Statistical Analysis Plan

[redacted]

Sponsor	Arcutis Biotherapeutics, Inc.
Protocol Title:	A Phase 2a, 8-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-154 Foam 0.3% Administered QD in Subjects with Seborrheic Dermatitis
Protocol Number:	ARQ-154-203
Premier Research PCN:	ARCU210100
Document Version:	Final 1.0
Document Date:	08-Sep-2020

Approvals

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

[redacted]

Document History

Not applicable.

[redacted]

Table of Contents

[redacted]

[redacted]

1. Overview

This statistical analysis plan (SAP) describes the planned analysis and reporting for Arcutis Biotherapeutics, Inc. protocol number ARQ-154-203 (A Phase 2a, 8-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-154 Foam 0.3% Administered QD in Subjects with Seborrheic Dermatitis), dated 07-Jan-2020 Amendment Version 1.0. 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 and included in the CSR 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 Biotherapeutics, Inc.'s study ARQ-154-203.

2. Study Objectives and Endpoints

2.1. Study Objectives

2.1.1. Primary Objective

The primary objective is to assess the safety and efficacy of ARQ-154 foam 0.3% (Roflumilast Foam 0.3%) administered once daily (QD) vs vehicle foam x 8 weeks in adult subjects with seborrheic dermatitis

2.2. Study Endpoints

2.2.1. Safety Endpoints

The safety endpoints of this study include the following:

- Local tolerability assessments
- Clinical laboratory parameters
- Adverse events (AEs)
- Columbia-Suicide Severity Rating Scale (C-SSRS)

- Patient Health Questionnaire depression scale (PHQ-8)
- Vital signs
- Physical examinations

2.2.2. Efficacy Endpoints

2.2.2.1. Primary Efficacy Endpoint

The primary efficacy endpoint of this study is achievement of an Investigator Global Assessment (IGA) score 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 an IGA score of "clear" or "almost clear" PLUS a 2-grade improvement from baseline at weeks 2 and 4.
- Change from baseline in Overall Assessment of Erythema score at weeks 2, 4, and 8.
- Change from baseline in Overall Assessment of Scaling score at weeks 2, 4, and 8.
- Achievement of an Overall Assessment of Erythema (0-3 scale) score of 0 or 1 PLUS a 2-grade improvement from baseline at weeks 2, 4, and 8.
- Achievement of an Overall Assessment of Scaling (0-3 scale) score of 0 or 1 PLUS a 2-grade improvement from baseline at weeks 2, 4, and 8.
- Change and percent change in Worst Itch – Numeric Rating Scale (WI-NRS) pruritus score at weeks 2, 4, and 8, as compared to baseline.
- In subjects with a baseline WI-NRS pruritus score of ≥4, achievement of a ≥4-point improvement from baseline in WI-NRS pruritus score at weeks 2, 4, and 8.

2.2.2.3. Exploratory Efficacy Endpoints

The exploratory efficacy endpoints of this study include the following:

- Change and percent change in Scalpdex total score from baseline at weeks 2, 4, and 8.
- Achievement of an Overall Assessment of Erythema (0-3 scale) score of 0 or 1 at weeks 2, 4, and 8.
- Achievement of an Overall Assessment of Scaling (0-3 scale) score of 0 or 1 at weeks 2, 4, and 8.
- Achievement of an Overall Assessment of Erythema (0-3 scale) score of 0 at weeks 2, 4, and 8.
- Achievement of an Overall Assessment of Scaling (0-3 scale) score of 0 at weeks 2, 4, and 8.
- Change and percent change from baseline in Dermatology Life Quality Index (DLQI) at weeks 2, 4, and 8.
- Change and percent change from baseline in body surface area (BSA) affected at week 8.
- A 2-grade improvement in IGA from Baseline.
- Incidence of a 2 point reduction from baseline in the WI-NRS (among subjects with WI-NRS ≥2 at baseline).
- Achievement of an IGA score of "clear" or "almost clear" PLUS a 2-grade improvement from Baseline at week 9.

- Change from baseline in Overall Assessment of Erythema score at week 9.
- Change from baseline in Overall Assessment of Scaling score at week 9.
- Achievement of an Overall Assessment of Erythema (0-3 scale) score of 0 or 1 PLUS a 2-grade improvement from baseline at week 9.
- Achievement of an Overall Assessment of Scaling (0-3 scale) score of 0 or 1 PLUS a 2-grade improvement from Baseline at week 9.

3. Overall Study Design and Plan

3.1. Overall Design

This is an 8-week, parallel group, double blind, vehicle-controlled study for the treatment of subjects with seborrheic dermatitis. This study will include both male and female adults having an IGA score of at least "Moderate" (3), and approximately 184 subjects will be enrolled. After having met all inclusion criteria, and none of the exclusion criteria, subjects will be randomized in a 2:1 ratio to ARQ-154 foam 0.3% QD (Roflumilast foam 0.3%) or Vehicle foam QD which will be applied to areas of lesions of seborrheic dermatitis. There will be screening for up to 4 weeks followed by 8 weeks of treatment phase. Subjects will have to apply the study drug once a day in the evening, except for on Day 0 and Week 2, the study drug is applied at the study site. Subjects have to record the date and time each dose has been applied, any missed doses, and any additional comments. There will be a follow-up visit approximately 1 week after treatment has been completed.

3.2. Sample Size and Power

A sample size of approximately 184 subjects is planned for the study. Subjects will be randomized in a 2:1 ratio to ARQ-154 foam 0.3% (Roflumilast Foam 0.3%): vehicle foam QD, stratified by study site and baseline disease severity. Approximately 121 subjects will receive ARQ-154 foam 0.3% QD (Roflumilast Foam 0.3%); approximately 63 subjects will receive vehicle foam QD. A sample size of 184 subjects will provide approximately 90% power to detect an active response of at least 58.5%, assuming 159 subjects complete the study and a vehicle response of 30%, based on two-group Chi-squared test of equal proportions (without continuity correction), using a 2-sided alpha of 0.10.

3.3. Study Population

Study population consists of male and female subjects of age 18 years or older with no more than 20% BSA of seborrheic dermatitis. Subjects should have a minimum IGA of 'Moderate' (3) at baseline.

3.4. Treatments Administered

Subjects who meet the eligibility criteria will be randomized to 1 of the 2 following treatment groups in a 2:1 ratio (active:vehicle):

- ARQ-154 foam 0.3% QD (Roflumilast Foam 0.3%)
- Matching vehicle foam QD.

[redacted]

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 according to a computer-generated randomization list. Randomization will be stratified by study site and baseline disease severity (IGA = 3 or IGA = 4).

3.6. Blinding and Unblinding

This is a double-blind study, therefore neither the subjects nor the Investigator, clinical personnel, or sponsor will be aware of which treatment an individual has received. Emergency unblinding will be done using the study internet-based randomization system (IWRS) system in consultation with the Medical Monitor and the sponsor's Chief Medical Officer (CMO).

3.7. Schedule of Events

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

[redacted]

Table 1: Schedule of Events

Study Procedure	Screen	Baseline Day 0	Wk 2 Day 14	Wk 4 Day 28	Wk 8 Day 56	Wk 9 Day 63
Visit	**1**	**2**	**3**	**4**	**5**	**6**
Visit Window	**-4 weeks**		**+/- 3 day**	**+/- 5 days**	**+/- 5 days**	**+/-5 days**
Informed consent/assent	X					
Medical history	X					
Physical examination[a]	X	X			X	
I/E criteria	X	X				
Randomization		X				
Hematology, Serum Chemistries, and Urine Analysis	X	X			X	
Vital signs, height, weight[b]	X	X	X	X	X	X
IGA[c], Overall Assessment of Erythema[c], Overall Assessment of Scaling[c]	X	X	X	X	X	X
WI-NRS, DLQI, Scalpdex	X	X	X	X	X	
BSA	X	X			X	
Application Site Reaction Assessment/Local Tolerability[d]		X		X	X	
Pigmentation Assessment[e]	X	X	X	X	X	X
C-SSRS, PHQ-8	X	X		X	X	
Medical Photography[f]		X	X	X	X	
Pregnancy test[g]	X	X		X	X	
PK draws[h]		X		X	X	
IP/vehicle application at the study site[i]		X	X			
Dispense study medication kit[j]		X	X	X		
Dispense/review diary		X	X	X	X	
Weigh study medication kit[k]		X	X	X	X	
Compliance calculation[l]		X	X	X	X	
Adverse event assessment	X	X	X	X	X	X
Concomitant medications	X	X	X	X	X	X

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

[b] Height will be collected at Baseline and Week 8. Weight will be collected at all study visits. 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% weight loss from Baseline should be reported to the medical monitor.

[c] IGA will be a 5-point scale ranging from clear (0) to severe (4). IGA should be completed prior to other physician assessments. Overall assessment of erythema (0-3 scale) and overall assessment of scaling (0-3 scale) will be completed.

[d] Local tolerability Assessments: The Investigator local tolerability assessment of skin irritation (Berger and Bowman skin irritation score) should be performed prior to the investigational product application at Baseline, and at Weeks 4 and 8. Reactions at the site of product application, which may occur post-Baseline, should be differentiated from the preexisting inflammation associated with the subject's seborrheic dermatitis. Subjects will perform the local tolerability

assessment 10-15 minutes post-drug application at Baseline, and recall assessments at Weeks 4 and 8 for the subject's '0-3' burning/stinging assessment.

[e] An assessment for hypopigmentation and hyperpigmentation will be performed by the investigator at all clinic visits.

[f] At selected sites, medical photography will be obtained for target lesions. 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.

[g] A pregnancy test will be administered to all females of child-bearing potential. A serum pregnancy test will be performed at the Screening visit only. A urine pregnancy test will be performed at Baseline, Week 4, and Week 8. 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.

[h] PK draws (trough / pre-dose) will be collected at Days 0, 28 and 56. At baseline, this draw will be pre-dose relative to drug application in the clinic. Ensure study medication is not applied in the area where PK will be drawn.

[i] Subjects to apply assigned IP at the study site at every designated visit, to confirm understanding of instruction on how to apply initially.

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

[k] The entire kit should be weighed and recorded at every visit. See IP Handling Manual for details.

[l] Compliance calculation is described in the IP Handling Manual.

[redacted]

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 for analyses conducted by Premier Research, the final statistical methodology 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.

Unless otherwise indicated, all statistical tests will be conducted at the 0.10 significance level using 2-tailed tests, and P values will be reported. Corresponding 90% confidence intervals (CIs) will be presented for statistical tests. In addition, 95% CIs have been added for the efficacy table summaries as requested by sponsor. Wilson confidence intervals for binomial proportions will be computed.

Statistical testing will be performed at the 0.10 level using two-tailed tests.

ARQ-154 0.3% foam will be described as “Roflumilast Foam 0.3%” throughout the tables, figures, and listings.

4.2. Interim Analysis and Data Monitoring

Based on protocol section 6.1.3, one interim futility analysis was planned when the first 60 subjects randomized and have had the opportunity to complete the 8 week disease assessment. This futility analysis will be nonbinding and will be used for decision making on further expansion of the clinical development program in seborrheic dermatitis. The sponsor considered that the analysis is no longer needed for decision making and interim analysis was not carried out upon their request. This interim analysis was not meant to change the accrual to or conduct of this trial.

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 1 confirmed dose of investigational product (IP). 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 used as sensitivity analysis of primary and secondary endpoints along with disposition.

- **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 the novel coronavirus disease (COVID-19) disruption. This population will be the primary analysis population for the analysis of efficacy endpoints.
- **Pruritus Population (PRU4):** The PRU 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 Population (PRU2):** The PRU population is a subset of the ITT population and includes subjects with WI-NRS pruritus score ≥2 at Baseline. This population will be used for the analysis of achievement of a 2-point reduction in WI-NRS pruritus score as compared to Baseline.
- **Pharmacokinetic Population (PK):** The PK population will include all subjects receiving the active drug with sufficient plasma concentrations of roflumilast to define a profile, as determined by the pharmacokineticist. This population will be used for analyses of PK parameters. Analyses using the PK Population will be presented in a separate PK report, provided by the pharmacokineticist.

6. General Issues for Statistical Analysis

6.1. Statistical Definitions and Algorithms

6.1.1. Baseline

The last observation recorded on or before the day of first dose of IP will be used as the baseline observation for all calculations of change from baseline for all the efficacy and vital signs data. For the laboratory and PHQ data, the last observation recorded before the first dose of IP will be used as the baseline observation for all calculations of change from 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.

Subgroup analyses may be generated for the baseline covariates.

6.1.3. Multiple Comparisons

No adjustments will be made for multiple comparisons; all analyses will be conducted at the $\alpha = 0.10$ level.

[redacted]

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.

For the primary efficacy endpoint of IGA score, 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, 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. MCMC method will use seed of 878508. 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
$\leq 2\%$	1
$> 2\%$ to $\leq 5\%$	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, baseline IGA score, treatment group, and investigational site at baseline using a seed of 633621. 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. The results obtained will be analyzed using the Cochran-Mantel-Haenszel (CMH) analysis for each of the complete analysis data sets stratified by baseline IGA score and site. The results will be combined into one multiple imputation inference (odds ratio, associated confidence interval and *P* value) using PROC MIANALYZE as illustrated[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. Other missing data will not be imputed, with the exception of incomplete dates as described in Section 6.1.8. For responder analysis, only observed data will be used. Only observed data will be summarized using descriptive statistics.

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

Step 1:

```
proc mi data=example seed=878508 nimpute=XX round=1 out=example_1;
   mcmc impute=monotone;
   var <baseline score> …… <visit8 score>;

run;
```

Step 2:

```
proc mi data=example_1 seed=633621 nimpute=XX  out=example_2;
  class <treatment> <site>;
  monotone regpmm(<baseline score> …… <visit8 score>);
  var <treatment> <site> <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 pooled site and baseline iga score on each completed dataset and combining the results using proc mianalyze.

```
proc freq data=example noprint;
 by <imputationnumber> <visit> ;
 tables <site>*<baseline score>*<treatment>*<outcome>/ cmh alpha=0.1;
 output out=example_stat cmh;

run;
```

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

[redacted]

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

For Week 9 visit, only lower limit will be used and if the study day falls on or after the lower limit, the unscheduled or early termination visit would be set to week 9 as appropriately.

Visit Name	Visit Number	Target Start Day	Lower Limit	Upper Limit
Week 2	3	14	2	22
Week 4	4	28	23	42
Week 8	5	56	43	61
Week 9	6	63	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. 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 1 subject in each treatment group.

6.1.7. Derived Variables

- **IGA success =** IGA of "Clear" or "Almost Clear" plus a 2-grade improvement from Baseline.
- **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.
    - **Number of expected IP applications** = calculated as last treatment date - first treatment date + 1.
    - Treatment end date = for subjects that completed the study it is calculated as week 8 date obtained from COMP CRF. If subjects are missing week 8 date, the last available date from SV will be used. If subject discontinued from study, the last available COMP date will be used. In the absence of date in COMP page, the last available date from SV with visit performed will be used.
    - **Number of IP applications =** number of expected IP applications – missed IP applications as collected in the CRF.

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

- **BMI (kg/m^2)** = (weight in kg)/[(height in cm/100)^2]. For Week 2, and 4, baseline height will be used to derive BMI. For Week 8, and 9, Week 8 height will be used to derive BMI 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

- **WI-NRS 4-point reduction** = achievement of a 4-point reduction in WI-NRS pruritus score at Weeks 2, 4, and 8 compared to baseline, calculated only for subjects with a pruritus score of ≥ 4 at baseline.

- **WI-NRS 2-point reduction** = achievement of a 2-point reduction in WI-NRS pruritus score at Weeks 2, 4, and 8 compared to baseline, calculated only for subjects with a pruritus score of ≥ 2 at baseline.

- **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, No=0, with Not relevant recorded to 0; 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.

- **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.

- **Scalpdex score transformation =** Scalpdex is rated on a 1 to 5 scale which will be transformed to 0 to 100 Scale where 1=0; 2=25; 3=50; 4=75; 5=100. This transformed score is used to calculate scale scores.

    - **Emotions Scale =** average of (Q2, Q4, Q5, Q6, Q7, Q9, Q10, Q11, Q12, Q14, Q16, Q17, Q19, Q20, Q22) after transforming to 0 to 100 scale as mentioned above.

      Q refers to question number. Q19 will be reverse scored i.e., 1=100; 2=75; 3=50; 4=25; 5=0.

    - **Symptoms Scale =** average of (Q1, Q3, Q8) after transforming to 0 to 100 scale as mentioned above. Q refers to question number.

- **Functioning Scale =** average of (Q13, Q15, Q18, Q21, Q23) after transforming to 0 to 100 scale as mentioned above. Q refers to question number.

- **Scalpdex Total Score =** calculated as mean of all the 23 scalpdex questions using the transformed scale of 0 to 100. Q19 will be reverse scored i.e., 1=100; 2=75; 3=50; 4=25; 5=0 while calculating the mean. Q refers to question number.

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

- **TEAE** = any AE with an onset date/time after the first application of IP.

- **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).

6.1.8. Data Adjustments/Handling/Conventions

All collected data will be presented in 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 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; similarly, if a *P* value greater than 0.9999 occurs it will be shown in tables as >0.9999.

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

A treatment-related AE is any AE with a relationship to the study drug of possibly, probably, likely, or missing.

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 (if present) is 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 telemedicine; 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 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
- Scalpdex
- C-SSRS
- PHQ-8
- Subject Local Tolerability

The following assessments cannot be completed via telemedicine/remotely:

- IGA
- BSA
- Investigator Local Tolerability
- Overall Assessment of Erythema
- Overall Assessment of Scaling
- Pigmentation Assessment

- 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 Subjects and Withdrawals

Disposition will include tabulations of the number of subjects randomized into each treatment group, the number of subjects who received treatment, subjects completing the study, tabulated reasons for discontinuation from the study overall and due to COVID-19 disruption, and number of subjects in each analysis population. Disposition will be summarized for all subjects who were entered into database by treatment group and overall.

7.2. Protocol Deviations

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

7.3. Demographics and Other Baseline Characteristics

Summary statistics for age, gender (including child-bearing potential), race, ethnicity, height, weight, baseline disease characteristics (IGA, Overall Assessment of Erythema, Overall Assessment of Scaling, Scalpdex, WI-NRS), percent BSA, and BMI will be presented by treatment group and overall.

A summary of treatment history, including history of response, intolerance, or contraindication to topical corticosteroids and/or topical antifungals will be provided.

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

For ordinal variables such as the IGA, and WI-NRS, summary statistics including the mean, median, and range of the ordinal variable will be presented as well as frequency counts of each level of the ordinal variable.

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

These analyses will be conducted for the mITT and Safety populations.

[redacted]

7.4. Exposure and Compliance

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

The amount of IP used by each subject based on can weight will be summarized descriptively by treatment group 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 IP application period and does not miss more than 3 consecutive doses.

Investigational product application compliance will be calculated based on number of applications divided by the expected number of IP applications. Compliance will be summarized descriptively by treatment group using the following categories:

> 100%

≥ 80% - ≤100%

< 80%

8. Efficacy Analysis

8.1. Primary Efficacy Analysis

The IGA is an ordinal scale with five severity grades which is reported only in integers.

Table 3 illustrates the description of each severity grade.

Table 3: IGA

Score	Description
0	**Completely clear:** No erythema, no scaling (hypo-hyperpigmentation can be present)
1	**Almost clear:** Residual slight erythema and/or trace amounts of scaling
2	**Mild:** Pink to red color and/or slight scaling
3	**Moderate:** Distinct redness and/or clearly visible scaling
4	**Severe:** Severe erythema (intense, fiery red) and/or severe scaling (coarse, thick scales with flaking onto clothes or skin)

For this study, the primary estimand is the odds ratio of achieving IGA success at 8 weeks; that is, the ratio of the odds of achieving IGA success at 8 weeks ARQ-154 (Roflumilast Foam 0.3%) relative to the odds of success at 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. However, the COVID-19 pandemic related issues may impact the integrity of the interpretation of the primary endpoint analysis; thus the "Principle Stratum Strategy" has been adopted to handle subjects impacted by COVID-19 related issues by removing them from the analysis entirely; thus the stratum of subjects included in this analysis are those who either have a Week 8 visit or were discontinued for a reason other than COVID-19 related issues. Discontinuations for any other reasons or other known or unknown intercurrent events will be handled using the "Treatment Policy Strategy" where the ITT principle will serve as the analytical basis for interpreting the estimand. In other words, the odds ratio of achieving IGA success for ARQ-154 (Roflumilast Foam 0.3%) relative to vehicle at 8 weeks will be evaluated regardless of the occurrence of any such intercurrent event, with the exception of subjects who are unable to attend the week 8 visit due to COVID-19 disruption. The study database will collect whether or not subjects missed a visit specifically due to COVID-19 disruption. It is the sponsor's assumption that data missing due to COVID-19 disruption are MCAR, and therefore, the mITT population will be no different than the ITT population for the purpose of generating this result. Under this assumption, the use of the mITT population will provide an unbiased estimate and facilitate the interpretation of the study in the presence of COVID-19 disruption. 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 effects of two classification factors – investigative site and baseline disease severity.

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 CMH test stratified by site (throughout all analyses that use stratification, it will be based on the pooling algorithm described in Section 6.1.6) and baseline IGA. Statistical significance will be concluded at the 10% significance level (2-sided). Additionally, both 90% and 95% Wilson CIs for proportion of successes in each treatment group will be presented.

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 CI and *P* value).

Sensitivity analyses of the primary endpoint will also be performed using the original (non-imputed) dataset. These will include a repeated measures logistic regression model (GEE) with IGA success as the dependent variable and treatment, site, baseline IGA score, and visit as the independent variables, as well as the above described CMH test on the non-imputed data.

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 will be based on mITT population and these analyses will be repeated for the ITT population.

If subjects were unable to attend the week 8 visit due to COVID-19 related issues they will be excluded from the primary analyses for efficacy, which is handled by the mITT population.

8.2. Secondary Efficacy Analysis

All secondary efficacy analyses will be performed using the mITT and ITT populations.

8.2.1. IGA

Achievement of an IGA score of "clear" or "almost clear" PLUS a 2-grade improvement from Baseline at weeks 2 and 4 will be analyzed similar to the primary efficacy endpoint using CMH test stratified by site , and baseline IGA with the exception that missing data will not be imputed. In addition, the number and percentage of subjects in each category will be summarized by treatment and visit. Additionally, both 90% and 95% Wilson CIs for proportion of successes in each treatment group will be presented.

8.2.2. Overall Assessment of Erythema

The assessment is performed at Screening, Baseline, and Weeks 2, 4, 8, and 9. The score is reported in ordinal scale with severity grades ranging from 0-3.

- Analysis of change and percent change from baseline in Overall Assessment of Erythema at Weeks 2, 4, and 8 will be performed using descriptive summaries (mean, median, inter quartile range) by treatment group and study visit. The treatment groups will be compared using an exact Wilcoxon rank-sum test as continuous variable. In addition, the number and percentage of subjects in each category will be summarized by treatment and study visit.
- Achievement of an Overall Assessment of Erythema (0-3 scale) score of 0 or 1 PLUS a 2-grade improvement from Baseline at Weeks 2, 4, and 8 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. Additionally, both 90% and 95% Wilson CIs for proportion of successes in each treatment group will be presented.

Table 4 illustrates description of erythema grades.

Table 4: Erythema Grades

Score	Description
0	**None:** No evidence of erythema
1	**Mild:** Barely perceptible erythema which is faint or patchy
2	**Moderate:** Distinct erythema,
3	**Severe:** Intense (fiery red) erythema

[redacted]

8.2.3. Overall Assessment of Scaling

This assessment is performed at Screening, Baseline, and Weeks 2, 4, 8, and 9. The score is reported in ordinal scale with severity grades ranging from 0-3.

- Analysis of change and percent change from baseline in Overall Assessment of Scaling score at weeks 2, 4, and 8 will be performed using descriptive summaries (median, inter quartile range) by treatment group and study visit. The treatment groups will be compared using an exact Wilcoxon rank-sum test as continuous variable.
- In addition, the number and percentage of subjects in each category will be summarized by treatment and visit. Achievement of an Overall Assessment of Scaling (0-3 scale) score of 0 or 1 PLUS a 2-grade improvement from Baseline at weeks 2, 4, and 8 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 replaced. Additionally, both 90% and 95% Wilson CIs for proportion of successes in each treatment group will be presented.

Table 5 illustrates description of scaling grades.

Table 5: Scaling Grades

Score	Description
0	**None:** No scaling evident on lesions
1	**Mild:** Barely detectable, scattered, small flaking scales
2	**Moderate:** Scales clearly visible and prominent
3	**Severe:** Coarse, thick scales, with flaking into clothes or skin

8.2.4. WI-NRS

WI-NRS scale ranges from 0 to 10 with 0 being "no itch" and 10 equaling "worst imaginable itch". This will be determined by asking the subject's assessment of worst itch over the past 24 hours.

- Analysis of change from baseline in WI-NRS pruritus score at Weeks 2, 4, and 8 will be performed using descriptive summaries (mean, median, inter quartile range) by treatment group and study visit. In addition, the number and percentage of subjects for each scale score will be summarized by treatment and visit.
- Analysis of change from baseline in WI-NRS pruritus score at Weeks 2, 4, and 8 will be analyzed using an analysis of covariance with the factors treatment, study site (grouped as specified in Section 6.1.6), baseline IGA, and baseline value of the respective scale as independent variables. The LS Means, standard errors, 90% CIs, 95% CIs, and P values for change from baseline and testing difference in treatments will be presented.
- In subjects with a baseline WI-NRS pruritus score of ≥4, achievement of a ≥4-point improvement from baseline in WI-NRS pruritus score at Weeks 2, 4, and 8 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 replaced. This will be summarized using PRU4 population. Additionally, both 90% and 95% Wilson CIs for

proportion of successes in each treatment group will be presented.

8.3. Exploratory Efficacy Analysis

Exploratory efficacy analysis will be performed for subjects belonging to mITT and ITT populations.

The exploratory endpoints include:

- Change and percent change in Scalpdex total score from baseline at Weeks 2, 4, and 8
- Achievement of an Overall Assessment of Erythema (0-3 scale) score of 0 or 1 at Weeks 2, 4, and 8
- Achievement of an Overall Assessment of Scaling (0-3 scale) score of 0 or 1 at Weeks 2, 4, and 8
- Achievement of an Overall Assessment of Erythema (0-3 scale) score of 0 at Weeks 2, 4, and 8
- Achievement of an Overall Assessment of Scaling (0-3 scale) score of 0 at Weeks 2, 4, and 8
- Change and percent change from baseline in DLQI at Weeks 2, 4, and 8
- Change and percent change from baseline in BSA affected at Week 8
- A 2-grade improvement in IGA from Baseline.
- Incidence of a 2 point reduction from baseline in the WI-NRS (among subjects with WI-NRS ≥ 2 at baseline
- Achievement of an IGA score of "clear" or "almost clear" PLUS a 2-grade improvement from Baseline at week 9.
- Change from baseline in Overall Assessment of Erythema score at week 9.
- Change from baseline in Overall Assessment of Scaling score at week 9.
- Achievement of an Overall Assessment of Erythema (0-3 scale) score of 0 or 1 PLUS a 2-grade improvement from baseline at week 9.
- Achievement of an Overall Assessment of Scaling (0-3 scale) score of 0 or 1 PLUS a 2-grade improvement from Baseline at week 9.

Scalpdex questionnaire consists of 23 questions which are categorized into emotions, symptoms and functioning scales. The scale scores and total score are calculated as described in the Section 6.1.7. Change and percent change from baseline for Scalpdex scale scores along with total score are summarized descriptively. In addition, number and percentages for each Scalpdex question is summarized by study visit and treatment group.

Change and percent change from baseline for DLQI, and BSA will be summarized descriptively. In addition change from baseline values for Scalpdex scale scores and total score, DLQI, and BSA will be analyzed using an analysis of covariance with the factors treatment, study site (grouped as specified in Section 6.1.6), baseline IGA, and baseline value of the respective scale as independent variables. Statistical comparisons between the treatment groups will be obtained using contrasts. The LS Means, standard errors, 90% CIs, 95% CIs, and P values for change from baseline and testing difference in treatments will be presented.

Analysis of 2 point reduction from baseline in the WI-NRS (among subjects with WI-NRS ≥ 2 at baseline) will be conducted similarly as described in Section 8.2.4; i.e., CMH test stratified by site and baseline IGA will be presented for Weeks 2, 4, and 8 for PRU2 population.

For efficacy analysis purposes, the data that is collected by ERT will be used for BSA summarization.

9. Safety and Tolerability Analysis

Safety will be evaluated from reported AEs, physical examinations, local tolerability assessments, changes in clinical laboratory values, changes in vital signs/weight, C-SSRS, and PHQ-8 results.

All safety analyses will be performed on the Safety population.

9.1. Adverse Event

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

A treatment emergent AE (TEAE) is defined as an AE that started post application of IP at the Baseline visit through study completion. 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 the study due to a TEAE, or had a TEAE resulting in death.

The number and percent of subjects reporting TEAEs, 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 (mild < moderate < severe) or strongest relationship (not related < related) 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).

In the AE data listings, all AEs will be displayed. AEs that are 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 study drug, 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 study drug 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 (SAEs) will be listed and tabulated by system organ class and preferred term and presented by treatment group, severity, and relationship to study treatment.

9.2. Local Tolerability Assessments

The investigator's assessment of the application site reaction will be summarized by study visit using both categorical methods (number and percentage of subject with each score) as well as continuous methods (e.g., mean, median, etc.). Categorical summaries will be provided for dermal response as well as other effects. 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 study visit as both observed values and change from baseline values for continuous hematology, chemistry, and urinalysis results. Categorical urinalysis results will be summarized using frequencies by study visit and treatment.

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

Abnormal results will be flagged in the listings. In addition, pregnancy test results and hormonal laboratory results will be listed.

9.4. Vital Signs

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

Changes in weight 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 by treatment group and visit.

BMI is derived as specified in Section 6.1.7. 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.

Upon sponsor's request, shift in BMI, and changes in weight are summarized by intentional and unintentional or missing weight loss question captured in the CRF.

[redacted]

9.5. PHQ-8

Data for PHQ-8 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 visit by treatment group will be presented.

9.6. 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.

9.7. 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.

9.8. Concomitant Medication

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

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

Medications will be coded using WHODrug Global B3, version September 2019.

[redacted]

10. Changes from Planned Analysis

Based on protocol amendment, the significance level has been changed from 0.05 to 0.10. However, protocol Section 6.3.1 indicates alpha = 0.05 but section 6.1 indicates significance level to be 0.10. The alpha level of 0.10 is the intended level, and this SAP uses 0.10 as significance level as documented throughout. The PRU4 and PRU2 populations have been added to facilitate WI-NRS analysis. These populations are not included in the protocol. Additional exploratory endpoints have been added in the SAP which are not specified in the protocol. They are:

- Incidence of a 2 point reduction from baseline in the WI-NRS (among subjects with WI-NRS ≥ 2 at baseline
- Achievement of an IGA score of "clear" or "almost clear" PLUS a 2-grade improvement from Baseline at week 9
- Change from baseline in Overall Assessment of Erythema score at week 9
- Change from baseline in Overall Assessment of Scaling score at week 9
- Achievement of an Overall Assessment of Erythema (0-3 scale) score of 0 or 1 PLUS a 2-grade improvement from baseline at week 9
- Achievement of an Overall Assessment of Scaling (0-3 scale) score of 0 or 1 PLUS a 2-grade improvement from Baseline at week 9.
- In addition, percentage change from baseline summaries have been added for Scalpdex, DLQI, BSA and WI-NRS.

Based on the sponsor's request the following changes have been made:

- 95% CIs have been added in the efficacy table summaries along with 90% CIs as the alpha level is 0.1.
- If a subject is unable to attend the week 8 visit due to COVID-19, they will be excluded from the primary analyses for efficacy. To facilitate this, mITT population has been added to replace the ITT population.
- The ITT population will be used as a sensitivity analysis of the primary, secondary, and exploratory efficacy analysis along with disposition.
- ANCOVA analysis has been added for change in baseline scores of Erythema, Scaling and WI-NRS.
- A new BMI shift table has been added.
- Shift from baseline in Weight and BMI tables are repeated based on the intentional/nonintentional or missing weight loss question captured in the CRF.
- For efficacy analysis purposes, data collected from ERT will be used for BSA summarization.
- Per Protocol population and associated analyses have been removed as they were considered to be uninformative.
- ANCOVA analysis have been added for change from baseline scores for WI-NRS and BSA.
- Wilcoxon rank sum test has been added to summarize the continuous data associated with Overall Assessment of Erythema and Scaling.

An interim analysis was planned to make a decision on further expansion of the clinical development program for seborrheic dermatitis. The sponsor considered the interim analysis as no longer needed for decision making and therefore no interim analysis was conducted. The interim analysis was not meant to change enrollment or conduct of this trial.

11. Other Planned Analysis

11.1. Pharmacokinetic Analysis

All PK collection information from the eCRF will be presented in a listing.

Concentration data and pk parameters will be summarized by timepoint (concentrations data) and treatment group using summary statistics. In addition, PK concentration and parameter listings will be presented as well.

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 TLFs 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 (e.g., listing number).

13.1. Planned Table Descriptions

The following are planned summary tables for protocol number ARQ-154-203. The table numbers and page numbers are place holders only and will be determined when the tables are produced.

[redacted]

13.1.1. Demographic Data

Table 6: Demographic Data Summary Tables

Number	Population	Title	Topline
Table 14.1.1	All Randomized	Summary of Subject Disposition	X
Table 14.1.2.1	Safety	Summary of Demographics and Baseline Characteristics	X
Table 14.1.2.2	mITT	Summary of Demographics and Baseline Characteristics	
Table 14.1.3	Safety	Previous Treatment History of Seborrheic Dermatitis	X
Table 14.1.4	Safety	Summary of Protocol Deviations	
Table 14.1.5	Safety	Summary of Prior Medications by ATC Class Level 4 and Preferred Term	
Table 14.1.6	Safety	Summary of Study Drug Exposure and Compliance	

13.1.2. Efficacy Data

Table 7: Efficacy Data Summary Tables

Number	Population	Title	Topline
Table 14.2.1.1	mITT	Summary and Change from Baseline in Investigator Global Assessment (IGA) Grades by Study Visit – Observed Data	
Table 14.2.1.2	ITT	Summary and Change from Baseline in Investigator Global Assessment (IGA) Grades by Study Visit – Observed Data	
Table 14.2.1.3	mITT	Summary and Change from Baseline in Investigator Global Assessment (IGA) Grades by Study Visit – Multiple Imputation	X
Table 14.2.1.4	ITT	Summary and Change from Baseline in Investigator Global Assessment (IGA) Grades by Study Visit – Multiple Imputation	
Table 14.2.1.6	mITT	Summary of Investigator Global Assessment (IGA) Grades and Success by Study Visit Categorical Results – Observed Data	

[redacted]

Number	Population	Title	Topline
Table 14.2.1.7	ITT	Summary of Investigator Global Assessment (IGA) Grades and Success by Study Visit Categorical Results – Observed Data	
Table 14.2.1.9	mITT	Summary of Investigator Global Assessment (IGA) Grades and Success by Study Visit – Multiple Imputation Categorical Results	X
Table 14.2.1.10	ITT	Summary of Investigator Global Assessment (IGA) Grades and Success by Study Visit – Multiple Imputation Categorical Results	
Table 14.2.1.11	mITT	Summary of Investigator Global Assessment (IGA) Grades and Success by Study Visit Tabulation of Fitted Point Estimates from Generalized Estimating Equations for Binary Response – Observed Data	
Table 14.2.1.12	ITT	Summary of Investigator Global Assessment (IGA) Grades and Success by Study Visit Tabulation of Fitted Point Estimates from Generalized Estimating Equations for Binary Response – Observed Data	
Table 14.2.1.13	mITT	Summary of Investigator Global Assessment (IGA) – 2-grade Improvement from Baseline – Observed Data	
Table 14.2.2.1	mITT	Summary and Change from Baseline in Overall Assessment of Erythema – Severity Grades by Study Visit – Wilcoxon Test	
Table 14.2.2.2	mITT	Summary of Overall Assessment of Erythema – Severity Grades and Success by Study Visit Categorical Results	X
Table 14.2.2.3	ITT	Summary and Change from Baseline in Overall Assessment of Erythema – Severity Grades by Study Visit	
Table 14.2.2.4	ITT	Summary of Overall Assessment of Erythema – Severity Grades and Success by Study Visit Categorical Results	

[redacted]

Number	Population	Title	Topline
Table 14.2.2.5	mITT	Summary of Overall Assessment of Erythema – Success Score of 0 or 1 and Success Score of 0 by Study Visit	
Table 14.2.2.6	ITT	Summary of Overall Assessment of Erythema – Success Score of 0 or 1 and Success Score of 0 by Study Visit	
Table 14.2.2.7	mITT	Summary and Change from Baseline in Overall Assessment of Erythema – Severity Grades by Study Visit – Wilcoxon Test	
Table 14.2.2.8	ITT	Summary and Change from Baseline in Overall Assessment of Erythema – Severity Grades by Study Visit – Wilcoxon Test	
Table 14.2.3.1	mITT	Summary and Change from Baseline in Overall Assessment of Scaling – Severity Grades by Study Visit	
Table 14.2.3.2	mITT	Summary of Overall Assessment of Scaling – Severity Grades and Success by Study Visit Categorical Results	X
Table 14.2.3.3	ITT	Summary and Change from Baseline in Overall Assessment of Scaling – Severity Grades by Study Visit	
Table 14.2.3.4	ITT	Summary of Overall Assessment of Scaling – Severity Grades and Success by Study Visit Categorical Results	
Table 14.2.3.5	mITT	Summary of Overall Assessment of Scaling – Success Score of 0 or 1 and Success Score of 0 by Study Visit	
Table 14.2.3.6	ITT	Summary of Overall Assessment of Scaling – Success Score of 0 or 1 and Success Score of 0 by Study Visit	
Table 14.2.3.7	mITT	Summary and Change from Baseline in Overall Assessment of Scaling – Severity Grades by Study Visit – Wilcoxon Test	
Table 14.2.3.8	ITT	Summary and Change from Baseline in Overall Assessment of Scaling – Severity Grades by Study Visit – Wilcoxon Test	
Table 14.2.4.1	mITT	Summary and Change from Baseline in Worst Itch – Numeric Rating Scale (WI-NRS) Pruritus Score by Study Visit	X
Table 14.2.4.2	ITT	Summary and Change from Baseline in Worst Itch – Numeric Rating Scale (WI-NRS) Pruritus Score by Study Visit	

[redacted]

Number	Population	Title	Topline
Table 14.2.4.3	PRU4-mITT	Worst Itch – Numeric Rating Scale (WI-NRS) Pruritus Score Success by Study Visit	X
Table 14.2.4.4	PRU2-mITT	Worst Itch – Numeric Rating Scale (WI-NRS) Pruritus Score Success by Study Visit	
Table 14.2.4.5	PRU4-ITT	Worst Itch – Numeric Rating Scale (WI-NRS) Pruritus Score Success by Study Visit	
Table 14.2.4.6	PRU2-ITT	Worst Itch – Numeric Rating Scale (WI-NRS) Pruritus Score Success by Study Visit	
Table 14.2.4.7	mITT	Summary of Worst Itch – Numeric Rating Scale (WI-NRS) by Study Visit - ANCOVA	
Table 14.2.4.8	ITT	Summary of Worst Itch – Numeric Rating Scale (WI-NRS) by Study Visit - ANCOVA	
Table 14.2.4.9	mITT	Summary of Worst Itch - Numeric Rating Scale (WI-NRS) by Study Visit - Categorical Scale	
Table 14.2.4.10	ITT	Summary of Worst Itch - Numeric Rating Scale (WI-NRS) by Study Visit - Categorical Scale	
Table 14.2.5.1	mITT	Change from Baseline in Scalpdex by Study Visit	
Table 14.2.5.2	mITT	Summary of Scalpdex by Study Visit - ANCOVA	
Table 14.2.5.3	ITT	Change from Baseline in Scalpdex by Study Visit	
Table 14.2.5.4	ITT	Summary of Scalpdex by Study Visit - ANCOVA	
Table 14.2.5.5	mITT	Categorical Summary of Individual Responses by Study Visit – Scalpdex Questionnaire	
Table 14.2.5.6	ITT	Categorical Summary of Individual Responses by Study Visit – Scalpdex Questionnaire	
Table 14.2.6.1	mITT	Change from Baseline in Dermatology Life Quality Index (DLQI) by Study Visit	
Table 14.2.6.2	mITT	Summary of Dermatology Life Quality Index (DLQI) by Study Visit – ANCOVA	
Table 14.2.6.3	ITT	Change from Baseline in Dermatology Life Quality Index (DLQI) by Study Visit	
Table 14.2.6.4	ITT	Summary of Dermatology Life Quality Index (DLQI) by Study Visit – ANCOVA	

Number	Population	Title	Topline
Table 14.2.7.1	mITT	Change from Baseline in Body Surface Area (BSA) by Study Visit	
Table 14.2.7.2	mITT	Summary of Body Surface Area (BSA) by Study Visit – ANCOVA	
Table 14.2.7.3	ITT	Change from Baseline in Body Surface Area (BSA) by Study Visit	
Table 14.2.7.4	ITT	Summary of Body Surface Area (BSA) by Study Visit – ANCOVA	

13.1.3. Safety Data

Table 8: Safety Data Summary Tables

Number	Population	Title	Topline
14.3.1 Displays of Adverse Events			
Table 14.3.1.1	Safety	Summary of Treatment Emergent Adverse Events	X
Table 14.3.1.2	Safety	Incidence of Treatment Emergent Adverse Events by System Organ Class and Preferred Term	
Table 14.3.1.3	Safety	Incidence of Treatment Emergent Adverse Events by System Organ Class, Preferred Term, and Maximum Severity	
Table 14.3.1.4	Safety	Incidence of Treatment Emergent Adverse Events by System Organ Class, Preferred Term, and Relationship to Study Drug	
Table 14.3.1.5	Safety	Incidence of Treatment Emergent Adverse Events Leading to Study Drug Discontinuation by System Organ Class and Preferred Term	
14.3.2 Summary of Deaths, Other Serious and Significant Adverse Events			
Table 14.3.2.1	Safety	Incidence of Serious Adverse Events by System Organ Class and Preferred Term	
Table 14.3.2.2	Safety	Incidence of Serious Adverse Events by System Organ Class, Preferred Term, and Relationship to Study Drug	
Table 14.3.2.3	Safety	Incidence of Serious Adverse Events by System Organ Class, Preferred Term, and Maximum Severity	
14.3.3 Narratives of Deaths, Other Serious and Certain Other Significant Adverse Events			
Table 14.3.3.1	Safety	Listing of Adverse Events Leading to Study Drug Discontinuation	
Table 14.3.3.2	Safety	Listing of Serious Adverse Events	
Table 14.3.3.3	Safety	Listing of Deaths	
14.3.4 Abnormal Laboratory Value			
NA			

Number	Population	Title	Topline
14.3.5 Laboratory Data Summary Tables			
Table 14.3.5.1.1	Safety	Summary of Serum Chemistry Laboratory Results by Study Visit	
Table 14.3.5.1.2	Safety	Shift from Baseline in Clinical Chemistry Laboratory Results by Study Visit	
Table 14.3.5.2.1	Safety	Summary of Hematology Laboratory Results by Study Visit	
Table 14.3.5.2.2	Safety	Shift from Baseline in Hematology Laboratory Results by Study Visit	
Table 14.3.5.3.1	Safety	Summary of Quantitative Urinalysis Laboratory Results by Study Visit	
Table 14.3.5.3.2	Safety	Shift from Baseline in Quantitative Urinalysis Laboratory Results by Study Visit	
Table 14.3.5.3.3	Safety	Summary of Qualitative Urinalysis Laboratory Results by Study Visit	
14.3.6 Other Safety Data Summary Tables			
Table 14.3.6.1	Safety	Shift from Baseline in Patient Health Questionnaire (PHQ-8) by Study Visit	
Table 14.3.6.2.1	Safety	Summary of Investigator Local Tolerability Assessment by Study Visit	
Table 14.3.6.2.2	Safety	Summary of Investigator Local Tolerability Assessment (Dermal Response) by Study Visit Categorical Results	
Table 14.3.6.2.3	Safety	Summary of Investigator Local Tolerability Assessment (Other Effects) by Study Visit Categorical Results	
Table 14.3.6.3.1	Safety	Summary of Subject Local Tolerability Assessment by Study Visit	
Table 14.3.6.3.2	Safety	Summary of Subject Local Tolerability Assessment by Study Visit Categorical Results	X
Table 14.3.6.4.1	Safety	Summary of Vital Signs by Study Visit	
Table 14.3.6.4.2	Safety	Shift from Baseline in Weight by Study Visit	
Table 14.3.6.4.3	Safety	Shift from Baseline in BMI by Study Visit	
Table 14.3.6.4.4	Safety	Shift from Baseline in Weight by Study Visit – by Weight Loss Intentional/Non-Intentional	

[redacted]

Number	Population	Title	Topline
Table 14.3.6.4.5	Safety	Shift from Baseline in BMI by Study Visit – by Weight Loss Intentional/Non-Intentional	
Table 14.3.6.5	Safety	Summary of Columbia-Suicide Severity Rating Scale (C-SSRS) Suicidal Ideation and Suicidal Behavior by Study Visit	
Table 14.3.6.6	Safety	Summary of Physical Examination by Study Visit	
Table 14.3.6.7	Safety	Summary of Concomitant Medications by ATC Class Level 4 and Preferred Term	
14.4 **Pharmacokinetic and Pharmacodynamic Summary Tables**			
14.4.1.1	PK	Summary of Pharmacokinetic Results by Study Visit	
14.4.1.2	PK	Summary of Pharmacokinetic Parameters by Study Visit	

[redacted]

13.2. Planned Listing Descriptions

The following are planned data and patient/subject data listings for protocol number ARQ-154-203.

In general, one listing will be produced per CRF domain. All listings will be sorted by site, and subject number. All calculated variables will be included in the listings. Screen failures will only be presented in Listing 16.2.2.1.

In all listings a blank line will be placed between each subject. Within a data listing, if an item appears line after line (e.g., 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 9: Planned Listings

Number	Population	Title
16.2.1 Subject Discontinuations/Completions		
Listing 16.2.1.1	All Subjects	Subject Disposition
Listing 16.2.1.2	All Subjects	Subject Visits
16.2.2 Protocol Deviations		
Listing 16.2.2.1	All Subjects	Inclusion and Exclusion Criteria Not Met
Listing 16.2.2.2	All Subjects	Protocol Deviations
16.2.3 Subjects Excluded from the Efficacy Analyses		
Listing 16.2.3	All Subjects	Analysis Populations
16.2.4 Demographic Data and Other Baseline Characteristics		
Listing 16.2.4.1	All Subjects	Subject Demographics
Listing 16.2.4.2	All Subjects	Medical History
16.2.5 Compliance and/or Drug Concentration Data		
Listing 16.2.5.1	All Subjects	Study Drug Application at the Study Site
Listing 16.2.5.2	All Subjects	Cans
Listing 16.2.5.3	All Subjects	Diary Dispensation
Listing 16.2.5.4	All Subjects	Compliance (CRF)
Listing 16.2.5.5	All Subjects	Missed Doses
Listing 16.2.5.6	PK	Pharmacokinetic Blood Collection
Listing 16.2.5.7	PK	Pharmacokinetic Calculated Parameters
16.2.6 Individual Efficacy Response Data		
Listing 16.2.6.1	All Subjects	Investigator Global Assessment (IGA)

[redacted]

Number	Population	Title
Listing 16.2.6.2	All Subjects	Overall Assessment of Erythema
Listing 16.2.6.3	All Subjects	Overall Assessment of Scaling
Listing 16.2.6.4	All Subjects	Worst Itch Numerical Rating Scale (WI-NRS)
Listing 16.2.6.5	All Subjects	Scalpdex
Listing 16.2.6.6	All Subjects	Dermatology Life Quality Index (DLQI)
Listing 16.2.6.7	All Subjects	Body Surface Area (BSA) including Scalp
16.2.7 Adverse Event Listings (by Subject)		
Listing 16.2.7.1	All Subjects	Adverse Events
16.2.8 Laboratory Values (by Subject)		
Listing 16.2.8.1.1	All Subjects	Clinical Laboratory Data: Clinical Chemistry
Listing 16.2.8.1.2	All Subjects	Clinical Laboratory Data: Hematology
Listing 16.2.8.1.3	All Subjects	Clinical Laboratory Data: Urinalysis
Listing 16.2.8.1.4	Female Subjects	Clinical Laboratory Data: Serum and Urine Pregnancy Test
16.2.9 Other Clinical Observations and Measurements (by Subject)		
Listing 16.2.9.1	All Subjects	Investigator Local Tolerability Assessments
Listing 16.2.9.2	All Subjects	Subject Local Tolerability Assessments
Listing 16.2.9.3	All Subjects	Pigmentation Assessment
Listing 16.2.9.4.1	All Subjects	Vital Signs
Listing 16.2.9.4.2	All Subjects	Vital Signs - Weight
Listing 16.2.9.5	All Subjects	Physical Examination
Listing 16.2.9.6	All Subjects	Medical Photography
Listing 16.2.9.7	All Subjects	Prior and Concomitant Medications
Listing 16.2.9.8	All Subjects	Patient Health Questionnaire (PHQ-8)
Listing 16.2.9.9	All Subjects	Columbia-Suicide Severity Rating Scale (C-SSRS)

[redacted]

14. Tables, Listings, and Listing Shells

14.1. Standard Layout for all Tables, Listings, and Figures

The following standard layout will be applied to all TLFs in support of this study. Note that programming notes may be added if appropriate after each TLF shell.

[redacted]

Appendix 1: Premier Research Library of Abbreviations

Abbreviation	Definition
aCRF	annotated case report form
AD	associated documents
ADR	adverse drug reactions
AE	adverse event
AESI	adverse events special interest
ANCOVA	analysis of covariance
ATC	anatomical therapeutic chemical
BLQ	beneath limit of quantification
BMI	body mass index
BRD	business requirements document
BSL	biostatistician lead

[redacted]

Abbreviation	Definition
CCGs	CRF completion guidelines
CD	compact disc
CDISC	clinical data interchange standards consortium
CEC	central ethics committee
CFR	code of federal regulations
CI	confidence intervals
CIOMS	council for international organizations of medical sciences
CIP	clinical investigational plan
CM	clinical manager
CMH	Cochran Mantel Haenszel
CMP	clinical monitoring plan
COV	close out visit
COVID-19	Novel coronavirus disease
CRA	clinical research associate

[redacted]

Abbreviation	Definition
CRF	case report form
CRO	contract research organization
CS	clinically significant
CSM	clinical supply manager
CSR	clinical study report
CTA	clinical trial administrator
CTM	clinical trial manager
CTMS	clinical trial management system
DB	database
DBL	database lock
DBP	diastolic blood pressure
DCRF	data change request form
DDE	drug dispensing error form
DEA	drug enforcement administration

[redacted]

Abbreviation	Definition
DIA	drug information association
DIS	data integration specification
DLT	dose limiting toxicity
DM	data management
DMB	data monitoring board
DMC	data monitoring committee
DML	data management lead
DMP	data management plan
DNA	deoxyribonucleic acid
DOB	date of birth
DS	document specialist
DSG	drug safety group
DSM	drug supply management (drug distributor)
DSMB	data safety monitoring board

[redacted]

Abbreviation	Definition
DSP	data safety plan
DSUR	development safety update report
DTS	data transfer specification
DVD	digital video disk
EC	ethics committee
ECD	edit check and derivation specifications
ECG	electrocardiogram
eCRF	electronic case report form
EDC	electronic data capture
EMA	European medicines agency
eTMF	electronic trial master file
EU	European Union
FA	full analysis
FDA	food and drug administration

[redacted]

Abbreviation	Definition
FMP	file management plan
FPFV	first patient first visit
FPI	first patient in
GCP	good clinical practice
GMP	good manufacturing practices
GPV	global pharmacovigilance
HR	heart rate
IB	investigator's brochure
IC or ICF	informed consent or informed consent form
ICH	international council for harmonization
ID	identification
IDM	independent drug monitoring
IEC	independent ethics committee
IGA	investigator global assessment

[redacted]

Abbreviation	Definition
IM	investigator meeting
IMV	interim monitoring visit
IND	investigational new drug
INDSR	investigational new drug safety reports
IP	investigational product
IRB	institutional review board
IRF	inventory release file
IRR	infusion related reactions
IRT	interactive response technology
ISF	investigator site file
ITT	intent-to-treat
IVRS	interactive voice response system
IWRS	interactive web response system
IxRS	interactive voice/web response system

[redacted]

Abbreviation	Definition
KPI	key performance indicator
LAN	local area network
LDM	lead data manager
LMS	learning management system
LLOQ	lower limit of quantification
LPI	last patient in
LPLV	last patient last visit
LPO	last patient out
MAAP	medical affairs and pharmacovigilance teams
MAH	marketing authorization holder
MedDRA	medical dictionary for regulatory activities
MHRA	medicines and healthcare products regulatory agency
MM	medical monitor
MMP	medical monitoring plan

[redacted]

Abbreviation	Definition
MMRM	mixed effect model repeat measurement
MTD	maximum tolerated dose
MVR	monitoring visit report
N	number
NA	not applicable
NCS	non-clinically significant
NF	non-functional
PD	protocol deviation
PDGP	protocol deviation guidance plan
PE	physical examination
PHQ	patient health questionnaire
PI	principal investigator
PIN	personal identification number
PK	pharmacokinetic

[redacted]

Abbreviation	Definition
PKAP	pharmacokinetic analysis plan
PM	project manager
PMP	project management plan
PP	per-protocol
PRIMS	Premier Research information management system
PS	project specialist
PV	pharmacovigilance
PVG	pharmacovigilance group
QA	quality assurance
QARC	quality assurance, risk and compliance
QC	quality control
QOL	quality of life
ROT	record of training
RR	respiratory rate or relative rate

[redacted]

Abbreviation	Definition
RSM	regional site monitor
SAE	serious adverse event
SAP	statistical analysis plan
SAS®	a software system used for data analysis
SBP	systolic blood pressure
SC	study coordinator
SCR	software change request
SD	standard deviation
SDS	study design specifications
SDTM	study data tabulation model
SDV	source data verification
SECC	self-evident correction conventions
SECP	self-evident correction plan
SF	screen failure

[redacted]

Abbreviation	Definition
SFT or SFTP	secure file transfer or secure file transfer plan
SIV	site initiation visit
SLA	service level agreement
SMP	safety management plan
SOC	system organ class
SOP	standard operating procedure
SOW	statement of work
SQV	site qualification visit
SUA	start-up associate
SUSAR	suspected, unexpected, serious adverse (drug) reaction
TA	trial assistant
TEAE	treatment-emergent adverse event
TMF	trial master file
TOM	task ownership matrix

[redacted]

Abbreviation	Definition
UAT	user acceptance testing
USA	United States of America
UTC	universal coordinated time
WAN	wide area network
WAR	work at risk
WG	working guideline
WHO	world health organization
WHO-DD	world health organization drug dictionary
WI-NRS	worst itch – numeric rating scale