with a decrease of at least 2 points from baseline at week 36;

(3) The percentage change from baseline in the SALT score at week 36. These secondary endpoints help complement the primary endpoints by incorporating both patients and clinician perspective on hair regrowth and providing a more comprehensive evaluation of the treatment.

### 2.2. Safety Endpoints

Safety assessments will include the monitoring and documentation of any adverse events (AEs) reported by participants throughout the 36-week treatment period. Those adverse events may include, but are not limited to, death, initial or extended hospitalization, life-threatening experiences, persistent or significant disability, or congenital anomalies and birth defects. Investigators will collect those information on the incidence, severity, and relationship of these events to the study drug.

In addition to self-reported adverse events, at each study visit, safety evaluations will be conducted by clinical investigators, which include measurement of vital signs and clinical laboratory testing, such as liver enzymes, lipid panels, and creatine kinase levels, to help detect potential adverse physiological effects. Such safety analysis will include all randomized participants who received at least one dose of study medication and did not discontinue due to loss to follow-up.

### 2.3. Covariates

Several variables will be included as covariates in the statistical analysis to adjust for potential confounding factors and improve the precision of treatment effect estimation. These include geographic region, treatment group, baseline SALT score, age group, body mass index (BMI), and duration since the onset of current episode of alopecia areata. Those covariates were selected based on their clinical relevance. For example, age and BMI are commonly associated with variability in autoimmune disease progression and thus may impact the treatment outcomes (Lepe et al., 2024). The inclusion of baseline SALT score helps account for the initial severity of disease, ensuring fair comparison across treatment groups. Adjusting for those factors in analyses can help minimize bias and enhance the interpretability of results.

## 3. HYPOTHESIS AND DECISION RULES

### 3.1. Statistical Hypothesis

$H_0$ : There is no difference in the proportion of patients achieving a SALT score  $\leq 20$  at week 36 between each baricitinib dose group (2 mg or 4 mg) and the placebo group.

$H_1$ : A higher proportion of patients in the baricitinib treatment groups will achieve a SALT score  $\leq 20$  at week 36 compared to placebo.

More specifically, let  $p_0$  be the proportion of patients achieving SALT  $\leq 20$  in the placebo group, and  $p_1$ ,  $p_2$  be the proportions in the 2 mg and 4 mg baricitinib groups, respectively. We test each dose group vs. placebo separately:

- $H_{0_1} : p_1 = p_0$  vs.  $H_{1_1} : p_1 > p_0$
- $H_{0_2} : p_2 = p_0$  vs.  $H_{1_2} : p_2 > p_0$

### 3.2. Statistical Decision Rules

The primary hypotheses will be tested at a two-sided significance level of 0.05. The primary endpoint will be assessed using a logistic regression model adjusting for treatment group, geographic region, baseline SALT score, and duration of the current alopecia areata episode (<4 years vs.  $\geq 4$  years). A two-sided p-value <0.05 will be considered statistically significant.

To control the familywise Type I error rate across the primary and key secondary endpoints, a graphical multiple testing procedure will be implemented. This approach allows for flexible allocation and reallocation of alpha across multiple hypotheses. Each hypothesis is assigned a portion of the overall alpha level (e.g., 0.05) and represented as a node in a graph. When a hypothesis is tested and found to be significant, its allocated alpha can be redistributed to other hypotheses according to predefined weights and edges. This method ensures that the overall Type I error rate does not exceed the pre-specified 5% threshold while preserving power for clinically important comparisons.

## 4. ANALYSIS SETS

### 4.1. Full Analysis Sets

All subjects randomised to initial treatment who were exposed to Baricitinib 2mg or 4mg are included in the full analysis set (FAS) and will be analysed for efficacy up to Week 36. For subjects not exposed to Baricitinib, the decision to withdraw can't be biased by knowledge of the assigned treatment due to the blinding. This definition of the FAS implements the consideration mentioned in the protocol regarding special excluded cases with reference to ICH E9, Section 5.2.1. For analysis of efficacy subjects will be included 'as randomised', regardless of treatment regimen followed, also known as Intent-to-Treat (ITT) analysis. The analysis of all endpoints, unless otherwise noted, will be conducted using ITT.

### 4.2. Safety Analysis Sets

Safety analysis set comprises all subjects randomised to initial treatment who were exposed to Baricitinib or placebo. The protocol further specifies to exclude subjects from the safety analysis set for whom no post-baseline safety data are available. Initial safety analysis sets will be conducted using ITT and those will be the ones presented in the final manuscript.

To address clinician concerns about drug safety, another analysis of safety using per-protocol (PP) methods will be conducted. If a subject is mistakenly given a therapy other than that to which they were randomized, they should be analyzed 'as-treated' in this analysis set, thus included in the group according to the therapy actually received by the subject. Subjects who received at least one dose of Baricitinib during the initial treatment period will be analysed in the baricitinib treatment group. All other subjects will be analyzed as placebo. Although this may dilute the AE rate in the Baricitinib treatment group slightly by including in the denominator

subjects who only received one dose of active treatment, it will ensure that no significant drug reactions to Baricitinib will erroneously be assigned to placebo.

### 5. HANDLING OF MISSING VALUES

Missing data for categorical outcomes will be handled using multiple imputation with 100 datasets. The imputation model will include treatment group, baseline covariates, and previous post-baseline observations. For continuous outcomes, sensitivity analyses will be performed using both multiple imputation and a modified last-observation-carried-forward (mLOCF) approach.

### 6. STATISTICAL METHODOLOGY AND ANALYSES

#### 6.1. Statistical Methods

All statistical analyses will be conducted using SAS Enterprise Guide v7.12 or equivalent software. Efficacy analyses will follow the intention-to-treat (ITT) principle. Safety analyses will include all randomized subjects who received at least one dose of study treatment and had any post-baseline safety follow-up.

**Logistic Regression:** The primary endpoint and categorical secondary endpoints will be analyzed using logistic regression, adjusting for geographic region, duration of current AA episode at baseline (< 4 years vs. ≥ 4 years), baseline value, and treatment group. A logistic regression model for the primary endpoint will be assessed for goodness-of-fit, with  $P > 0.05$  indicating adequate fit.

$$\log \left( \frac{\Pr(Y_i = 1)}{1 - \Pr(Y_i = 1)} \right) = \beta_0 + \beta_1 \cdot \text{Treatment}_i + \beta_2 \cdot \text{Region}_i + \beta_3 \cdot \text{Duration}_i + \beta_4 \cdot \text{Baseline}_i$$

**ANCOVA (Analysis of Covariance):** For continuous efficacy endpoints, ANCOVA will be used with the same covariates as in the logistic regression. It will also be used to compare continuous safety parameters (e.g., hemoglobin, CK, LDL, HDL) between treatment groups, adjusting for baseline values.

$$Y_i = \alpha_0 + \alpha_1 \cdot \text{Treatment}_i + \alpha_2 \cdot \text{Region}_i + \alpha_3 \cdot \text{Duration}_i + \alpha_4 \cdot \text{Baseline}_i + \varepsilon_i$$

$$Y_i = \gamma_0 + \gamma_1 \cdot \text{Treatment}_i + \gamma_2 \cdot \text{Baseline}_i + \varepsilon_i$$

**Fisher's Exact Test:** Fisher's exact test will be used to compare categorical safety outcomes such as adverse events, treatment discontinuations, and laboratory abnormalities between treatment groups.

#### 6.2. Statistical Analysis

Efficacy analyses will be performed by intention-to-treat. On the other hand, safety analysis will include all patients who are randomized and receive at least one dose of the baricitinib or placebo, regardless of whether they completed the study or were lost to follow-up. For categorical efficacy analyses, patients who do not meet any clinical response criteria will be considered as having no response. To control for a two-sided 5% significance level familywise error rate, the primary and secondary multiple comparison testing will be adjusted using

graphical testing schemes. Comparisons between the two treatment doses will not be included in the analyses.

### 6.2.1. Primary Analysis

Both trials are designed to have 90% power in testing the superiority of either treatment doses or placebo in the primary outcome, assuming a response rate of 30% for the 4-mg baricitinib, 20% for the 2-mg baricitinib, and 5% for the placebo. This will be analyzed using logistic regression with adjustments for treatment group, geographic region, duration of current alopecia areata episode at baseline (<4 years vs. ≥4 years), and baseline SALT score.

### 6.2.2. Secondary Analysis

For multiple comparison correction using graphical testing schemes, the order of secondary endpoints will be adjusted for BRAVE-AA1 after the results of BRAVE-AA2 are made available. For categorical secondary endpoints, a logistic regression will again be fitted, with the same covariates as in the primary analysis. For continuous secondary endpoints, ANCOVA testing will be used, with the same covariates in the categorical analyses.

### 6.2.3. Safety Analysis

Adverse events will be coded using a standardized medical dictionary and summarized by treatment group in terms of incidence, severity (mild, moderate, severe). Descriptive statistics will be used to present the frequency and proportion of participants experiencing any AE. comparison across treatment groups will be conducted using Fisher's exact test for categorical safety outcomes. Those continuous safety outcomes such as laboratory values and vital signs will be analyzed using analysis of covariance (ANCOVA) models. All safety data will be summarized over the entire treatment period and any significant trends or imbalances between groups will be noted.

## 7. REFERENCES

- Atanaskova Mesinkovska, N., King, B. A., Vañó-Galván, S., Shimomura, Y., Jedynak, J., McCollam, J., et al. (2024). Visualizing Severity of Alopecia Tool (SALT) scores in the clinical setting using patient images from a clinical trial. *JEADV Clinical Practice*, 3(2), 528–535. <https://doi.org/10.1002/jvc2.327>
- King, B., Ohyama, M., Kwon, O., Zlotogorski, A., Ko, J., Mesinkovska, N. A., Hordinsky, M., Dutronc, Y., Wu, W.-S., McCollam, J., Chiasserini, C., Yu, G., Stanley, S., Holzwarth, K., DeLozier, A. M., & Sinclair, R. (2022). Two Phase 3 Trials of Baricitinib for Alopecia Areata. *New England Journal of Medicine*, 386(18). <https://doi.org/10.1056/nejmoa2110343>
- Lepe, K., & Zito, P. M. (2024). Alopecia Areata. In *StatPearls*. StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK537000/>
- Pfizer Inc. [Internet]. SALT Evaluation. LITFULO™ (ritlicitinib) HCP. [cited 2025 Apr 29]. Available from <https://litfulo.pfizerpro.com/about/salt-evaluation>

- Yamaoka, K., Saharinen, P., Pesu, M., Holt, V. E., Silvennoinen, O., & O'Shea, J. J. (2004). The Janus kinases (Jaks). *Genome Biology*, 5, 253. <https://doi.org/10.1186/gb-2004-5-12-253>
- Wang, Y., Qiao, J., & Fang, H. (2023). Comparison of the latest Chinese and international guidelines/consensus on diagnosis and treatment of alopecia areata. *Chinese Journal of Dermatology*, 56(9), 885–888. <https://doi.org/10.35541/cjd.20210184>

## **8. APPENDICES**

### **APPENDICES A: Team Member Contribution**

Jason Dagher: Analysis Sets, Reference

Adeena Moghni: Statistical Analysis, Reference

Shehzrin Shah: Handling of Missing Values, Hypothesis and Decision Rules, Study Design, Sample Size Estimation, References

Puyuan Zhang: Statistical Methods, Reference

Longyi Zhao: Safety Analysis, Endpoints and Covariates, Reference

Yixin Zheng: Background, Study Objective, Eligibility Criteria, Sample Size Estimation, Reference