Sponsor	*Arcutis Biotherapeutics, Inc.*
Protocol Title:	*A Phase 3, 8-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.3% Administered QD in Subjects with Chronic Plaque Psoriasis*
Protocol Number:	*ARQ-151-301*
Premier Research PCN:	*ARCU9902*
Document Version:	*Final 1.0*
Document Date:	*16-Dec-2020*

Approvals

Role	Signatures	Date (dd-Mmm-yyyy)
Biostatistician	Print Name: [redacted] Sign Name: [redacted]	
Arcutis, Inc. Representative	Print Name: [redacted] Sign Name: [redacted]	
Arcutis, Inc. Representative	Print Name: [redacted] Sign Name: [redacted]	

Document History

Not applicable.

Table of Contents

List of Tables

1. Overview

This statistical analysis plan (SAP) describes the planned analysis and reporting for Arcutis, Inc. protocol number ARQ-151-301 (*A Phase 3, 8-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.3% Administered QD in Subjects with Chronic Plaque Psoriasis*), Amendment 1 version dated 21-Feb-2020. Reference materials for this statistical plan include the protocol and the accompanying sample data collection documents. Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analysis.

The structure and content of this SAP provides sufficient detail to meet the requirements identified by the Food and Drug Administration (FDA), European Medicines Agency (EMA), and International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Guidance on Statistical Principles in Clinical Trials[1]. All work planned and reported for this SAP will follow internationally accepted guidelines, published by the American Statistical Association[2] and the Royal Statistical Society[3], for statistical practice.

The planned analyses identified in this SAP may be included in clinical study reports (CSRs), regulatory submissions, or future manuscripts. Also, post-hoc exploratory analyses not necessarily identified in this SAP may be performed to further examine study data. Any post-hoc or unplanned, exploratory analysis performed will be clearly identified as such in the final CSR.

The statistical plan described hereafter is an *a priori* plan. It will be submitted to file prior to any unblinded inferential or descriptive analysis of data pertaining to Arcutis, Inc.'s study ARQ-151-301.

2. Study Objectives and Endpoints

2.1. Study Objectives

Psoriasis is a chronic inflammatory skin disease characterized by raised, well-demarcated, erythematous oval plaques with adherent silvery scales.

The advent of biological therapies has caused a transformation in the systemic treatment of moderate to severe psoriasis. However, for patients with milder forms of disease, the therapeutic landscape has not significantly changed in several decades. Lower potency topical steroids are not effective and the higher potency steroids can cause local skin atrophy and the potential for hypothalamic-pituitary axis suppression. Another typical therapy, vitamin D, is irritating and is not suitable for use on the face or intertriginous areas. There is substantial medical need for additional topical approaches in the treatment of mild to moderate psoriasis. The Phase 2 results suggest that ARQ-151 (Roflumilast Cream 0.3%) may be a highly efficacious and well-tolerated topical treatment for psoriasis.

2.1.1. Primary Objective

The primary objective is to assess the safety and efficacy of ARQ-151 cream 0.3% (Roflumilast Cream 0.3%) vs. vehicle administered QD for 8 weeks to individuals with 2-20% BSA of chronic plaque psoriasis.

2.2. Study Endpoints

2.2.1. Safety Endpoints

The safety endpoints of this study include the following:

- Adverse events (AEs)
- Local tolerance assessments
- Patient Health Questionnaire depression scale (PHQ-8) and Modified PHQ-9 for Adolescents (PHQ-A)
- Columbia-Suicide Severity Rating Scale (C-SSRS)
- Children's Depression Inventory 2nd Edition (CDI-2)
- Clinical laboratory results
- Vital signs
- Physical examination

2.2.2. Efficacy Endpoints

2.2.2.1. Primary Efficacy Endpoint

The primary efficacy endpoint of this study is success in Investigator Global Assessment (IGA) of disease severity, defined as an IGA of 'Clear' or 'Almost Clear' plus a 2-grade improvement from Baseline at Week 8.

2.2.2.2. Secondary Efficacy Endpoints

The secondary efficacy endpoints of this study include the following:

- Achievement of Psoriasis Area Severity Index-75 (PASI-75; subjects who achieve a 75% reduction in PASI from Baseline) at week 8.
- For subjects with intertriginous area involvement, and with severity of the intertriginous lesions at least 'mild' (intertriginous IGA (I-IGA) ≥2) at Baseline, achievement of 'I-IGA' score of 'clear' or 'almost clear' PLUS a 2-grade improvement from Baseline at week 8.
- In subjects with Worst Itch – Numeric Rating Score (WI-NRS) pruritus score ≥ 4 at baseline, achievement of a 4-point reduction in WI-NRS pruritus score at week 8 as compared to Baseline.
- In subjects with WI-NRS pruritus score ≥ 4 at baseline, achievement of a 4-point reduction in WI-NRS pruritus score at week 4 as compared to Baseline.
- In subjects with WI-NRS pruritus score ≥ 4 at baseline, achievement of a 4-point reduction in WI-NRS pruritus score at week 2 as compared to Baseline.
- Change from Baseline in total Psoriasis Symptoms Diary (PSD) score at week 8.
- Change from Baseline in total PSD score at week 4.
- Time to achieving Psoriasis Area Severity Index-50 (PASI-50; subjects who achieve a 50% reduction in PASI from Baseline)
- For subjects with intertriginous area involvement, and with severity of the intertriginous lesions at least 'mild' (I-IGA ≥2) at Baseline, achievement of 'I-IGA' score of 'clear' at week 8.

- Achievement of PASI-90 (subjects who achieve a 90% reduction in PASI from Baseline) at week 8.

3. Overall Study Design and Plan

3.1. Overall Design

3.2. Sample Size and Power

A sample size of approximately 400 subjects are planned for the study.

Approximately 267 subjects will receive ARQ-151 cream 0.3% (Roflumilast Cream 0.3%) QD; approximately 133 subjects will receive vehicle cream QD. The randomization scheme will be 2:1 (ARQ-151 cream 0.3% (Roflumilast Cream 0.3%) QD: matching vehicle QD).

This sample size provides >99% power to detect a 22.4% difference between treatment groups on IGA success at α=0.05 using a 2-sided Chi-squared test. The results from a recent phase 2b study (ARQ-151-201) of ARQ-151 (Roflumilast Cream 0.3%) compared to vehicle treatment were used to estimate the treatment difference. Specifically, in this trial 32.2% of subjects reported IGA success in the ARQ-151 0.3% (Roflumilast Cream 0.3%) group and 9.8% of subjects reported IGA success in the vehicle group.

The number of subjects to be enrolled will also provide sufficient power for the first 5 secondary endpoints. Additionally, the larger study size is included in order to provide additional/sufficient numbers of subjects on ARQ-151 (Roflumilast Cream 0.3%) treatment for a safety database.

3.3. Study Population

Males and females ages 12 years and older with a clinical diagnosis of psoriasis vulgaris on the face, extremities, trunk, and/or intertriginous areas involving 2% to 20% BSA (excluding the scalp, palms and soles) of at least 6 months duration that has been stable for the past 4 weeks, with an Investigator's Global Assessment of disease severity (IGA) of at least Mild ('2') and a PASI score of at least 2 (excluding the scalp, palms and soles) at Baseline.

3.4. Treatments Administered

Subjects will be randomized to one of the two following treatment groups in a 2:1 ratio (active:vehicle):

- ARQ-151 cream 0.3% (Roflumilast Cream 0.3%) QD
- vehicle cream QD.

3.5. Method of Assigning Subjects to Treatment Groups

Subjects will be randomized and assigned to active drug or vehicle in a 2:1 ratio (active:vehicle) according to a computer-generated randomization list. The randomization schedule will be stratified by study site, baseline IGA (IGA=2 vs. IGA≥3), and intertriginous involvement at baseline (I-IGA ≥2, yes vs no).

3.6. Blinding and Unblinding

This study is double-blind. In the event of a serious safety concern where the situation requires emergency unblinding this will be done by investigator using the study IWRS system after discussion with Medial Monitor and the Sponsor's CMO.

3.7. Schedule of Events

A detailed schedule of events for the study is provided in Table 1.

Table 1: Schedule of Events

Study Procedure	Screen	Baseline Day 1	Wk 2 Day 15	Wk 4 Day 29	Wk 6 Day 43	Wk 8[r] Day 57	Wk 9 Day 64
Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Visit Window	**-35 days**		**+/- 3 days**	**+/- 5 days**	**+/- 5 days**	**+/- 7 days**	**+/- 7 days**
Informed consent/assent	X						
Medical history	X						
Physical examination[a]	X	X				X	
I/E criteria	X	X					
Hematology, Serum Chemistries, and Urine Analysis[b]	X	X		X		X	
Vital signs, height, weight[c]	X	X	X	X	X	X	X
IGA[d], BSA[d], PASI/mPASI[d]	X	X	X	X	X	X	X
Intertriginous area IGA (I-IGA)[e]		X	X	X	X	X	X
WI-NRS[f]	X	X	X	X	X	X	
DLQI/CDLQI[g]	X	X	X	X		X	
Local Tolerability Assessments[h]		X		X		X	
C-SSRS, PHQ-8 / modified PHQ-A (12 years and older)	X	X		X		X	
CDI-2 parent report (6-11 y/o inclusive)	X	X		X		X	
PSD[i]	X	X	X	X	X	X	
Photography[j]		X	X	X	X	X	
Serum pregnancy test	X						
Urine pregnancy test[k]		X		X		X	
PK draws[l]		X		X		X	

IP application in clinic[m]		X	X	X	X	X	
Assign investigational product kit[n]		X					
Dispense/review diary		X	X	X	X	X	
Weigh investigational product tubes[n]		X	X	X	X	X	
Compliance calculation[o]		X	X	X	X	X	
Adverse event assessment[p]	X	X	X	X	X	X	X
Concomitant medications	X	X	X	X	X	X	X
Study Exit[q]							X

[a] Limited physical examination: skin, lungs, and heart only

[b] If Baseline visit occurs within 14 days of Screening, the Screening lab results may be utilized.

[c] Height will be collected at Baseline only. Weight will be collected at every visit. Subject to void prior to weight being taken. Remove any jackets, outerwear and shoes. Remove any objects of significant weight (i.e. cell phones, wallet, key chains). A 5% unintentional weight loss should be reported to the medical monitor on Page 1.

[d] IGA (based on whole body involvement) will be a 5-point scale ranging from clear (0) to severe (4). **IGA should be completed prior to other physician assessments and by the same Evaluator at each study visit.** Total BSA affected by psoriasis will be determined by the subject's hand method, where the subject's hand (including fingers) surface area is assumed to equal 1% of the BSA. PASI/mPASI will be determined by standard methods. NOTE: while palms and soles will be treated with IP, they are not counted towards IGA, PASI/mPASI, or BSA assessments.

[e] For subjects with intertriginous area involvement of at least 'mild' severity by IGA (IGA≥2) at Baseline (using the IGA scale but evaluating intertriginous areas ONLY and NOT whole body involvement), an IGA for the intertriginous region alone (I-IGA) will be recorded. **This 'intertriginous area IGA' should be done AFTER the 'standard whole body IGA' (primary endpoint) in subjects who qualify.**

[f] Subjects 12 years and older will complete WI-NRS pruritus assessment.

[g] Subjects ≥17 years of age will complete DLQI. For subjects 2-16 years of age, CDLQI will be completed.

[h] Tolerability assessments should be recorded prior to investigational product application for Investigator assessment (Berger and Bowman skin irritation score) and 10-15 minutes post-investigational product application for subject '0-3' burning/stinging assessment. Parents may report for children. **Note for investigator tolerability assessments: reactions at the site of product application, which may occur post-Baseline, should be differentiated from the preexisting inflammation associated with the subject's Psoriasis.**

[I] Adult subjects only

[j] Photography will be performed using Canfield equipment on all subjects at all sites. All efforts will be made to de-identify the subjects. Subjects who are unwilling to participate in the medical photography will be allowed to opt out of this procedure and documented on the informed consent. See Photography Manual for details.

[k] A urine pregnancy test will be administered to all females of child-bearing potential. A negative result is required for continued participation in the study, and results must be available prior to dispensing of investigational product at each visit.

[l] PK draws will be collected from all subjects at Days 1, 29 and 57. The draws will be pre-dose investigational product application in the clinic (i.e., trough levels). Ensure investigational product is not applied in the area where PK will be drawn.

[m] Subjects to apply assigned IP in clinic at every visit. The time of application will be documented.

[n] Kits will be dispensed based on % BSA affected. See IP Handling Manual for details.

[o] Each IP tube should be weighed and recorded at every visit. See IP Handling Manual for details.

[p] Any emergent AEs will be followed in the clinic for up to one month at the Investigator's discretion until resolved or otherwise judged as clinically stable.

[q] Subjects who consent to/enroll into the open label extension study (ARQ-151-306) will complete the study at Week 8; subjects that do not consent to/enroll into ARQ-151-306 will return at Week 9 to complete the study.

[r] Subjects that terminate early should return to the clinic for the week 8 assessments.

4. Statistical Analysis and Reporting

4.1. Introduction

Data processing, tabulation of descriptive statistics, calculation of inferential statistics, and graphical representations will be performed primarily using SAS (release 9.4 or higher). If the use of other software is warranted, the final clinical study report will detail what software was used for what purposes.

Continuous (quantitative) variable summaries will include the number of subjects (n) with non-missing values, mean, standard deviation (SD), median, minimum, and maximum.

Categorical (qualitative) variable summaries will include the frequency and percentage of subjects who are in the particular category or each possible value. In general, the denominator for the percentage calculation will be based upon the total number of subjects in the study population for the treatment group, unless otherwise specified.

The minimum and maximum will be reported with the same degree of precision (ie, the same number of decimal places) as the observed data. Measures of location (mean and median) will be reported to 1 degree of precision more than the observed data and measures of spread (SD) will be reported to 2 degrees of precision more than the observed data, unless otherwise specified.

Percentages will be presented to 1 decimal place, unless otherwise specified.

Unless otherwise indicated, all statistical tests will be conducted at the 0.05 significance level using 2-tailed tests, and P values will be reported. Corresponding *95%* confidence intervals (CIs) will be presented for statistical tests where appropriate.

A *p*-value of ≤ 0.10 but > 0.05 will be considered evidence of a trend.

4.2. Interim Analysis and Data Monitoring

No interim analyses or data safety monitoring are planned.

5. Analysis Populations

The following analysis populations are planned for this study:

- **Safety Population (SAF):** The Safety Population includes all subjects who are enrolled and received at least one confirmed dose of investigational product. This population will be used for all safety analyses.
- **Intent-To-Treat Population (ITT):** The ITT population includes all randomized subjects. This population will be the primary analysis population for the analysis of efficacy endpoints.

- **Modified Intent-To-Treat Population (mITT):** The mITT population includes all randomized subjects with the exception of subjects who missed the week 8 IGA assessment specifically due to COVID-19 disruption. This population will be used for sensitivity of the primary endpoint.
- **I-IGA ITT Population (I-IGA-ITT):** The I-IGA population is a subset of the ITT population and includes subjects with intertriginous area involvement, and with severity of the intertriginous lesions at least 'mild' (intertriginous IGA (I-IGA) ≥2) at Baseline.
- **I-IGA mITT Population (I-IGA-mITT):** The I-IGA population is a subset of the mITT population and includes subjects with intertriginous area involvement, and with severity of the intertriginous lesions at least 'mild' (intertriginous IGA (I-IGA) ≥2) at Baseline.
- **Pruritus ITT Population (PRU4-ITT):** The PRU4 population is a subset of the ITT population and includes subjects with WI-NRS pruritus score ≥ 4 at Baseline. This population will be used for the analysis of achievement of a 4-point reduction in WI-NRS pruritus score as compared to Baseline.
- **Pruritus mITT Population (PRU4-mITT):** The PRU4 population is a subset of the mITT population and includes subjects with WI-NRS pruritus score ≥ 4 at Baseline. This population will be used for the analysis of achievement of a 4-point reduction in WI-NRS pruritus score as compared to Baseline.
- **Pharmacokinetic Population (PK):** The PK population includes all subjects receiving active drug and had a PK draw (concentration data available) at Baseline Day 0 visit. This population will be used for the analysis of PK concentrations.

6. General Issues for Statistical Analysis

6.1. Statistical Definitions and Algorithms

6.1.1. Baseline

The last observation recorded prior to the first dose of investigational product being administered will be used as the baseline observation for all calculations of change from baseline except subject tolerability, WI-NRS, DLQI and PSD. The value for date and time of IP administration recorded in the EDC database will be used as the anchor for baseline. For assessments where no time is recorded (e.g., DLQI/CDLQI and PHQ/mPHQ-A) and the date of the assessment is the same as the date of first IP, it is assumed that the assessment took place prior to IP application, per study site training, and the last assessment taken on or before the date of first IP is used as the baseline.

For subject tolerability assessments, baseline is derived as the last non-missing measurement taken on the day of first application of study drug.

6.1.2. Adjustments for Covariates

If there is a statistical difference among treatment groups with respect to baseline characteristics, that variable may be added to the statistical models as a blocking factor or covariate to determine the effect on treatment.

6.1.3. Multiple Comparisons

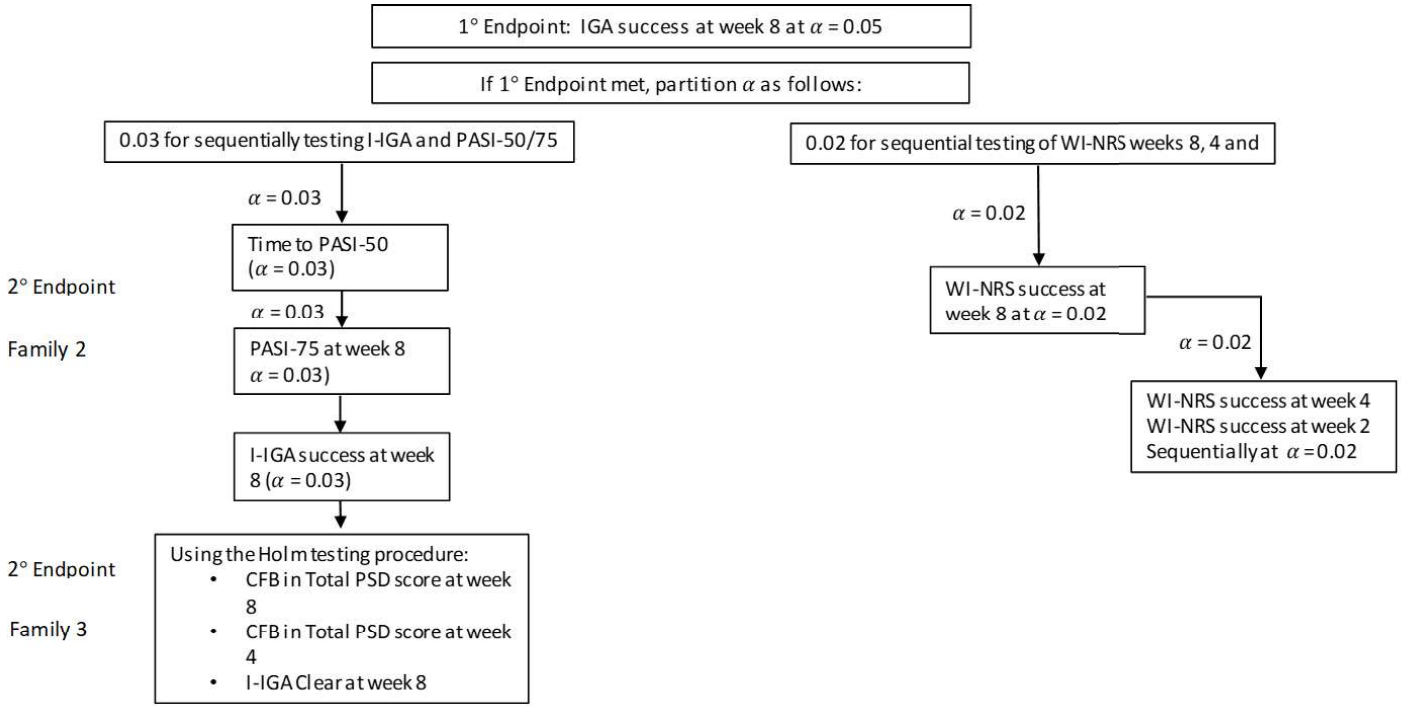
The type I error rate of $\alpha = 0.05$ will be maintained for all secondary efficacy endpoints included in the hierarchical testing strategy by assigning the endpoints into families and only proceeding to the next family in accordance with the rules of the pre-specified sequential testing strategy.

Secondary endpoints will only be tested statistically if the primary endpoint is considered statistically significant. To control for multiple comparisons among the secondary endpoints, the following multiplicity procedure will be used:

Upon successful testing of the primary endpoint (Family 1) the alpha will be partitioned to test secondary endpoint Families 2 and 3 (Partition 1) and to test WI-NRS timepoints (Partition 2).

Partition 1 of alpha = 0.03 will be allocated to test Family 2 and Family 3. The endpoints in Family 2 will be tested hierarchically in the order depicted below; the next 3 secondary endpoints (Family 3) will be tested using the alpha available after testing endpoint Family 2. The Holm procedure will be used to control for multiple comparisons in endpoint Family 3.

Partition 2 of alpha = 0.02 will be allocated to sequentially test WI-NRS success at week 8, then week 4 and subsequently week 2.



No adjustments will be made for multiple comparisons for other endpoints.

6.1.4. Handling of Dropouts or Missing Data

Any subject who prematurely withdraws from the study will have their last available data assigned to an analysis window as described in Section 6.1.5.

6.1.4.1. Imputation of Missing Data

For the primary efficacy endpoint of IGA score and tested secondary endpoints, the primary analysis will impute missing values using a regression-based multiple imputation model. This is a three step process.

1. The first step is to understand the pattern of missingness. In order to perform the multiple imputation, a monotone missing pattern has to be achieved. For example, if there exists values for baseline and Week 8 visits, but missing values for the Week 2 or 4 visits, the Markov-Chain Monte-Carlo (MCMC) method will be used to impute the small amount of missing data that may be missing at the intermediate visits that is required to make the missing data pattern monotone before applying the multiple imputation algorithm. This method uses a non-informative Jeffreys prior to derive the posterior mode from the expectation-maximization (EM) algorithm as the starting values for the MCMC method. The MCMC method will use the seed 59726314. The IGA score will be treated as a continuous variable for this step. To avoid values that could not be observed in practice, imputed values will be constrained to be integers in the range of 0 to 4.

a. The table below will determine the number of datasets to be imputed in this step. Determine the proportion of datapoints with non-monotone pattern across all visits and subjects which could be derived as a percentage of number of nonmonotone data points/total number of expected data points.

This can be determined as $\frac{number\ of\ non\ montone\ visits}{total\ number\ of\ visits\ across\ all\ subjects} * 100$

Non-monotone Missing Data	Number of Imputed Datasets
≤ 2%	1
> 2% to ≤ 5%	3
> 5%	10

2. Once the monotone pattern is achieved, the next step is to implement the imputation algorithm. For this, the Predictive Mean Matching method (PMM) will be used. This method is particularly helpful if the normality assumption is violated. For subjects with complete data up to a particular visit, a PMM model will be fit that includes the outcome at that visit as the dependent variable and as independent variables, IGA score outcomes at previous visits, baseline IGA score, treatment group, and investigational site using a seed of 461903. This process will be repeated 25 times, resulting in a total of 25 to 250 complete analysis datasets, depending on the number of imputed monotone datasets that are required. The seed may be changed after unblinding in case of any issues with the imputation process, and it will be documented in the CSR if any change is required.
3. For each completed dataset, compute the necessary derived variables. The dichotomous success rate (clear or almost clear with at least a 2-point change from baseline) will be derived using these datasets. The results obtained will be analyzed using the Cochran-Mantel-Haenszel (CMH) test. Analysis will be performed separately for each of the 25 to 250 complete analysis data sets, and the results will be combined into one multiple imputation inference (odds ratio, associated confidence interval and *p*-value) using PROC MIANALYZE[4].

This approach to imputation should be superior to other strategies such as carrying forward the last available observation, which may yield unrealistic imputed values. Also, the use of multiple imputation avoids the problem of artificially increasing power through data imputation associated with single-imputation methods because it accounts for the uncertainty associated with the imputation.

Data for secondary endpoints intended for hypothesis testing will be imputed. No other data, with the exception of incomplete dates as described in Section 6.1.8, will be imputed.

The SAS pseudo code for the multiple imputation process is listed below:

Step 1:

```
proc mi data=example seed=59726314 nimpute=XX round=1 out=example_1;
  mcmc impute=monotone;
  var <baseline score> …… <visit8 score>;
run;
```

Step 2:

```
proc mi data=example_1 seed=461903 nimpute=XX  out=example_2;
  class <treatment> <site>;
  monotone regpmm(<baseline score> …… <visit8 score>);
  var <treatment> <country> <baseline score> …… <visit8 score>;
run;

XX will be the determined based on the proportion of missing data across visits.
```

Step 3: This step involves running CMH test stratified by baseline IGA score, and site on each completed dataset and combining the results using PROC MIANALYZE.

```
proc freq data=example noprint;
  by <imputationnumber> <visit>;
  tables <site> * <BL IGA> * <treatment> *
         <outcome> / cmh alpha=0.05;
  output out=example_stat cmh;
run;
```

In order to apply PROC MIANALYZE, normalizing transformations have to be applied to the odds ratio. *P* values are obtained using the Wilson Hilferty transformation[4].

6.1.4.2. Tipping Point Analysis

As a sensitivity analysis to the multiple imputation analysis as described in Section 6.1.4.1 for the IGA primary endpoint, a tipping point analysis will be performed in order to determine the inflection point at which the inference under the MNAR assumption changes substantially. This will be used to check the robustness of the imputation.

The sensitivity analysis will be performed by using a specified sequence of shift parameters, which will adjust the imputed values for observations in the active treatment group. The range of shift parameters to be included in this analysis are -1.5 to 1.5 by 0.2. Once the likely point of the shift is determined, the analysis will be rerun using an expanded range around the suspected tipping point, with greater precision (i.e., to 2 decimal places, by 0.01). Thus, the value at which the results of the analysis are shifted from significant (i.e., $\alpha \leq 0.05$) to non-significant (i.e., $\alpha > 0.05$) will be determined.

Steps 1 and 3 of the analysis will be the same as for the multiple imputation analysis as described in Section 6.1.4.1. However, Step 2 of the analysis is where the shift parameters will be applied. Pseudo-code for Step 2 is as follows:

```
proc mi data=example_1 seed=461903 nimpute=XX  out=example_2;
  class <treatment> <site> <BL IGA>;
  monotone regpmm(<baseline score> …… <visit8 score>);
  var <treatment> <site> <baseline score> …… <visit8 score>;
  mnar adjust( <visit2 score> / shift=YY adjustobs=(treatment = 'Roflumilast
Cream 0.3%'));
```

```
  mnar adjust( <visit4 score> / shift=YY adjustobs=(treatment = 'Roflumilast
Cream 0.3%'));
  mnar adjust( <visit8 score> / shift=YY adjustobs=(treatment = 'Roflumilast
Cream 0.3%'));
run;

XX will be the determined based on the proportion of missing data across visits.
YY will encompass the range of shift parameters as pre-specified above.
```

6.1.5. Analysis Visit Windows

Visits will be analyzed as scheduled. Unscheduled, early termination visits, and/or repeated measurements will only be included if a scheduled measurement is not available and the early termination or unscheduled/repeated measurement falls within the analysis visit windows as described in Table 2. The windows follow the Schedule of Events in Table 1. Unscheduled/repeated measurements will be listed.

Table 2: Analysis Visit Windows

Visit Name	Visit Number	Target Start Day	Lower Limit	Upper Limit
Week 2	2	15	2	22
Week 4	3	29	23	35
Week 6	4	43	36	50
Week 8	5	57	51	61
Week 9	6	64	62	

6.1.6. Pooling of Sites

Sites will be pooled for statistical analysis as follows. For analysis, sites should have a minimum of 10 randomized subjects (ITT population). The smallest sites will be grouped sequentially in order of smallest to largest, restricting to those sites that did not meet the minimum enrollment of 10, until each pooled site has a minimum of 10 subjects with at least one subject in each treatment group. Prior to pooling, descriptive statistics on the outcome by site will be assessed to evaluate clinically meaningful differences in outcomes between sites.

6.1.7. Derived Variables

All questionnaire scores (including PASI and mPASI) will be derived by Biostatistics in the ADaM datasets using the formulas defined below, even if calculated scores are present in the EDC or ERT databases. All pre-calculated scores will be ignored for analysis.

- Investigator Global Assessment (IGA) success = IGA of 'Clear' or 'Almost Clear' plus a 2-grade improvement from Baseline.
- Intertriginous IGA (I-IGA) success = I-IGA of 'Clear' or 'Almost Clear' plus a 2-grade improvement from Baseline.

- Worst Itch – Numeric Rating Score (WI-NRS) 4-point reduction = achievement of a 4-point reduction in WI-NRS pruritus score compared to baseline, calculated only for subjects with a pruritus score of ≥ 4 at baseline.

- PASI = 0.1 ($E_h + T_h + S_h$) A_h + 0.2 ($E_a + T_a + S_a$) A_a + 0.3 ($E_t + T_t + S_t$) A_t + 0.4 ($E_l + T_l + S_l$) A_l where E, T, and S are erythema (redness), thickness (induration), and scaling (desquamation), respectively, scored on a scale of 0 to 4, A is estimated area of skin involved, graded on a scale of 0 to 6, and h, e, t, and l are head, arms, trunk, and legs, respectively (range for total score 0 to 72). If any one of the components/scores is missing, the PASI score is set to missing.

- mPASI = same as PASI above, except that for subjects with < 10% of any particular involved anatomic area, the mPASI will be calculated using the actual percentage of the anatomical area involved rather than the 0 to 6 estimated area score (e.g. 0.1 for 1%, 0.2 for 2%, 0.3 for 3%, … 0.9 for 9%). If any one of the components/scores is missing, the mPASI score is set to missing.

- PASI-50(mPASI-50) = achievement of a 50% reduction in PASI(mPASI) from Baseline.

- Time to PASI-50 Success (days) = date of PASI-50 achievement – Day 1 date + 1.

- PASI-75(mPASI-75) = achievement of a 75% reduction in PASI(mPASI) from Baseline.

- PASI-90(mPASI-90) = achievement of a 90% reduction in PASI(mPASI) from Baseline.

- PASI-100(mPASI-100) = achievement of a 100% reduction in PASI(mPASI) from Baseline.

- PSD Total Score = sum of the 16 questions (individual questions scored 0 to 10; range for total score 0 to 160). If 1 or more items are missing, the score is not calculated.

- DLQI Score = sum of the 10 questions (individual questions scored as Very much=3, A lot=2, A little=1, Not at all=0, Not relevant=0, Question 7: Yes=3; range for score 0 to 30). If 1 item is missing, it is scored as 0 for that item. If 2 or more items are missing, the score should not be calculated. Note that 'Not Relevant' is scored as 4 in the ERT database and must be converted to 0 for the calculation of the total score.

- CDLQI Score = sum of the 10 questions (individual questions scored as Very much=3, Quite a lot=2, Only a little=1, Not at all=0; Question 7: if the last week was school time, the question was scored as Very much=3, Quite a lot=2, Only a little=1, Not at all=0, with Prevented school recoded to 3, and if the last week was vacation, the standard responses apply; range for score 0 to 30). If 1 item is missing, that item is scored as 0. If 2 or more items are missing, the score should not be calculated.

- PHQ-8 = sum of the 8 questions (individual questions scored as Not at all=0, Several days=1, More than half the days=2, and Nearly every day=3, (range for score 0 to 24). If

more than 1 item is missing the score should not be calculated. If 1 item is missing the score is calculated as (sum of answered items*8)/number of answered items.

- Modified PHQ-A = sum of the 8 questions (individual questions scored as Not at all=0, Several days=1, More than half the days=2, and Nearly every day=3, (range for score 0 to 24). If 1 or more items are missing the score should not be calculated. If 1 item is missing, the score is calculated as (sum of answered items*8)/number of answered items.

- CDI-2 total score is a sum of the 17 questions (individual questions scored as much or most of the time=3, often=2, some of the time=1, Not at all=0; range for score 0 to 51).

    - CDI-2 emotional problem scale is a sum of 9 questions (Q1, Q3-6, Q8, Q10 -12)

    - CDI-2 function problem scale is a sum of 8 questions (Q2, Q7, Q9, Q13-17)

- Change from baseline = value at current time point – value at Baseline.

- TEAE = any adverse event with an onset date/time after the first application of investigational product through study completion; or onset date at/after date of the first IP when time is missing

- C-SSRS Suicidal ideation = A "yes" answer at any time during treatment to any one of the five suicidal ideation questions (Categories 1-5: Wish to be Dead, Non-specific Active Suicidal Thoughts, Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act, Active Suicidal Ideation with Some Intent to Act, without Specific Plan Active Suicidal Ideation with Specific Plan and Intent).

- C-SSRS Suicidal behavior = A "yes" answer at any time during treatment to any one of the five suicidal behavior questions (Categories 6-10: Preparatory Acts or Behavior, Aborted Attempt, Interrupted Attempt, Actual Attempt (non-fatal), Completed Suicide).

- C-SSRS Suicidal ideation or behavior = A "yes" answer at any time during treatment to any one of the ten suicidal ideation and behavior questions (Categories 1-10) on the C-SSRS.

- Weight of IP (g) = dispensed can weight – returned can weight.

- Compliance = number of applications divided by the expected number of IP applications for each subject. Compliance will be calculated using drug accountability data over the entire treatment period for each subject, up to treatment completion or discontinuation. Both the EDC and ERT databases will be used to determine missed doses, etc.

- Number of expected IP applications = calculated as the last treatment date – first treatment date +1.

- Number of IP applications = number of expected IP applications – missed IP applications as collected on the CRF.

- BMI (kg/m^2) = (weight in kg)/[(height in cm/100)^2]. Baseline height will be used to derive BMI for each visit since height is not collected at all visits.

- BMI Categories

    - Underweight: BMI < 18.5
    - Normal: 18.5 <= BMI <= 24.9
    - Overweight: 25.0 <= BMI < = 29.9
    - Obese: BMI >= 30.0

- Intentional Weight Loss = for the analyses of subjects losing >=5%/>=10% of their baseline body weight, if the question 'was the subject intentionally dieting or losing weight'? is answered 'yes' at each visit.

- Completion of Study = Completion of the primary efficacy assessment (IGA) at Week 8. This may differ from what was recorded in the eCRF.

- BSA (body surface area) = sum of coverage areas of head, limbs, upper limbs and trunk/genitalia.

6.1.8. Data Adjustments/Handling/Conventions

All collected data will be presented in listings or CDISC datasets. Data not subject to analysis according to this plan will not appear in any tables or graphs but will be included only in the data listings or CDISC datasets.

All *P* values will be displayed in four decimals and rounded using standard scientific notation (eg, 0.XXXX). If a *P* value less than 0.0001 occurs it will be shown in tables as <0.0001.

All analyses that include the stratification factors (site, baseline IGA, and baseline intertriginous involvement) will use the data as collected in the IVR system. If the overall number of discrepancies between the IVR and the eCRF database exceeds 10%, a sensitivity analysis will be conducted in which the stratification will be based on the stratification values as collected in the eCRF database.

In the event that data from a questionnaire for a particular visit is present in both the EDC database and the ERT database, the values present in the ERT database will be used for all analyses. If a questionnaire for a visit is only present in the ERT database, then the ERT values will be used, and similarly if only present in the EDC database then the EDC values will be used.

For laboratory values that are < X (X = a number), <= X, > X or >=X then X will be used in the analysis.

Adverse events will be coded using the MedDRA version 23.0 thesaurus.

A treatment related AE is any AE with a relationship to IP of possibly, probably, or likely.

For partial AE and medication start dates:

- If the year is unknown, then do not impute the date but assign a missing value.
- If the year is known, but the month or month and day is unknown, then:
    - If the year matches the year of first dose date and the end date is missing or after first dose date, then impute as the month and day of the first dose date.
    - Otherwise, assign 01 January.
- If the year and month are known, but the day is unknown, then:
    - If the month and year match the month and year of the first dose date, then impute as the day of the first dose date.
    - Otherwise, assign 01.

For partial AE and medication end dates:

- If the year is unknown, then do not impute the date but assign as missing value.
- If the year is known, but the month or month and day is unknown, then:
    - If the year matches the year of the last date of the study (date of last contact if subject lost to follow-up; date of completion or early termination otherwise), then impute as the month and day of the last date of the study.
    - Otherwise, assign 31 December.
- If the year and month are known, but the day is unknown, then:
    - If the month and year match the month and year of the last date of the study, then impute as the day of the last date of the study.
    - Otherwise, assign the last day of the month.

6.2. Special Handling for COVID-19 Disruptions

In some cases, study visits will have to be delayed/not performed as a result of COVID-19 disruptions (e.g., sites were closed or subjects were under stay-at-home orders). Where possible, study sites may collect post-baseline data from subjects remotely by telephone, traditional mail, and/or email; this will be clearly documented in the source. If possible, sites should adhere to the protocol visit window for remote data collection.

Investigator assessments and subject questionnaires normally completed directly in the electronic patient reported outcomes (ePRO) tablet during on-site visits should be completed on the appropriate paper source documents. The following assessments/questionnaires are approved to be collected via telemedicine/remotely:

- WI-NRS
- DLQI/CDLQI
- C-SSRS
- PHQ-8/PHQ-A
- Subject Local Tolerability

The following assessments cannot be completed via telemedicine/remotely:

- IGA/I-IGA
- BSA
- Investigator Local Tolerability
- PASI/mPASI
- PSD
- Subject Weight

Study visits and procedures must be followed per protocol whenever possible. Any specific changes in study conduct that deviate from the protocol should be communicated to the institutional review board and Sponsor. All protocol deviations which occurred as a result of COVID-19 disruptions (e.g., visits out of window, missed assessments, etc.) will be differentiated from other protocol deviations.

Subjects who were affected by COVID-19 disruptions by either missing their Week 8 visit or being discontinued before having a Week 8 visit due to COVID-19 related disruptions will be excluded from the mITT population, as described in Section **Error! Reference source not found.**.

7. Study Patients/Subjects and Demographics

7.1. Disposition of Patients/Subjects and Withdrawals

Disposition will include tabulations of the number of subjects randomized into each treatment group, the number of subjects who received treatment, tabulated reasons for discontinuation from the study, and number of subjects in each analysis population.

7.2. Protocol Violations and Deviations

The number of subjects with major protocol deviations will be summarized in categories by treatment group and overall for subjects in Safety population.

Protocol deviations will be listed.

7.3. Demographics and Other Baseline Characteristics

Summary statistics for age, age groups (2- 11, 12 – 17 and ≥18 years old), gender, race, ethnicity, height, weight, percent BSA covered with plaque psoriasis, IGA, I-IGA, PASI, mPASI, WI-NRS, number of subject with WI-NRS >= 4, DLQI, CDLQI, PSD, PHQ-8, PHQ-A and BMI will be presented by treatment group and overall.

For the continuous variables, the number of non-missing values and the mean, standard deviation, minimum, median and maximum will be tabulated.

For the categorical variables, the counts and proportions of each value will be tabulated.

These analyses will be conducted for the ITT, mITT, I-IGA-ITT, PRU4-ITT, and Safety populations.

Medical history will be listed.

The number and percent of subjects reporting inadequate response, intolerance or contraindication to topical corticosteroids, vitamin D Derivatives, Apremilast (Otezla), taken conventional systemic therapy or been on phototherapy will be tabulated by treatment group and overall.

The number and percent of subjects reporting psoriasis involvement in knee, elbow, knee and/or elbow, facial and genitalia will be tabulated by treatment group and overall.

7.4. Exposure and Compliance

The number of investigational product applications by each subject based on diary data will be summarized using descriptive statistics appropriate for continuous variables.

The amount of investigational product used by each subject based on tube weight will be summarized descriptively by treatment using continuous methods.

A subject will be considered compliant with the dosing regimen if the subject applies at least 80% of the expected applications during the investigational product application period and does not miss more than 3 consecutive doses. The number of subjects that missed more than 3 consecutive doses will also be displayed. The number and percent of subjects that missed more than 3 consecutive doses will be tabulated by treatment group and overall.

Investigational product application compliance will be calculated as described in section 6.1.7. Compliance will be summarized descriptively by treatment group using the following categories:

> 100%

≥ 80% - ≤100%

< 80%.

8. Efficacy Analysis

All efficacy analyses that include the stratification factors (site, baseline IGA, and baseline intertriginous involvement) will use the data as collected in the IVR system. If the overall

number of discrepancies between the IVR and the eCRF database exceeds 10%, a sensitivity analysis will be conducted in which the stratification will be based on the stratification values as collected in the eCRF database.

8.1. Primary Efficacy Analysis

For this study, the primary estimand is the ratio of the odds of achieving IGA success after 8 weeks of using ARQ-151 (Roflumilast Cream 0.3%), relative to the odds of success after 8 weeks of using a matching vehicle cream. In the course of the 8-week randomized treatment period, subjects may be exposed to possible known or unknown inter-current events that could possibly impact the estimand, such as treatment discontinuation due to a specific adverse effect or perhaps a lack of effect. The "Treatment Policy Strategy" has been adopted for handling all known or unknown inter-current events in this study. To this end, the intent-to-treat (ITT) principle will serve as the analytical basis for interpreting the estimand. In other words, the odds ratio of achieving IGA success for ARQ-151 (Roflumilast Cream 0.3%) relative to vehicle after 8 weeks will be evaluated regardless of the occurrence of any such inter-current event. This estimand shall be estimated using the CMH approach. This approach produces an estimate which is the combined odds ratio resulting from adjusting for the possible confounding effects of three classification factors – investigative site, baseline IGA and baseline intertriginous involvement.

The primary efficacy endpoint is success in IGA of disease severity, defined as an IGA of 'Clear' or 'Almost Clear' plus a 2-grade improvement from Baseline at Week 8.

The primary endpoint will be analyzed using a Cochran-Mantel-Haenszel (CMH) test stratified by site, baseline IGA, and baseline intertriginous involvement. Statistical significance will be concluded at the 5% significance level (2-sided).

For the primary analysis, missing IGA scores will be imputed using multiple imputation as described in Section 6.1.4. These imputations will result in a minimum of 25 to a maximum of 250 complete analysis datasets, depending on the number of imputed monotone datasets that are required.

The CMH analyses will be performed separately for each of the complete analysis data sets, and the results will be combined into one multiple imputation inference (estimated treatment effect and associated confidence interval and p-value).

Sensitivity analyses of the primary endpoint will also be performed using the original (non-imputed, observed data) dataset. These will include a repeated measures logistic regression model (GEE) with IGA success as the dependent variable and treatment, and visit as the independent variables, and a tipping point analysis. The tipping point analysis will be performed as described in Section 6.1.4.26.1.4.2. For the tipping point analysis, a figure will be provided that displays the *P* values in the analysis plotted against the range of shift parameters; included in this figure will be the shift parameters and *P* values for the range of shift parameters to provide information on where the analysis tips from significant to non-significant.

All other missing data for all other analyses and summaries will remain missing and will not be imputed. Only observed data will be included in the summaries showing descriptive statistics.

The primary efficacy analysis and the sensitivity analyses will be based on the ITT/mITT population. The analysis of the mITT population will use the MI datasets already created for the

ITT population (i.e. no new MI datasets will be generated).

8.2. Secondary Efficacy Analysis

The secondary endpoints are:

- Time to achieving Psoriasis Area Severity Index-50 (PASI-50; subjects who achieve a 50% reduction in PASI from Baseline)Achievement of Psoriasis Area Severity Index-75
- (PASI-75; subjects who achieve a 75% reduction in PASI from Baseline) at week 8.
- For subjects with intertriginous area involvement, and with severity of the intertriginous lesions at least 'mild' (intertriginous IGA (I-IGA) ≥2) at Baseline, achievement of 'I-IGA' score of 'clear' or 'almost clear' PLUS a 2-grade improvement from Baseline at week 8.
- In subjects with Worst Itch – Numeric Rating Score (WI-NRS) pruritus score ≥ 4 at baseline, achievement of a 4-point reduction in WI-NRS pruritus score at week 8 as compared to Baseline.
- In subjects with WI-NRS pruritus score ≥ 4 at baseline, achievement of a 4-point reduction in WI-NRS pruritus score at week 4 as compared to Baseline.
- In subjects with WI-NRS pruritus score ≥ 4 at baseline, achievement of a 4-point reduction in WI-NRS pruritus score at week 2 as compared to Baseline.
- Change from Baseline in total Psoriasis Symptoms Diary (PSD) score at week 8.
- Change from Baseline in total PSD score at week 4.
- For subjects with intertriginous area involvement, and with severity of the intertriginous lesions at least 'mild' (I-IGA ≥2) at Baseline, achievement of 'I-IGA' score of 'clear' at week 8.

For the continuous variables, the number of non-missing values and the mean, standard deviation, minimum, median and maximum will be tabulated by treatment group and visit and similarly, categorical variables will have the counts and proportions of each value will be tabulated by treatment group and visit.

Each of the binary endpoints - achievement of PASI-75/50 at week 8, achievement of PASI-90 at week 8, achievement of 'I-IGA' score of 'clear' or 'almost clear' PLUS a 2-grade improvement from Baseline at week 8 in the I-IGA-ITT population, achievement of a 4-point reduction in WI-NRS pruritus score from baseline to weeks 8, 4, and 2 in the PRU4-ITT population, and achievement of 'I-IGA' score of 'clear' at week 8 in the I-IGA population - will be analyzed using CMH tests stratified by site, and baseline IGA, similar to the primary analysis above, with the exception that missing data will not be imputed. Analyses of variables mentioned above with no population specified will be performed on the ITT population. Only observed data will be included in the descriptive statistics.

The continuous endpoints change from baseline in total PSD score at week 4 and at week 8 will be analyzed using an analysis of covariance (ANCOVA: total score only) with the factors treatment, site, baseline IGA, baseline intertriginous involvement, and baseline of the variable under analysis . Statistical comparisons between the treatment groups will be obtained using contrasts. The LS Means, standard errors, 95% confidence intervals, and p-values will be presented. These analyses will be performed on the ITT population.

The secondary endpoint time to achieving PASI-50 will be summarized using Kaplan-Meier methods and the difference between treatment groups evaluated using the log-rank statistic and 95% confidence intervals of the median for each treatment group will be presented. In addition, a Cox proportional hazards model that includes the stratification factors will be run and the hazard ratio and associated 95% confidence interval will be presented. These analyses will be performed on the ITT population.

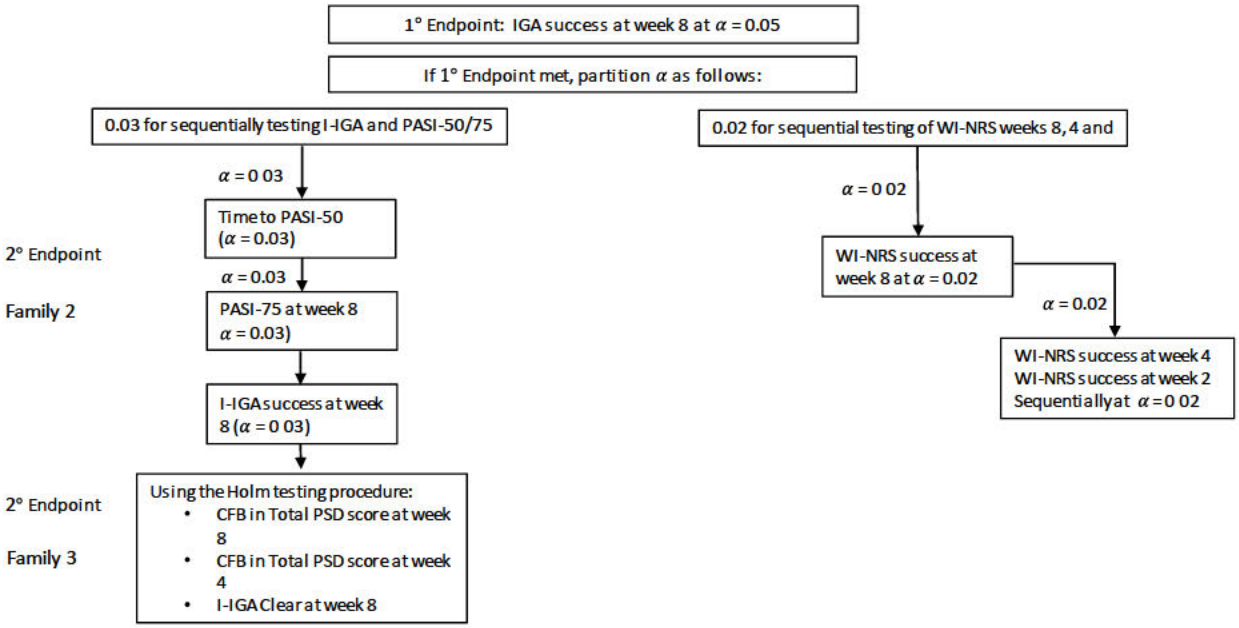
Secondary endpoints will only be tested statistically if the primary endpoint is considered statistically significant. To control for multiple comparisons among the secondary endpoints, the following multiplicity procedure will be used:

Upon successful testing of the primary endpoint (Family 1) the alpha will be partitioned to test secondary endpoint Families 2 and 3 (Partition 1) and to test WI-NRS timepoints (Partition 2).

Partition 1 of alpha = 0.03 will be allocated to test Family 2 and Family 3. The endpoints in Family 2 will be tested hierarchically in the order depicted below; the next 3 secondary endpoints (Family 3) will be tested using the alpha available after testing endpoint Family 2. The Holm procedure will be used to control for multiple comparisons in endpoint Family 3.

Partition 2 of alpha = 0.02 will be allocated to sequentially test WI-NRS success at week 8, then week 4 and subsequently week 2.

A figure summarizing the endpoint testing is presented below.



Failure to successfully pass any testing gate in the hierarchy of testing families at any stage in the sequence implies automatic failure at subsequent stages. The p-value for each comparison will be reported in the summaries for informational purposes regardless of whether the previous comparison reaches significance.

8.3. Other Efficacy Analysis

The binary variables at weeks other than the primary/secondary endpoints – IGA success, WI-NRS success, IGA score of "clear', PASI-50, PASI-75, PASI-90, and PASI-100 – as well as mPASI-50, mPASI-75, mPASI-90, and mPASI-100 at all weeks, will be analyzed using the same CMH model described above on the ITT population.

I-IGA success at weeks 2, 4, and 6 will also be analyzed using the same CMH methodology on the I-IGA population.

Continuous endpoints including change and percent change from baseline in PASI, change from baseline in affected percent BSA, change and percent change from baseline in mPASI, change and percent change in PSD, change and percent change in WI-NRS, change and percent change in DLQI, and change and percent change in CDLQI will be analyzed using an analysis of covariance (ANCOVA: total score only) with the factors treatment, site, baseline IGA, baseline intertriginous involvement, and baseline of the variable under analysis. Statistical comparisons between the treatment groups will be obtained using contrasts. The LS Means, standard errors, 95% confidence intervals, and p-values will be presented. These analyses will be performed on the ITT and mITT populations.

For the primary endpoint of IGA success (IGA of 'Clear' or 'Almost Clear' plus a 2-grade improvement from Baseline at Week 8), subgroup analysis will be performed on the following:

- Topical Corticosteroids – Inadequate Response, intolerance or contraindication (Y/N)
- Topical Vitamin D Derivatives – Inadequate Response, intolerance or contraindication (Y/N)
- Apremilast (Otezla) – subjects with any prior use/no prior use
- Psoriasis involvement on the elbow (yes)
- Psoriasis involvement on the knee (yes)
- Psoriasis involvement on the knee or elbow (either one)
- By site
- By age categories (2- 11, 12 -17, >= 18)

Subgroups based on baseline BSA severity, grouped into mild, moderate, and severe categories will be analyzed for the following variables:

- IGA and I-IGA success
- PASI-50, PASI-75, PASI-90
- investigator local tolerability and subject local tolerability assessments
- WI-NRS

where the BSA groups are defined as follows:

1. Mild BSA Severity: up to 5% BSA at baseline
2. Moderate BSA Severity: >=5-10% BSA at baseline
3. Severe BSA Severity: >=10% BSA at baseline

For the subgroup of subjects with intertriginous area involvement or face involvement (defined medical history assessment of psoriasis) analyses of the following variables will be performed:

- IGA and I-IGA success
- PASI-50, PASI-75, PASI-90
- Investigator local tolerability and subject local tolerability assessments
- WI-NRS

All analyses of these endpoints will be performed on the observed data with no imputation.

8.4. Patient Reported Outcomes

The continuous endpoints change from baseline in WI-NRS, change from baseline in PSD, change from baseline in total DLQI score, and change from baseline in total CDLQI will be analyzed using an analysis of covariance (ANCOVA: total score only) with the factors treatment, site, baseline IGA, baseline intertriginous involvement, and baseline of the variable under analysis. Statistical comparisons between the treatment groups will be obtained using contrasts.

The LS Means, standard errors, 95% confidence intervals, and p-values will be presented. These analyses will be performed on the ITT population.

Observed and change from baseline scores in WI-NRS itch severity, the total and individual PSD items, and the total and individual for both DLQI and CDLQI items will be summarized descriptively by treatment group and time point using continuous summary statistics. These summaries will be presented for both the ITT and mITT populations.

All analyses will be performed on the observed data with no imputation.

9. Safety and Tolerability Analysis

Safety will be evaluated from reported AEs, local tolerability assessments, changes in clinical laboratory values, changes in vital signs, C-SSRS, PHQ-8/modified PHQ-A and CDI-2 results.

All safety analyses will be performed on the Safety population.

9.1. Adverse Events

All AEs, TEAEs, and SAEs will be coded using the MedDRA dictionary v. 23.0.

An overall summary of TEAEs will be provided; this will present number and percent of subjects who reported at least 1: TEAE (including all TEAEs, TEAEs by maximum severity, and TEAEs by greatest relationship), SAE, discontinued IP due to a TEAE, or had a TEAE resulting in death.

The number and percent of subjects reporting treatment emergent AEs, grouped by MedDRA system organ class and preferred term, will be tabulated by severity or greatest relationship to study IP and treatment group. In the case of multiple occurrences of the same TEAE within the same subject, each subject will only be counted once for each preferred term.

In the summaries showing severity and relationship to study medication the event with the maximum severity (death related to AE > life-threatening consequences > severe > moderate > mild) or strongest relationship (unrelated < unlikely < possibly < probably < likely) will be reported. If a particular event is missing the severity and/or relationship, then the strongest possible severity or relationship will be assumed for analysis (severity = severe, relationship = likely). The number and percent of subjects reporting related treatment emergent AEs will be tabulated by preferred term and severity.

In the AE data listings, all AEs will be displayed. AEs that are not treatment-emergent will be flagged.

9.1.1. Adverse Events Leading to Withdrawal

A summary of incidence rates (frequencies and percentages) of TEAEs leading to withdrawal of IP, by treatment group, SOC, and preferred term will be prepared for the Safety Population. No inferential statistical tests will be performed.

A data listing of AEs leading to withdrawal of IP will also be provided, displaying details of the event(s) captured on the CRF.

9.1.2. Deaths and Serious Adverse Events

Any deaths that occur during the study will be listed.

Serious adverse events will be listed and also tabulated by system organ class and preferred term and presented by treatment as well as by relationship to IP and maximum severity.

9.1.3. Adverse Events of Special Interest (AESIs)

AESIs may be identified and multiple MedDRA preferred terms may be grouped together to calculate the subject incidence of adverse events of interest.

9.2. Local Tolerance Assessments

The investigator's assessment of the application site reaction will be summarized by visit using both categorical methods (number and percentage of subject with each score) as well as continuous methods (e.g., mean, standard deviation, etc.). No inferential statistical tests will be performed.

The subject's assessment of the application site reaction will be summarized similarly.

9.3. Clinical Laboratory Evaluations

Laboratory test results will be summarized descriptively by treatment and time point as both observed values and changes from baseline.

The number of subjects with clinical laboratory values below, within, or above the normal range by time point and in relation to baseline will be tabulated for each clinical laboratory analyte by treatment group (shift table).

9.4. Vital Signs

Descriptive summaries of observed values and changes from baseline will be calculated for body weight, height, BMI, systolic blood pressure, diastolic blood pressure, heart rate, and oral body temperature by treatment group and time point.

Shift table (overall, subjects who indicated weight loss was intentional and subjects weight loss was not intentional at each visit) by treatment group will summarize the number of subjects who gain or lose >=5% of their baseline body weight during the course of the study, as well as subjects who gain or lose >=10% of their baseline body weight over the course of the study.

BMI is derived as specified in Section **Error! Reference source not found.**. Shift tables (overall, subjects who indicated weight loss was intentional and subjects weight loss was not intentional at each visit) by treatment group for subjects who shift from their baseline BMI category (underweight, normal, overweight, obese) to a different BMI category throughout the course of the study will be provided by treatment group and visit.

9.5. PHQ and Modified PHQ-A

Data for PHQ-8 and Modified PHQ-A will be classified using each subject's total score at a time point into a category based on the following scoring system:

- None – Minimal depression (0 to 4)
- Mild depression (5 to 9)
- Moderate depression (10 to 14)
- Moderately severe depression (15 to 19)
- Severe depression (20 to 24)

Shift tables showing the category of severity at each time point by treatment group will be presented.

9.6. CDI-2

For Children's Depression Inventory 2(CDI-2), descriptive summaries of observed values and changes from baseline will be calculated for the total score and the 2 scales emotional problems and functional problems by treatment group and visit.

All CDI-2 data will be listed.

9.7. C-SSRS

The C-SSRS is a questionnaire that prospectively assesses Suicidal Ideation and Suicidal Behavior. At the Screening study visit, "Baseline/Screening" version of the C-SSRS will be used. This version assesses Suicidal Ideation and Suicidal Behavior during the subject's lifetime and during the past 6 months. For the Screening visit, "lifetime" experience of the subject with Suicidal Ideation and Suicidal Behavior will be summarized. From Baseline visit, the "Since Last Visit" version will be used.

Suicidality data collected on the C-SSRS will be listed for all subjects. Tables will include results from the Suicidal Ideation and Suicidal Behavior sections of the C-SSRS. Frequencies and percentages of subjects with a response of "Yes" at any point on the Suicidal Ideation and Suicidal Behavior items will be summarized by study visit and treatment group.

All C-SSRS data will be listed.

9.8. Physical Examination

The number and percentage of subjects with normal and abnormal findings in the physical examination will be displayed at each study visit and treatment group.

9.9. Concomitant Medication

Prior and concomitant medications will be summarized descriptively by treatment group, overall, ATC level 4, and preferred term using counts and percentages.

Prior medications will be presented separately from concomitant medications. Medications that started prior to the first application of IP will be considered prior medications whether or not they were stopped prior to the first application of IP. Any medications continuing or starting post the first application of IP will be considered to be concomitant. If a medication starts prior to the first application of IP and continues after the first application of IP it will be considered both prior and concomitant.

Medications will be coded using WhoDrug Global B3, vSep2019.

10. Changes from Planned Analysis

Per Protocol population is described in the protocol however that population and the associated analyses have been removed as they were considered to be uninformative.

Modified Intent-To-Treat Population has been added into account for COVID-19 disruption.

Additional exploratory endpoints have been added in the SAP which are not specified in the protocol. They are:

- For the primary endpoint of IGA success (IGA of 'Clear' or 'Almost Clear' plus a 2-grade improvement from Baseline at Week 8), subgroup analysis will be performed on the following, only summary statistics:
    - Topical Corticosteroids – Inadequate Response, intolerance or contraindication (Y/N)
    - Topical Vitamin D Derivatives – Inadequate Response, intolerance or contraindication (Y/N)
    - Prior Apremilast (Otezla) Use – subjects with any prior use / no prior use
    - Psoriasis involvement on the elbow (yes)
    - Psoriasis involvement on the knee (yes)
    - Psoriasis involvement on the knee or elbow (either one)
    - By site
    - By age categories (2- 11, 12- 17, >= 18)

- Subgroups based on baseline BSA severity, grouped into mild, moderate, and severe categories will be analyzed for the following variables, only summary statistics:
    - IGA and I-IGA success
    - PASI-50, PASI-75, PASI-90
    - investigator local tolerability and subject local tolerability assessments
    - WI-NRS

  where the BSA groups are defined as follows:

  1. Mild BSA Severity: up to 5% BSA at baseline
  2. Mod BSA Severity: >=5-10% BSA at baseline
  3. Severe BSA Severity: >=10% BSA at baseline

- For the subgroup of subjects with intertriginous area involvement or face involvement (defined medical history assessment of psoriasis) analyses of the following variables will be performed only summary statistics:
    - IGA and I-IGA success
    - PASI-50, PASI-75, PASI-90
    - Investigator local tolerability and subject local tolerability assessments
    - WI-NRS

- For PSD, DLQI and CDLQI individual questions will be summarized descriptively by treatment group and time point using both continuous summary statistics for the ITT population.
- Shift tables by treatment group for subjects who shift from their baseline BMI category (underweight, normal, overweight, obese) to a different BMI category throughout the course of the study will be provided by treatment group and visit.
- The number and percentage of subjects with normal and abnormal findings in the physical examination will be displayed at each study visit and treatment group.

- Tipping point analysis has been included as a sensitivity analysis for the primary endpoint
- The mITT population will also be used as a sensitivity analysis for the primary endpoint
- Analysis of covariance (ANCOVA) analyses of the following:
    - change and percent change from baseline in PASI
    - change from baseline in affected percent BSA
    - change and percent change from baseline in mPASI
    - change and percent change in WI-NRS
    - change and percent change in DLQI
    - change and percent change in CDLQI
- Achievement of PASI-90 (subjects who achieve a 90% reduction in PASI from Baseline) at week 8 was a planned secondary endpoint in the protocol and has been moved to an exploratory endpoint
- The testing of the secondary endpoints in section 6.1.3 has been modified from what is in in the protocol.
- Shift from baseline in Weight and BMI tables are repeated based on the intentional/nonintentional or missing weight loss question captured in the CRF at each visit.
- Incidence of TEAEs leading to study discontinuation by SOC and preferred term
- Incidence of related TEAEs by preferred term
-

11. Other Planned Analysis

11.1. Pharmacokinetic Analysis

Concentration data will be summarized by timepoint and treatment group using summary statistics.

Data will be listed in the data listing.

11.2. Intertriginous Subjects

Only a small percentage of subjects enrolled into the ARQ-151-301 and ARQ-151-302 studies have intertriginous involvement. Since these small numbers make drawing conclusions about this population difficult, it is planned to pool the data for the subjects with intertriginous involvement from these two studies together for a subgroup analysis of this population in the integrated summary of efficacy.

12. References

1. US Federal Register. (1998) International Conference on Harmonization; Guidance on

Statistical Principles for Clinical Trials. Department of Health and Human Services: Food and Drug Administration [Docket No. 97D-0174]. Federal Register Volume 63, Number 179, pages 49583-49598. September 16, 1998.

2. ASA. (2016) Ethical Guidelines for Statistical Practice. Prepared by the Committee on Professional Ethics, April 2016. http://www.amstat.org/about/ethicalguidelines.cfm
3. RSS. (2014) The Royal Statistical Society: Code of Conduct, 2014. http://www.rss.org.uk/Images/PDF/join-us/RSS-Code-of-Conduct-2014.pdf.
4. Ratitch, B., Lipkovich, I., & O'Kelly, M. (2013). *Combining Analysis Results from Multiply Imputed Categorical Data*. PharmaSUG. https://www.pharmasug.org/proceedings/2013/SP/PharmaSUG-2013-SP03.pdf

13. Tables, Listings, and Figures

All listings, tables, and graphs will have a header showing the sponsor company name and protocol and a footer showing the version of SAS, the file name and path, and the source of the data (CRF page or listing number).

13.1. Planned Table Descriptions

The following are planned summary tables for protocol number ARQ-151-301. The table numbers and page numbers are place holders only and will be determined when the tables are produced.

Table 3: Demographic Data Summary Tables and Figures

Table Number	Population	Table Title/Summary
13.1 Demographic Data Summary Tables and Figures		

13.2. Efficacy Data

Table 4: Efficacy Data

Table Number	Population	Table Title / Summary

13.3. Safety Data

Table 5: Safety Data

Table Number	Population	Table Title / Summary
		14.3.1 Displays of Adverse Events **14.3.2 Summary of Deaths, Other Serious and Significant Adverse Events**
		14.3.3 Narratives of Deaths, Other Serious and Certain Other Significant Adverse Events
		14.3.4 Abnormal Laboratory Value **14.3.5 Laboratory Data Summary Tables**
		14.3.6 Other Safety Data Summary Tables

13.4. Pharmacokinetic/Pharmacodynamic Data

Table 6: Pharmacokinetic/Pharmacodynamic Data

Table Number	Population	Table Title / Summary
		14.4 Pharmacokinetic and Pharmacodynamic Data Summary Tables

13.5. Other Data Summary Tables

Table 7: Other Data Summary Tables

Table Number	Population	Table Title / Summary

13.6. Planned Listing Descriptions

The following are planned data and patient/subject data listings for protocol number ARQ-151-301.

In general, one listing will be produced per CRF domain. All listings will be sorted by treatment, site, and subject number. All calculated variables will be included in the listings.

In all listings a blank line will be placed between each subject. Within a data listing, if an item appears line after line (eg, repetition of subject number), then only the first occurrence will be displayed.

In data listings, the information for one subject will be kept on one page if at all possible, rather than splitting a subject's information across pages.

Table 8: Planned Listings

Data Listing Number	Population	Data Listing Title / Summary
16.2 Patient/Subject Data Listings		
16.2.1 Patient/Subject Discontinuations/Completions		
16.2.2 Protocol Deviations		
16.2.3 Patients/Subjects Excluded from the Efficacy Analyses		
16.2.4 Demographic Data and Other Baseline Characteristics		
16.2.5 Compliance and/or Drug Concentration Data		
16.2.6 Individual Efficacy Response Data		
16.2.7 Adverse Event Listings (by Patient/Subject)		
16.2.8 Laboratory Values (by Patient/Subject)		
16.2.9 Other Clinical Observations and Measurements (by Patient/Subject)		
16.2.10 Other Study Measurements or Assessments (by Patient/Subject)		

13.7. Planned Figure Descriptions

The following are planned summary figures for protocol number ARQ-151-301. The figure numbers and page numbers are place holders only and will be determined when the figures are produced.

Table 9: Planned Figures

Figure Number	Population	Figure Title/Summary	
14.x.x			16.2.x.x

Appendix 1: Premier Research Library of Abbreviations

Abbreviation	Definition
AE	adverse event
ANCOVA	analysis of covariance
ATC	anatomical therapeutic chemical
BMI	body mass index
BSA	body surface area
CDISC	clinical data interchange standards consortium
CDLQI	Children's Dermatology Life Quality Index
CI	confidence intervals
CMH	Cochran-Mantel-Haenszel
COVID-19	novel coronavirus disease-19
CRF	case report form
CS	clinically significant
C-SSRS	Columbia Suicide Severity Rating Scale
DBP	diastolic blood pressure
DLQI	Dermatology Life Quality Index
eCRF	electronic case report form
EDC	electronic data capture
EMA	European medicines agency
ePRO	electronic patient report outcomes
FDA	food and drug administration

Abbreviation	Definition
HR	heart rate
ICH	international council for harmonization
IP	investigational product
IRT	interactive response technology
ITT	intent-to-treat
LS	least-squares
MedDRA	medical dictionary for regulatory activities
mPASI	Modified Psoriasis Scalp Severity Index
N	number
NA	not applicable
NCS	non-clinically significant
PASI	Psoriasis Scalp Severity Index
PD	protocol deviation
PE	physical examination
PHQ-A	Modified PHQ-9 for Adolescents
PHQ-8	Patient Health Questionnaire depression scale
PP	per-protocol
PSD	Psoriasis Symptoms Diary
RR	respiratory rate or relative rate
SAE	serious adverse event
SAP	statistical analysis plan

Abbreviation	Definition
SAS®	a software system used for data analysis
SBP	systolic blood pressure
SD	standard deviation
SOC	system organ class
TEAE	treatment-emergent adverse event
WHO	world health organization
WHO-DD	world health organization drug dictionary
WI-NRS	worst itch – numeric rating scale