Statistical Analysis Plan

Study Title:	**A Phase 3, Multicenter, Open-Label Extension Study of the Long-Term Safety of ARQ-151 Cream 0.15% and ARQ-151 Cream 0.05% in Subjects with Atopic Dermatitis**
Protocol Number and Version:	**ARQ-151-313, Original dated 20 December 2020** **ARQ-151-313, Amendment 1 dated 17 September 2021**
Product:	**ARQ-151 Cream 0.15% or 0.05%**
Sponsor:	**Arcutis Biotherapeutics, Inc.** **3027 Townsgate Road, Suite 300** **Westlake Village, CA 91361**
Date:	**20-Dec-2022**
Version:	**Final V1.0**
Prepared by:	**Innovaderm Research Inc.** **3530 boul. Saint-Laurent, suite 300,** **Montreal, Quebec, Canada, H2X 2V1**

STATISTICAL ANALYSIS PLAN REVISION SUMMARY			
Version	**Version Date**	**Author**	**Summary of Changes**
Original V1.0	20-Dec-2022		Initial version

This statistical analysis plan will be reviewed and revised as needed. The most recent approved version will replace the previous version in place.

SIGNATURE PAGE

AUTHOR:

Name	Title	Signature	Date (DD-MMM-YYYY)
	Principal Biostatistician, Innovaderm Research		

APPROVALS:

Name	Title	Signature	Date (DD-MMM-YYYY)
	Associate Director, Biostatistics Innovaderm Research		
	Vice President, Biometrics Arcutis Biotherapeutics, Inc.		
	Vice President, Clinical Development Arcutis Biotherapeutics, Inc.		

TABLE OF CONTENT

ABBREVIATIONS

AD	Atopic Dermatitis
AE	Adverse Event
ATC	Anatomical Therapeutic Chemical
BIW	Twice Weekly
BMI	Body Mass Index
BSA	Body Surface Area
CDI-2	Children's Depression Inventory 2
CDLQI	Children's Dermatology Life Quality Index
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
COVID-19	Coronavirus Disease-19
CRF	Case Report Form
CRO	Contract Research Organization
C-SSRS	Columbia-Suicide Severity Rating Scale
CSR	Clinical Study Report
DFI	Dermatitis Family Impact
DLQI	Dermatology Life Quality Index
EAIR	Exposure-adjusted Incidence Rate
EASI	Eczema Area and Severity Index
eCRF	Electronic Case Report Form
EDC	Electronic Data Capture
EM	Expectation-Maximization
ET	Early Termination
HR	Heart Rate
IP	Investigational Product
MCMC	Markov-Chain Monte-Carlo
MedDRA	Medical Dictionary for Regulatory Activities
NRI	Non-Responder Imputation
OLE	Open-Label Extension
PHQ-8	Patient Health Questionnaire-8
PHQ-A	Modified PHQ-9 for Adolescents
PMM	Predictive Mean Matching
POEM	Patient-Oriented Eczema Measure
PK	Pharmacokinetic
PT	Preferred Term
Q1	First Quartile
Q3	Third Quartile
QD	Once Daily
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System®

SCORAD	Scoring Atopic Dermatitis
SD	Standard Deviation
SE	Standard Error
SOC	System Organ Class
TEAE	Treatment Emergent Adverse Event
TESAE	Treatment Emergent Serious Adverse Event
TLF	Tables, Listings, and Figures
WHO-DD	World Health Organization Drug Dictionary
WI-NRS	Worst Itch - Numeric Rating Score
vIGA-AD	validated Investigator Global Assessment scale for Atopic Dermatitis

1 INTRODUCTION

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for Arcutis Biotherapeutics, Inc. clinical protocol ARQ-151-313. The analyses described in the SAP are based upon the protocol Amendment 1 dated 17 September 2021. In case of changes (note that any such changes are described in section 3.6 below) between the protocol and the SAP, the SAP will be used to guide the statistical analysis. Deviations from the SAP are not expected. If any, they will be described and justified in the final Clinical Study Report (CSR), as appropriate.

This version of SAP has been developed and finalized prior to ARQ-151-313 database lock for the first Interim Analysis (IA).

Each IA will be performed only after the following steps have been completed: approval of this SAP, clinical trial data up to a pre-specified data cut-off date are entered into the database for subjects from parent studies ARQ-151-311 and ARQ-151-312 and any discrepancies in the data are resolved, finalization of the classification as important or non-important of each protocol deviation that occurred before or on the pre-specified data cut-off date for subjects from parent studies ARQ-151-311 and ARQ-151-312, and the database is locked. That is, subjects from parent study ARQ-151-315 will not be included in the IAs.

Similarly, the final analysis will be performed only after the following steps have been completed: approval of this SAP, clinical trial data are entered into the database for all subjects and any discrepancies in the data are resolved, finalization of the classification as important or non-important of each protocol deviation, and the database is locked.

ARQ-151 Cream 0.15% and ARQ-151 Cream 0.05% will be described as “roflumilast cream 0.15%” and “roflumilast cream 0.05%” throughout this document and the tables, listings, and figures (TLFs).

2 STUDY OBJECTIVES AND ENDPOINTS

OBJECTIVES	ENDPOINTS
Safety	
To assess long-term safety in subjects with atopic dermatitis (AD) treated with Roflumilast cream 0.15%	Primary safety endpoints: • Treatment-emergent Adverse Events (TEAEs) • Serious Adverse Events (SAEs)

OBJECTIVES	ENDPOINTS
once daily (QD; subjects ≥6 years of age) or Roflumilast cream 0.05% QD (subjects 2 to 5 years of age) after completing one of three Phase 3 studies (ARQ-151-311 or ARQ-151-312 or ARQ-151-315).	Other safety endpoints: • Changes and percent change in clinical laboratory results • Changes and percent change in vital signs • Local tolerability assessments • Change from baseline of Patient Health Questionnaire depression scale (PHQ-8) and Modified PHQ-9 for Adolescents (PHQ-A) total score • Change from baseline of Children's Depression Inventory 2nd Edition (CDI-2) total score • Columbia-Suicide Severity Rating Scale (C-SSRS)
Efficacy	
To evaluate the long-term efficacy in subjects with AD treated with roflumilast cream 0.15% QD (subjects ≥6 years of age) or roflumilast cream 0.05% QD (subjects 2 to 5 years of age) after completing one of three Phase 3 studies (ARQ-151-311 or ARQ-151-312 or ARQ-151-315).	Secondary efficacy endpoints: • Validated Investigator Global Assessment-Atopic Dermatitis (vIGA-AD) score of clear (0) or almost clear (1) at each assessment • vIGA-AD success (defined as vIGA-AD score of clear (0) or almost clear (1) plus at least a 2-grade improvement from baseline) at each assessment • In subjects ≥ 12 years old at the start of the parent study, change and percent change from baseline in Worst Itch - Numeric Rating Score (WI-NRS) over time [a] • Change and percent change from baseline in Eczema Area and Severity Index (EASI) total score over time Other efficacy endpoints: • Duration of the first twice weekly (BIW) dosing interval • Cumulative time (days) on BIW dosing

OBJECTIVES	ENDPOINTS
	• Percent of time on BIW dosing • Duration (days) of the first disease control interval • Cumulative time (days) with disease control • Percent of time with disease control • Duration (days) of the first vIGA-AD score of Clear (0) • Cumulative time with vIGA-AD score of Clear (0) • Percent of time with vIGA-AD score of Clear (0) • Duration (days) of the first vIGA-AD score of Clear (0) or Almost Clear (1) • Cumulative time with vIGA-AD score of Clear (0) or Almost Clear (1) • Percent of time with vIGA-AD score of Clear (0) or Almost Clear (1) • Change and percent change from baseline in vIGA-AD score over time • In subjects ≥ 12 years old at the start of the parent study with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS over time [a] • Achievement of at least a 50%, 75%, 90%, 100% reduction in the EASI total score (EASI-50, EASI-75, EASI-90, EASI-100) over time • Change and percent change from baseline Body Surface Area (BSA) affected by AD over time • Change and percent change from baseline in Dermatology Life Quality Index (DLQI), Children's Dermatology Life Quality Index (CDLQI), Infants' Dermatitis Quality of Life (IDQOL) over time • Change and percent change from baseline in the Dermatitis Family Impact (DFI) over time • Change and percent change from baseline in the Scoring Atopic Dermatitis (SCORAD) over time

OBJECTIVES	ENDPOINTS
	• Change and percent change from baseline in the Patient-oriented Eczema Measure (POEM) over time.
Pharmacokinetic	
To assess the systemic exposure of roflumilast and its N-oxide metabolite	• Plasma concentrations of roflumilast and its N-oxide metabolite

[a] Subjects < 12 years old will be included in subgroup analyses (refer to Section 6.6).

3 STUDY DESIGN

3.1 Overall Design

This is a Phase 3, open-label, single-arm, long-term safety study in which roflumilast cream 0.15% or roflumilast cream 0.05% is applied QD for 24 (Cohort-Week 24) or 52 (Cohort-Week 52) weeks to subjects with mild to moderate AD (except for time on BIW maintenance therapy).

Eligible subjects will enroll into the long-term safety study on the same day as the Week 4 visit of the preceding study (ARQ-151-311 or ARQ-151-312 or ARQ-151-315).

Subjects/caregivers will apply roflumilast cream 0.15% for subjects ≥ 6 years old or roflumilast cream 0.05% for subjects 2-5 years old.

If a 5 year-old subject turns 6 years-old during this study, the subject will be switched from roflumilast cream 0.05% to roflumilast cream 0.15% at his/her first scheduled post-birthday visit. Dose reduction back to 0.05% is not permitted.

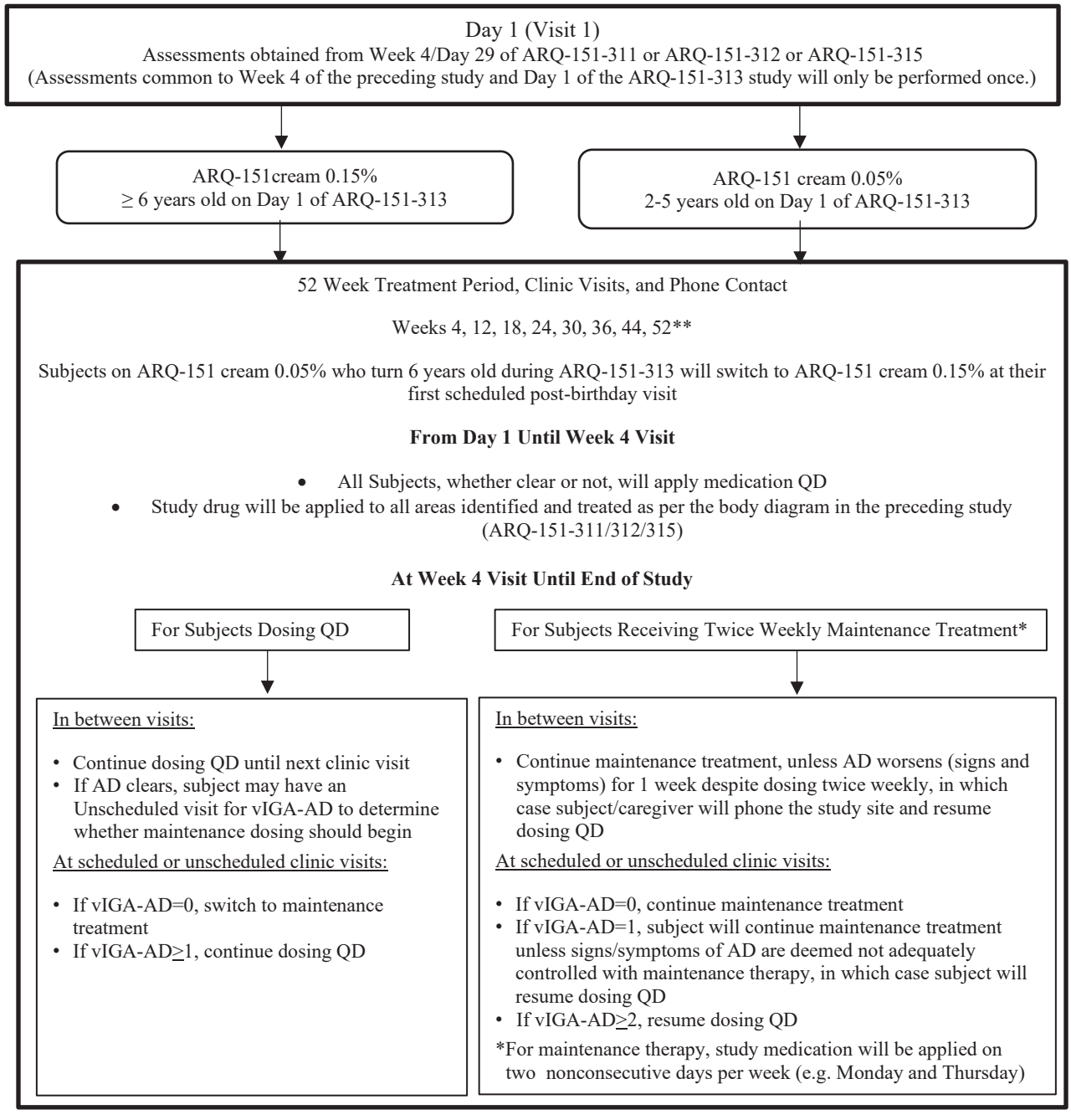
The study drug will be applied QD for 24 or 52 weeks (Cohort-Week 24 and Cohort-Week 52, respectively) to all AD affected areas and any newly appearing AD lesions that arise during the study, except on the scalp (unless entering BIW maintenance therapy, as described below).

For the first 4 weeks of study ARQ-151-313, all subjects will apply study medication QD in the areas identified and treated as per the body diagram in the preceding study (ARQ-151-311/312/315). Beginning at the Week 4 visit, any subject who achieves vIGA-AD of '0-clear' as confirmed by the Investigator at a scheduled or unscheduled clinic visit, will switch to BIW maintenance treatment. For maintenance therapy, study medication will be applied on two nonconsecutive days per week (e.g., Monday and Thursday) to areas commonly and/or most recently affected by AD, and any areas where AD is developing. Subjects receiving maintenance

treatment who experience AD worsening (signs and symptoms) for 1 week despite dosing twice weekly will phone the clinical study site, resume daily dosing, and complete the subject diary accordingly. The phone call (but not a clinic visit) is required to resume daily dosing. At clinic visits, subjects receiving maintenance treatment may continue twice weekly maintenance dosing if vIGA-AD is either '0-clear' or '1-almost clear'. Subjects will resume daily dosing if vIGA-AD ≥ 2 and can resume daily dosing if signs/symptoms of AD are not adequately controlled with maintenance therapy despite vIGA-AD of '1-almost clear'.

Cohort-Week 24 was created under Amendment 1, to allow enrollment of additional subjects 6-17 years of age that have completed the 4-week treatment period in ARQ-151-311 or ARQ-151-312. Subjects in this cohort will receive treatment for 24 weeks with ARQ-151 cream 0.15% (Cohort-Week 24).

3.2 Study Schema



A Phase 3, Multicenter, Open-Label Extension Study of the Long-Term Safety of ARQ-151 Cream 0.15% and ARQ-151 Cream 0.05% in Subjects with Atopic Dermatitis

Up to approximately 1500 subjects that have completed the 4-week treatment period in studies ARQ-151-311 or ARQ-151-312 or ARQ-151-315 and are willing and qualified to enroll in this long-term extension study will receive ARQ-151 cream 0.15% (ARQ-151-311/312 roll-overs) or ARQ-151 cream 0.05% (ARQ-151-315 roll-overs).

** Under Amendment 1, 24 Week Treatment Period for Cohort-Week 24, Clinic Visits, and Phone Contact Weeks 4, 12, 18, 24

3.3 Schedule of Events

Table 1 **Schedule of Visits and Assessments**

Study Procedure	Day 1	Wk 4	Wk 12	Wk18	Wk 24[t]	Wk 30	Wk 36	Wk 44	Wk 52/ET
Visit #	**1 (Week 4) of ARQ-151-311 or ARQ-151-312 or ARQ-151-315**	**2**	**3**	**Phone Visit**	**4**	**Phone Visit**	**5**	**Phone Visit**	**6**
Visit Window		**+/- 3 days**	**+/- 7 days**	**+/- 7 days**	**+/- 7 days**	**+/- 7 days**	**+/- 7 days**	**+/- 7 days**	**+/- 7 days**
Informed consent/assent	X								
Physical examination[a]	X[s]	X			X				X
I/E criteria	X								
Hematology, Serum Chemistries, and Urine Analysis[b]	X[s]				X				X
Vital signs, height, weight[c]	X[s]	X	X		X		X		X
vIGA-AD, EASI, BSA, SCORAD[d]	X[s]	X	X		X		X		X
WI-NRS pruritus[e]	X[s]								
POEM[f]	X[s]	X	X		X		X		X
Local Tolerability Assessment[g]	X	X	X		X		X		X
CDI-2, PHQ-8, PHQ-A, C-SSRS[h]	X[s]	X	X		X		X		X
DLQI, CDLQI, IDQOL, DFI[i]	X[s]	X	X		X		X		X
Medical Photography[j]	X[s]	X	X		X		X		X
Urine pregnancy test[k]	X[s]	X	X		X		X		X
PK draws[l]					X				X
Dispense study medication kit[m]	X	X	X		X		X		

Study Procedure	Day 1	Wk 4	Wk 12	Wk18	Wk 24[t]	Wk 30	Wk 36	Wk 44	Wk 52/ET
Visit #	**1 (Week 4) of ARQ-151-311 or ARQ-151-312 or ARQ-151-315**	**2**	**3**	**Phone Visit**	**4**	**Phone Visit**	**5**	**Phone Visit**	**6**
Visit Window		**+/- 3 days**	**+/- 7 days**	**+/- 7 days**	**+/- 7 days**	**+/- 7 days**	**+/- 7 days**	**+/- 7 days**	**+/- 7 days**
Drug application and Parents/Caregivers training in clinic	X								
Dispense/review diary[n]	X	X	X		X		X		X
Maintenance Dosing Determination[o]		X	X		X		X		
Weigh study medication[p]	X	X	X		X		X		X
Compliance Determination[q]		X	X		X		X		X
Adverse event assessment[r]	X	X	X	X	X	X	X	X	X
Concomitant medications	X	X	X	X	X	X	X	X	X
Study Exit					X[t]				X

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

[b] Hematology, Serum Chemistries, and Urine Analysis will only be done in subjects ≥ 12 years of age at time of enrollment in the preceding study.

[c] Height will be collected at Day 1 only for subjects ≥18 years old and at every clinic visit for subjects <18 years old. Weight should be obtained using a calibrated weight scale and the same scale should be used for a subject throughout the duration of the study. The subject should remove shoes and heavy clothing (sweaters or jackets), and empty pockets. The subject should stand with both feet in the center of the scale with their arms at their side and hold still. Record the weight to the nearest decimal fraction (for example, 25.1 kilograms). For subjects <18 years of age, measure the weight in triplicate and report the average weight in EDC. A 5% or greater weight loss (whether or not intentional or otherwise explained) should be reported to the medical monitor.

[d] The vIGA-AD assessment will be a 5-point scale ranging from clear (0) to severe (4) and is evaluated for the entire body except the scalp, palms, and soles. EASI takes into account overall severity of erythema, infiltration/papulation, excoriation, and lichenification, in addition to extent of BSA affected. The 4 clinical signs will be graded on a 4-point scale (0 [absent] to 3 [severe]) for 4 body regions (head and neck, upper extremities, lower extremities, and trunk). Total EASI score will be calculated as a sum of scores of all 4 body regions. EASI total score will range from 0 (absent) to 72 (severe). Total BSA affected by AD will be determined for all body surfaces except the scalp, palms and soles. vIGA-AD should be completed prior to other physician assessments. SCORAD total score will range between 0 and 103.

[e] The WI-NRS pruritus questionnaire will be completed once a week at home utilizing a diary. **The WI-NRS will be self-reported by subjects for subjects ≥ 6 years of age, and reported by parent/guardian for subjects <6 years of age.** During the Phone Visits at Weeks 18, 30, and 44 a reminder will be given to subjects to complete the WI-NRS pruritis questionnaire once a week.

[f] POEM will be completed by all subjects either by self or by proxy completion (for children unable to read and/or understand the POEM questionnaire, the parent/guardian/caregiver will complete the questionnaire).

[g] At Day 1 only, local tolerability assessments should be recorded prior to study drug application for the Investigator assessment of skin irritation (Berger and Bowman skin irritation score) and 10-15 minutes post-drug application for the Subject's '0-3' burning/stinging assessment. **Note for investigator tolerability assessments: reactions at the site of product application should be differentiated from the preexisting inflammation associated with the subject's atopic dermatitis. For subject tolerability assessments after Day 1, the subject will assess burning/stinging (0-3 score) 10-15 minutes post drug application based on recall of the last drug application.** For subjects <6 years of age, parents will complete.

[h] Adolescents and adults (12 years and older) will complete the C-SSRS (12 years old of age and older). Adults will complete the PHQ-8. Adolescents (ages 12 to 17 years of age, inclusive) will complete the PHQ-A (PHQ-9 modified for Adolescents). Parents/caregivers will complete the parent report (CDI-2) for children 6-11 years of age, inclusive. Subjects will continue with the assessments specific to their age group at the time of consent/assent at Screening of the preceding study.

[i] The DLQI will be completed by subjects ≥17 years of age (based on age at start of preceding studies ARQ-151-311/312). The CDLQI will be completed for subjects ≥ 4 years of age and ≤16 years of age (based on age at start of preceding studies ARQ-151-311/312/315). The Infants' Dermatitis Quality of Life (IDQOL) will be completed by parents/caregivers for subjects < 4 years (based on age at start of parent preceding studies ARQ-151-315). The Dermatitis Family Impact Questionnaire (DFI) will be completed by parents/caregivers for all subjects ≥ 2 years of age and ≤17 years of age. Subjects will continue with the assessments specific to their age group at the time of consent/assent at Screening of the preceding study. DFI is not to be conducted for subjects who were ≤17 years of age in the ARQ-151-311/312 studies and turned 18 years of age prior to enrollment into ARQ-151-313.

[j] Photography will be performed using Canfield equipment on all subjects at all study 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 study drug at each visit.

[l] For all subjects ≥12 years old entering this study under Amendment 1 a single PK trough draw will be collected at Week 52/ET for Cohort-Week 52 or Week 24/ET for Cohort 24 subjects. Ensure study medication was not applied in the area where PK will be drawn.

[m] It is expected that kits will be dispensed based on %BSA affected and subject's age. Subjects who turn 6 years old during the study will switch to ARQ-151 cream 0.15% at their first scheduled post-birthday visit. See IP Handling Manual for details. Under Amendment 1, this is the final visit for Cohort-Week 24 and no kits will be dispensed at this visit.

[n] Dispensing of IP is allowed, if needed. Study site should review IP kit to ensure sufficient IP is available until the next visit and only dispense additional IP if needed. **All subjects will continue study medication application in the areas identified and treated as per the body diagram in the preceding study for the first 4 weeks of the current study (ARQ-151-313).**

[o] Beginning at the Week 4 visit, any subject who achieves vIGA-AD of '0-clear' as confirmed by the Investigator at a scheduled or unscheduled clinic visit, will switch to twice weekly (BIW) maintenance treatment. For maintenance therapy, study medication will be applied on two nonconsecutive days per week (e.g. Monday and Thursday) to areas commonly and/or most recently affected by AD, and any areas where AD is developing. Subjects receiving maintenance treatment who experience AD worsening (signs and symptoms) for 1 week despite dosing twice weekly will phone the clinical study site, resume daily dosing, and complete the subject diary accordingly. The phone call (but not a clinic visit) is required to resume daily dosing. At clinic visits, subjects receiving maintenance treatment may continue twice weekly maintenance dosing if vIGA-AD is either '0-clear' or '1-almost clear'. Subjects will resume daily dosing if vIGA-AD ≥2, and can resume daily dosing if signs/symptoms of AD are not adequately controlled with maintenance therapy despite vIGA-AD of '1-almost clear'. Under Amendment 1, maintenance dosing determination will not be conducted at Week 24 for Cohort-Week 24.

[p] The entire kit (every tube) should be weighed and recorded at every visit. See IP Handling Manual for details.
[q] Compliance calculation is described in the IP Handling Manual
[r] 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.
[s] This data will be obtained from Week 4 of ARQ-151-311/312/315 Study and used as the Day 1 data for this long-term safety study (ARQ-151-313).
[t] Under Amendment 1, this is the final visit for Cohort-Week 24 from ARQ-151-311 and ARQ-151-312.

3.4 Treatment

Initial treatment with the IP in the ARQ-151-313 will occur on Day 1 (Day 29 from the preceding study). Roflumilast cream 0.15% or roflumilast cream 0.05% is administered once daily (unless entering BIW maintenance therapy after Week 4) as a topical product to cover the skin surface at an application rate of approximately 2 mg/cm^2.

- Roflumilast cream 0.15%, QD (≥ 6 years old on Day 1 of ARQ-151-313)
- Roflumilast cream 0.05%, QD (2 -5 years old on Day 1 of ARQ-151-313)

Subjects who turn 6 years old during the study will switch to roflumilast cream 0.15% at their first scheduled post-birthday visit.

3.5 Randomization, and Unblinding Procedures

This is an open-label study, therefore the subjects, the Investigator, clinical personnel, and the sponsor will be aware of which treatment an individual subject receives; however, the treatment received during the prior study will remain blinded while the prior study is ongoing.

3.6 Changes to the Analysis from the Protocol

Rationale for change	Description of the change
A categorical summary of amount of investigational product used by each subject based on tube weights was excluded from the SAP as it was determined that it was not necessary.	Original text: The amount of investigational product used by each subject based on tube weight will be summarized by treatment using descriptive statistics, and categorically. Changed to: The amount of investigational product used by each subject based on tube weight will be summarized by treatment using descriptive statistics.
Analyses to be performed by cohorts (i.e., 24 vs. 52 weeks) were excluded from the SAP as it was determined that it was not necessary.	Original text: The analysis will be performed by cohorts and overall over time. Changed to:

Rationale for change	Description of the change
	For analysis purposes, the TLFs will be summarized overall and by previous treatment group in the parent studies. Subgroup analyses by Cohort will be performed for specific endpoints (refer to Section 6.6)
Update compliant criteria for subject who are on BIW dosing schedule.	Original text: A subject will be considered compliant with the dosing regimen if the subject meets both of the following requirements: • applies at least 80% of the expected applications during the study drug application period • does not miss more than 3 consecutive doses Changed to: A subject will be considered compliant with the dosing regimen if the subject meets both of the following requirements: • applies at least 80% of the expected applications during the study drug application period • do not miss more than 3 consecutive doses while on the QD dosing schedule or at least 2 consecutive dosing on the BIW dosing schedule.

4 POPULATIONS FOR ANALYSIS

4.1 Full Analysis Population

The full analysis population is defined as all subjects who are enrolled in this study.

4.2 Safety Population

The safety population is defined as all subjects who are enrolled and received at least one dose of roflumilast in the parent study or at least one confirmed dose of roflumilast in this study.

4.3 Pharmacokinetic Population (PK)

The PK population includes all subjects who had a PK draw (concentration data available) post-baseline. This population will be used for the analysis of PK concentrations.

Of note, PK is only to be drawn on subjects aged 12 years or older who first enrolled under protocol amendment 1.

5 GENERAL CONSIDERATIONS

Formats and layouts of tables, listings, and figures (TLFs) will be provided in a separate document (output general layout is described in **Appendix 1**).

For analysis purposes, the TLFs will be summarized overall and by previous treatment group in the parent studies as indicated below. Rollovers from the ARQ-151-311/312 studies will be summarized separately from rollovers from the ARQ-151-315 study.

- Rollovers from ARQ-151-311/312 studies
    - Roflumilast cream 0.15%
    - Vehicle / Roflumilast cream 0.15%
    - Pooled Roflumilast cream 0.15%
- Rollovers from ARQ-151-315 study
    - Roflumilast cream 0.05% or 0.15%
    - Vehicle / Roflumilast cream 0.05% or 0.15%
    - Pooled Roflumilast cream 0.05% or 0.15%

5.1 Sample Size

Up to approximately 1500 subjects that have completed the 4-week treatment period in studies ARQ-151-311, ARQ-151-312, or ARQ-151-315 and are willing and qualified to enroll in this long-term extension study will receive roflumilast cream 0.15% (ARQ-151-311/312 roll-overs) or roflumilast cream 0.05% (ARQ-151-315 roll-overs) QD.

The sample size will provide a sufficient number of subjects to evaluate the long-term safety of roflumilast cream 0.15% or roflumilast cream 0.05% over 24 or 52 weeks of treatment (Cohort-Week 24 and Cohort-Week 52, respectively).

5.2 Baseline

There are two baseline definitions for the analyses:

- Primary baseline is defined as the last observation before or on the first dose of roflumilast cream 0.15% or 0.05% in ARQ-151 parent studies for subjects administered roflumilast previously and before or on the first dose of roflumilast cream 0.15% or 0.05% in this study* for subjects administered vehicle previously.

- OLE baseline is defined as the last observation before or on the first dose of roflumilast cream 0.15% or 0.05%* in this study.

If the last non-missing assessment was performed on the same date as the first dose of roflumilast cream* and time is not available, the assessment will be considered as baseline, except for adverse events (AEs) and medications starting on the first dose date which will be considered post-baseline. For subject local tolerability assessments, if the last non-missing assessment was performed on the same date as the first dose of roflumilast cream*, the assessment will be considered as baseline, regardless of the time of the assessment.

For the subjects administered vehicle in the parent study, the definition for Primary and OLE baselines are the same.

* Enrollment date (usually the Week 4 visit in the parent study) for enrolled subjects who discontinued early from this study before the first application of roflumilast cream 0.15% or 0.05%.

The following table (Table 2) described the baseline used in the analysis:

Table 2 Summary of Primary/OLE Baseline in the Analyses

Analyses	**Baseline***
Demographics and Baseline Characteristics	Primary baseline and OLE baseline, as applicable
Efficacy Analysis	Primary baseline and OLE baseline
Safety analysis for AEs and concomitant medications	OLE baseline
Safety analysis for clinical laboratory, vital signs, local tolerability assessments. PHQ-8, Modified PHQ-A, CDI-2, C-SSRS, and physical examination	Primary baseline

* Refer to Section 8 for more details.

5.3 Study Day

Study Day will be calculated from the first dose of roflumilast cream 0.15% or 0.05%* in this study, and will be used to show start/end day of assessments or events.

Study Day = (Date of assessment/event – Date of first dose of roflumilast cream in this study) + 1 if date of assessment/event is on or after the date of first dose of roflumilast cream in this study;

= (Date of assessment/event – Date of first dose of roflumilast cream in this study) if date of assessment/event is before the date of first dose of study roflumilast cream in this study.

* Enrollment date (usually the Week 4 visit in the parent study) for enrolled subjects who discontinued early from this study before the first application of roflumilast cream 0.15% or 0.05%.

In the situation where the assessment/event date is partially or completely missing, Study Day will be missing.

5.4 Windowing Conventions

The scheduled visit will be used to report measurements for summaries by visit regardless of whether this visit was performed within or outside of the analysis visit windows. Unscheduled visit, early termination (ET) visit, and/or retest measurements will only be used for the summaries by visit if a scheduled measurement is not available and the unscheduled/retest/ET measurement

falls within the analysis visit windows described in Table 3, Table 4, Table 5, and Table 6. Scheduled, unscheduled, ET, and retest measurements will be listed.

If there is more than one assessment for a given analysis visit when a scheduled measurement is not available, the assessment closest to the target day will be considered. If two assessments are equidistant to the target day, the latest assessment will be used.

Table 3 OLE Analysis Visit Windows – vIGA-AD, EASI, BSA Affected by AD, DLQI, CDLQI, IDQOL, DFI, SCORAD, POEM, Vital Signs, Local Tolerability Assessments, PHQ-8, PHQ-A, CDI-2, and C-SSRS

Analysis Visit	Target Day	Window (Days)
OLE Week 4	29	15 to 57
OLE Week 12	85	58 to 127
OLE Week 24	169	128 to 211
OLE Week 36	253	212 to 309
OLE Week 52	365	≥ 310

Table 4 OLE Analysis Visit Windows – WI-NRS

Analysis Visit	Target Day	Window (Days)
OLE Week 1	8	5 to 11
OLE Week 2	15	12 to 18
OLE Week 3	22	19 to 25
OLE Week 4	29	26 to 32
OLE Week x	y	z1 to z2
OLE Week 52	365	≥ 362

where: x= 5 to 51; y= (x * 7) + 1; z1 = y – 3, and z2 = y + 3.

Table 5 OLE Analysis Visit Windows – Clinical Laboratory

Analysis Visit	Target Day	Window (Days)
OLE Week 24	169	15 to 267
OLE Week 52	365	≥ 268

Table 6 OLE Analysis Visit Windows – Physical Examination

Analysis Visit	Target Day	Window (Days)
OLE Week 4	29	15 to 127
OLE Week 24	169	128 to 267
OLE Week 52	365	≥ 268

5.5 Derived Variables

All questionnaire scores will be derived by Biostatistics in the ADaM datasets using the formulas defined below, even if calculated scores are present in the EDC database. All pre-calculated scores will be ignored for analysis.

- With the following exception, vIGA-AD Score of Clear (0) or Almost Clear (1) = vIGA-AD of 'Clear' (0) or 'Almost Clear' (1). The exception is that for subjects who discontinued early from study due to an AE or lack of efficacy, the subject will be considered as not having a vIGA-AD score of Clear (0) or Almost Clear (1) for all pre-specified analysis visits (refer to Section 5.4) for which the subject's day of last dose falls within or is before the start of the analysis visit window.

- With the following exception, vIGA-AD Success = vIGA-AD of 'Clear' (0) or 'Almost Clear' (1) plus at least a 2-grade improvement from baseline. The exception is that for subjects who discontinued early from study due to an AE or lack of efficacy, the subject will be considered as not having a vIGA-AD success for all pre-specified analysis visits (refer to Section 5.4) for which the subject's day of last dose falls within or is before the start of the analysis visit window. vIGA-AD Success will be derived for both the Primary and OLE baselines (refer to Section 5.2)

- Date of last contact = the latest of last study visit date, last dose date, end-of-study date and end-of-treatment date.

- Post-OLE Week 4 study duration (days): The duration can be represented as End Date – Start Date +1. The start date of the Post-OLE Week 4 interval will be the date of the OLE Week 4 visit or OLE Day 29 if the subject missed the OLE Week 4 visit. The end date will be the date of last contact. This endpoint will be calculated only for subjects who enter the post-OLE Week 4 study period.

- Duration (days) of the first BIW dosing will be calculated from the first BIW dosing to the start of the succeeding QD dosing interval. Subject's last contact date will be used to censor

this endpoint (time-to-event analysis) or as the end date of this duration (analyzed as a continuous endpoint) for subjects who do not revert to a QD dosing regimen after the first BIW dosing. This endpoint will be calculated only for subjects who enter the post-OLE Week 4 study period with at least one BIW dosing interval.

- Cumulative time (days) on BIW dosing: Subjects may have more than 1 BIW dosing interval. The duration of succeeding BIW dosing intervals will be derived in a manner similar to the first and durations will be summed across all intervals for a subject. This endpoint will be calculated only for subjects who enter the post-OLE Week 4 study period with at least one BIW dosing interval.

- Percent of time on BIW dosing: The cumulative time (days) on BIW dosing will be divided by the Post-OLE Week 4 study duration and multiplied by 100 to measure the percentage of study days on BIW dosing. This endpoint will be calculated only for subjects who enter the post-OLE Week 4 study period with at least one BIW dosing interval.

- Duration (days) of first disease control will be calculated from the first disease control to the time subjects revert to a QD dosing or have a vIGA-AD score ≥ 2, whichever comes first. Disease control is defined as achievement of a vIGA-AD score of 0, a switch to BIW dosing, and maintenance of a vIGA-AD score of 0 or 1. Disease control starts with the later date of a vIGA-AD score of Clear (0) on or after OLE Week 4 and the switch to BIW dosing. Subject's last contact date will be used to censor this endpoint (time-to-event analysis) or as the end date of this duration (analyzed as a continuous endpoint) for subjects who never lose disease control after the start of the first disease control interval. This endpoint will be calculated only for subjects who enter the post-OLE Week 4 study period and had at least one disease control interval.

- Cumulative time (days) with disease control. Subjects can come in and out of disease control. The duration for each interval will be derived as described above for the first interval. Cumulative time with disease control will sum the durations for all intervals where disease was under control as defined in the previous bullet. This endpoint will be calculated only for subjects who enter the post-OLE Week 4 study period and had at least one disease control interval.

- Percent of time with disease control. The cumulative time with disease control will be divided by the Post-OLE Week 4 study duration and multiplied by 100 to express cumulative time as a percentage of the duration of time that each subject spent in the study after OLE Week 4. This endpoint will be calculated only for subjects who enter the post-OLE Week 4 study period and had at least one disease control interval.

- Duration (days) of the first vIGA-AD of Clear (0) will be calculated from the first vIGA-AD score of 0 during the study to the next vIGA-AD score > 0. Subject's last contact date will be used to censor this endpoint (time-to-event analysis) or as the end date of this duration (analyzed as a continuous endpoint) for subjects who never revert to a vIGA-AD score > 0. This endpoint will be calculated only for subjects who had at least one vIGA-AD score of Clear (0) interval.

- Cumulative time (days) with vIGA-AD of Clear (0). Subjects can have more than 1 vIGA-AD of Clear (0) interval. The duration for each interval will be derived as described above for the first interval. Cumulative time with vIGA-AD of 0 will sum the durations for all intervals. This endpoint will be calculated only for subjects who had at least one vIGA-AD score of Clear (0) interval.

- Percent of time with vIGA-AD of Clear (0). The cumulative time will be divided by the study duration and multiplied by 100 to express cumulative time as a percentage of the duration of time that each subject spent in the study. This endpoint will be calculated only for subjects who had at least one vIGA-AD score of Clear (0) interval.

- Duration (days) of the first vIGA-AD of Clear (0) or Almost Clear (1) will be calculated from the first vIGA-AD score of 0 or 1 during the study to the next vIGA-AD score > 1. Subject's last contact date will be used to censor this endpoint (time-to-event analysis) or as the end date of this duration (analyzed as a continuous endpoint) for subjects who never revert to a vIGA-AD score > 1. This endpoint will be calculated only for subjects who had at least one VIGA-AD with a score of Clear (0) or Almost Clear (1) interval.

- Cumulative time (days) with vIGA-AD of Clear (0) or Almost Clear (1). Duration with vIGA-AD score of 0 or 1 measures the duration for the first interval with a vIGA-AD score of 0 or 1. There can be more than 1 interval. Cumulative time with vIGA-AD of 0 or 1 will sum the durations for all intervals. The duration for each interval will be derived as described above for the first interval. This endpoint will be calculated only for subjects who had at least one VIGA-AD with a score of Clear (0) or Almost Clear (1) interval.

- Percent of time with vIGA-AD of Clear (0) or Almost Clear (1). The cumulative time will be divided by the study duration and multiplied by 100 to express cumulative time as a percentage of the duration of time that each subject spent in the study. This endpoint will be calculated only for subjects who had at least one VIGA-AD with a score of Clear (0) or Almost Clear (1) interval.

- WI-NRS 4-point Reduction = achievement of a ≥ 4-point reduction in weekly? WI-NRS pruritus score from baseline. WI-NRS 4-point reduction will be derived for both the Primary and OLE baselines (refer to Section 5.2)

- EASI Total Score:

  Four anatomic sites—head, upper extremities, trunk, and lower extremities—are assessed for erythema, induration/infiltration (papules), excoriation, and lichenification as seen on the day of the examination. The severity of each sign is assessed using a 4-point scale (half steps **are** allowed e.g., 0.5, 1.5 and 2.5):

  0 = none

  1 = mild

  2 = moderate

  3 = severe

  The area affected by atopic dermatitis within a given anatomic site is estimated as a percentage of the total area of that anatomic site and assigned a numerical value according to the degree of atopic dermatitis involvement as follows:

  0 = no involvement

  1 = 1-9%

  2 = 10-29%

  3 = 30-49%

  4 = 50-69%

  5 = 70-89%

  6 = 90-100%

  The EASI score is obtained by using the formula below for subjects ≥8 years old:

  $$EASI = 0.1\ (E_h + I_h + Ex_h + L_h)\ A_h + 0.2\ (E_u + I_u + Ex_u + L_u)\ A_u + 0.3\ (E_t + I_t + Ex_t + L_t)\ A_t + 0.4\ (E_l + I_l + Ex_l + L_l)\ A_l$$

  The EASI score is obtained by using the formula below for subjects <8 years old:

  $$EASI = 0.2\ (E_h + I_h + Ex_h + L_h)\ A_h + 0.2\ (E_u + I_u + Ex_u + L_u)\ A_u + 0.3\ (E_t + I_t + Ex_t + L_t)\ A_t + 0.3\ (E_l + I_l + Ex_l + L_l)\ A_l$$

Where E, I, E*x*, L, and A denote erythema, induration, excoriation, lichenification and area, respectively, and h, u, t, and l denote head, upper extremities, trunk, and lower extremities, respectively.

Note: Palms and soles may be treated with investigational product in this study, but will not be counted towards IGA, EASI, or BSA assessments. EASI is evaluated for the entire body except the scalp, palms, and soles.

Note: If a subject turns 8 years old during the study, the formula used at Screening of the preceding study will continue to be used through the duration of the subject's participation in this study.

- EASI-50, EASI-75, EASI-90, EASI-100 - Achievement of at least a 50%, 75%, 90%, or 100% reduction from baseline in the EASI total score.
- DLQI Total 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, if No, then follow the same score as A lot, A little, Not at all), ranging from 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. The DLQI will be performed for the subject with age 17+ years based on age at start of preceding studies ARQ-151-311/312/315.
- CDLQI Total 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 holiday time, the standard responses apply), ranging from 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. The CDLQI will be performed for the subjects with age >4 and <16 years based on age at start of preceding studies ARQ-151-311/312/315.
- IDQOL Total Score = sum of the 10 questions (individual question 1 scored as All the time= 3, A lot= 2, A little= 1, None= 0; individual question 2 scored as: Always crying, extremely difficult= 3, Very fretful= 2, Slightly fretful= 1, Happy= 0; individual question 3 scored as More than 2 hours= 3, 1 – 2 hours= 2, 15 minutes – 1 hour= 1, 0 – 15 minutes= 0; individual question 4 scored as 5 hours or more= 3, 3 – 4 hours= 2, 1 – 2 hours= 1, Less than 1 hour= 0; individual questions 5 to 10 scores as Very much= 3, A lot= 2, A little= 1, Not at all/None= 0), ranging from 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. Of note, the severity of eczema question is scored and reported separately than the IDQOL total score. The IDQOL

will be completed by parents/caregivers for subjects with age < 4 years based on age at start of the preceding study ARQ-151-315.

- DFI Total Score = sum of the 10 questions (individual questions scored as Very much=3, A lot=2, A little=1; Not at all=0) ranging from 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. The DFI will be completed by parents/caregivers for all subjects ≥ 2 years of age and ≤ 17 years of age. Please note that the DFI is not to be conducted for subjects who were ≤ 17 years of age in the ARQ-151-311/312 studies and turned 18 years of age prior to enrollment into ARQ-151-313.

- PHQ-8 Total Score = 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), ranging from 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 (i.e, 7). The PHQ-8 Assessment will be performed in adult subjects based on age at start of preceding studies ARQ-151-311/312.

- Modified PHQ-A Total Score = 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), ranging from 0 to 24. If more than 1 items 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 (i.e., 7). The Modified PHQ-A Assessment will be performed in adolescent subjects (12-17 years old, inclusive) based on age at start of preceding studies ARQ-151-311/312.

- CDI-2 Total Score = sum of the 17 questions (individual questions Q2, Q7, Q13, Q14, and Q16 scored as Much or most of the time=0, Often=1, Some of the time=2, Not at all=3; individual questions Q1, Q3 to Q6, Q8 to Q12, Q15, and Q17 scored as Much or most of the time=3, Often=2, Some of the time=1, Not at all=0), ranging from 0 to 51. If more than 2 items are missing the total score should not be calculated. If 1 item is missing the total score is calculated as (sum of answered items*17)/number of answered items (i.e., 16). If 2 items are missing the total score is calculated as (sum of answered items*17)/number of answered items (i.e., 15). The CDI-2 Assessment will be performed for subjects 6 to 11 years old, inclusive based on age at start of preceding studies ARQ-151-311/312/315.

- CDI-2 Emotional Problem Subscale Score = sum of 9 questions (Q1, Q3-6, Q8, Q10-12). If more than 1 item is missing the subscale score should not be calculated. If 1 item is missing the subscale score is calculated as (sum of answered items*9)/number of answered items (i.e., 8).

- CDI-2 Functional Problem Subscale Score = sum of 8 questions (Q2, Q7, Q9, Q13-17). If more than 1 item is missing the subscale score should not be calculated. If 1 item is missing the subscale score is calculated as (sum of answered items*8)/number of answered items (i.e., 7).
- POEM Total Score = sum of the 7 questions (individual questions scored as No days = 0, 1 to 2 days = 1, 3 to 4 days = 2, 5 to 6 days = 3, Every day = 4), ranging from 0 to 28. If 1 question is left unanswered, this question is scored 0 and the total score is calculated as usual out of a maximum of 28. If 2 or more questions are left unanswered, the questionnaire total score should not be calculated.
- SCORAD Total Score = [Overall BSA affected by AD / 5] + [Intensity score * 7/ 2] + Subjective symptoms score (i.e., pruritus + sleep loss), ranging from 0 to 103. SCORAD total score will be set to missing if information for any of the three measures is missing.
- Change from baseline = Assessment value at post-baseline visit X – baseline value.
- Percent change from baseline = (Assessment value at post-baseline visit X – baseline value) × 100% / (baseline value)

  Percent change from baseline will be missing in situation where baseline value equals to 0.
- Exposure-adjusted incidence rate (EAIR) = (number of subjects with an event / total within-arm exposure [days]) * 365.25 days/year * 100.
- BMI (kg/m^2) = (weight in kg)/ [(height in cm/100)^2]. For child and adolescent (2 to 17 years), after BMI is calculated, it is expressed as a percentile using data files and instructions provided by the CDC (https://www.cdc.gov/growthcharts/percentile_data_files.htm).

5.6 Descriptive Statistics

All continuous variables will be summarized by presenting the number of subjects ('n'), mean, standard deviation (SD), median, Q1, Q3, minimum, and maximum. For PK endpoints, the coefficient of variation (CV), geometric mean, and geometric CV will also be provided.

Categorical variables will be presented as frequencies and percentages. The AE tables will also present the within-arm summary measure of EAIR; expressed as number of subjects with events per 100 subject exposure years.

Summary tables will be presented by visit, when applicable, and treatment group (refer to Section 5).

5.7 Statistical Tests

No hypothesis tests are planned for this study; therefore p-values will not be provided. For estimation purposes, two-sided 95% confidence interval (CI) will be supplied for secondary efficacy endpoints and EAIRs.

5.8 Handling of Retests, Unscheduled Visits, and Early Termination Data

Retests measurements, unscheduled measurements, and ET visit assessments will be included in analysis and be summarized via analysis visit windowing according to the windowing conventions in section 5.4.

5.9 Software Version

All analyses will be performed using SAS® software Version 9.4 or higher.

6 STATISTICAL CONSIDERATIONS

6.1 Adjustments for Covariates

There will be no adjustment for covariates when performing summaries for this study.

6.2 Handling of Dropouts or Missing data

See **Appendix 2** for handling of completely or partially missing dates for prior and concomitant medications and AEs.

Unless otherwise specified, missing safety data will not be imputed.

6.2.1 Multiple Imputation

For the vIGA-AD secondary efficacy endpoints (refer to Section 12.2), vIGA-AD non-imputed values collected during the parent studies and this study (including the last available data of subjects who prematurely withdraws from this study) will be appended together after having been assigned to pre-specified analysis visit using the analysis windows defined in their respective study (parent study SAPs or this SAP Section 5.4, as applicable). Doing so, each subject will have between 4 (subjects who enrolled in OLE and discontinued early before the first application of

roflumilast cream during the OLE study) to 7 (subjects in Cohort-Week 24 who completed the OLE) or 9 (subjects in Cohort-Week 52 who completed the OLE) vIGA-AD non-imputed values:

OLE Analysis Visit	Study/Analysis Visit
OLE Week -4	Parent study derived baseline value (refer to parent study Section 5.2)
OLE Week -3	Parent study / analysis visit Week 1
OLE Week -2	Parent study / analysis visit Week 2
OLE Day 1	Parent Study / analysis visit Week 4
OLE Week 4	OLE study / analysis visit Week 4
OLE Week 12	OLE study / analysis visit Week 12
OLE Week 24	OLE study / analysis visit Week 24
OLE Week 36*	OLE study / analysis visit Week 36
OLE Week 52*	OLE study / analysis visit Week 52

* Cohort-Week 52 subjects only.

Note: No missing data are expected for the OLE Week -4 to OLE Day 1 analysis visits since subjects must have completed a parent study to be eligible for enrollment in this OLE study.

Then, to comply with the definition of the parent studies' primary estimand (refer to parent studies Section 12.1.1), efficacy data assigned to a pre-specified analysis visit will be removed from the source data used for the multiple imputation process if subject's last dose day falls within the analysis visit window or is prior to the start of the analysis visit window used to assign data to a pre-specified analysis visit (refer to Section 12.2) for subjects who discontinued due to lack of efficacy or adverse event. This procedure will ensure that the data collected on or after intercurrent events are not used in the imputation process.

vIGA-AD missing values will be imputed before deriving the vIGA-AD secondary efficacy endpoints (refer to Section 12.2), using a Predictive Mean Matching (PMM) sequential-regression multiple imputation model for the Full Analysis Set population as per the following 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 exist values for OLE Week 4 and OLE Week 24 visits, but missing values for the OLE 12 visit, 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 6457149. The model will include the parent study randomized vIGA-AD stratification variable and parent study randomized treatment group (refer to Section 5.5), both as categorical variables, as well as all vIGA-AD scores collected at scheduled visits during both the parent and OLE studies, all as continuous variables. To avoid values that could not be observed in practice, imputed values will be rounded to the nearest integer (Round=1 option in PROC MI) in the range of 0 to 4.

To determine the number of multiply imputed datasets to be created at this step, the proportion of datapoints with non-monotone pattern across all visits and subjects will first be derived as follows:

$$\frac{number\ of\ non\ monotone\ visits\ accross\ all\ visits\ and\ subjects}{total\ number\ of\ expected\ visits\ across\ all\ subjects} * 100$$

Of note, the total number of expected visits across all subjects is not the same for the Cohort-Week 24 subjects and Cohort-Week 52 subjects i.e., 7 visits during the OLE study for Cohort-Week 24 subjects and 9 visits during the OLE study for the Cohort-Week 52 subjects.

Then, the following table will be used:

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 PMM regression method 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 regression model will be fit that includes the outcome at that visit as the dependent variable and as independent variables, the parent study randomized vIGA-AD stratification variable and parent study randomized treatment group, both as categorical variables, as well as the vIGA-AD score outcomes at previous analysis visits, all as continuous variables, using a seed of 482371. 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 multiply imputed dataset, Primary or OLE Baseline values, as applicable, will be derived as defined Section 5.2 and analysis visits from parent studies (OLE Week -4 to OLE Day 1) not selected as Primary or OLE Baseline, as applicable, will be dropped from each multiple imputed dataset.

   For Cohort-Week 24 subjects, imputed OLE Week 36 and Week 52 analysis visit values will be set to missing given that these subjects will complete study at Week 24 or discontinue early from study before Week 24, whichever occurs first.

   Finally, for each completed dataset, the vIGA-AD secondary efficacy endpoints will be derived by visit, as described in Section 12.2. Common proportion of success/achievement (and associated standard error [SE]) will then be calculated overall and by treatment group for each OLE analysis visit and combined using PROC MIANALYZE based on Rubin's rule to obtain a combined common proportion of success/achievement for each OLE analysis visit overall and treatment group. The combined common two-sided 95% CI will be calculated as per Lott and Reiter (2020)[1] multiple imputation Wilson interval method

6.2.2 Non-Responder Imputation

For the vIGA-AD secondary efficacy endpoints (refer to Sections 12.2), missing endpoint values will also be imputed as a non-responder i.e., not achieving the endpoint criteria.

6.3 Interim Analyses

An IA of data collected in this study will be performed after the database lock and study unblinding of parent studies ARQ-151-311 and ARQ-151-312. Only data from subjects that rolled over from one of these two parent studies into study ARQ-151-313 will be included in this IA. That is, in order to preserve the blind of parent study ARQ-151-315, which will still be ongoing at the time of ARQ-151-311 and ARQ-151-312 database lock and study unblinding, no data from subjects enrolled in parent study ARQ-151-315 that have rolled over into study ARQ-151-313 will be included in this IA.

Another IA may be conducted at the time of rollovers from these two parent studies complete all scheduled visits in ARQ-151-313. If conducted, similarly to the first IA, only data from subjects

that rolled over from one of these two parent studies into study ARQ-151-313 will be included in the IA in order to preserve the blind of parent study ARQ-151-315.

All analyses planned in this SAP will be performed at the time of the IA.

6.4 Multicenter Studies

Data from all sites will be pooled in the summaries.

6.5 Multiple Comparisons/Multiplicity

No adjustments for multiple comparisons will be made for this study.

6.6 Examination of Subgroups

Subgroup analyses will be performed for the vIGA-AD secondary efficacy endpoints (refer to Section 12.2.3) as well as for the WI-NRS , EASI-50, and EASI-75 endpoints in the other efficacy endpoints (refer to Section 12.3.3) for the following subgroups relative to Primary Baseline only (refer to Section 5.2):

- Age group at informed consent in parent study (2 – 5 years, 6 – 11 years, 12 - 17 years, and ≥ 18 years), including the WI-NRS endpoints. That is, although the main analysis for the WI-NRS endpoint in the other efficacy endpoints will be performed on subjects ≥ 12 years old, a subgroup analysis will be performed for each of the previously defined age groups, including the 2 – 5 years and 6 – 11 years subgroups.
- Sex (Male and Female)
- Race (White, Black or African American, and Asian vs. Other)
- Ethnicity (Hispanic or Latino and Non-Hispanic or Latino)
- Actual Baseline vIGA-AD score (2- mild and 3- moderate) in parent study
- Prior inadequate response, intolerance, or contraindication to Topical Corticosteroids (yes and no)
- Facial Involvement (yes and no)

Subgroup analyses for rollovers from ARQ-151-311/312 will be performed for the duration, cumulative time, and percent of time other efficacy endpoints (refer to Section 12.3.3) for the following subgroups:

- Cohort (Cohort-Week 24 and Cohort-Week 52)

Furthermore, to investigate long-term maintenance of the treatment effect, additional subgroup analyses will be conducted on subgroups below that are defined based on post-treatment characteristics. More specifically, subgroup analyses for rollovers from ARQ-151-311/312 will be performed for the BIW dosing and disease control other efficacy endpoints (refer to Section 12.3.3) for the following subgroups (by Cohort and overall):

a. Subjects whose first interval started at OLE Week 4

b. Subjects with vIGA-AD success at OLE baseline

c. Subjects with vIGA-AD score of Clear (0) at OLE baseline

The subgroups criteria a and b as well as a and c will be combined for further subgroup analyses for the BIW dosing and disease control endpoints.

Additionally, the overall summary of treatment-emergent adverse events (TEAE) and summary of TEAEs by system organ class (SOC) and preferred term (PT) (refer to Section 13.1) will be presented by the age groups specified above.

Finally, in order to better understand the efficacy and safety profile of ARQ-151-315 rollover subjects who start this study 5 years old and who turn 6 years old during this study and so, will be switched from roflumilast cream 0.05% to roflumilast cream 0.15% at his/her first scheduled post-birthday visit, analyses will also be performed for this subgroup of subjects. The list of efficacy and/or safety endpoints for which analysis will be performed for this subgroup of subjects will be determined, finalized and approved before the final database lock of parent study ARQ-151-315.

7 STUDY SUBJECTS

7.1 Disposition of Subjects

Disposition summary will include tabulation of the number of the subjects who enrolled into the OLE (overall and by Cohort) as well as the number and percentage of subjects who received treatment during the OLE (overall and by Cohort), completed the OLE, prematurely discontinued during the OLE (including reasons for discontinuation and ET due to COVID-19 disruption) and in each analysis population by previous treatment group in parent studies as defined in Section 5. The percentages will be calculated using the number of subjects included in the full analysis population.

Number of days in the study will be summarized with descriptive statistics by previous treatment group in parent studies as defined in Section 5. For each subject, the number of days in the study will be calculated as following:

Date of completion/discontinuation – Date of enrollment into OLE +1

A listing of subject's disposition and inclusion/exclusion from analysis population will be provided.

7.2 Protocol Deviations

A data review will be conducted before the IA database lock as well as before the final database lock by the Medical Monitor and the Sponsor to classify protocol deviations as important or non-important. The number and percentage of subjects with at least one protocol deviation, at least one protocol deviation associated with COVID-19, and at least one important protocol deviation will be summarized by previous treatment group in parent studies based on the full analysis set. Protocol deviations associated with COVID-19 and important protocol deviations will be further broken down by deviation category.

A listing of all protocol deviations will also be provided. The protocol deviations associated with COVID-19 and important PDs will be flagged.

8 DEMOGRAPHIC AND BASELINE CHARACTERISTICS

Demographics and baseline characteristics will be summarized with descriptive statistics by previous treatment group in parent studies as defined in Section 5 using the safety population.

The list of demographics to be summarized includes:

- Age (years) at the start of the parent study
- Age group at the start of the parent study: 2 – 5 years, 6 – 11 years, 12 - 17 years, ≥ 18 years
- Age (years) at the start of the OLE study
- Age group at the start of the OLE study: 2 – 5 years, 6 – 11 years, 12 - 17 years, ≥ 18 years
- Sex at birth
- Childbearing potential at the start of the OLE study
- Race*

- Ethnicity
- Fitzpatrick Skin Type
- Prior failure of Topical Corticosteroids, Topical Calcineurin Inhibitors, Eucrisa;
- Atopic Dermatitis involvement on the face, on the eyelids

The following efficacy assessments will be summarized at both Primary Baseline and OLE Baseline:

- Baseline vIGA-AD (continuous and categorical)
- Baseline WI-NRS
- Baseline BSA affected by AD (%)
- Baseline EASI total score
- Baseline SCORAD

The following assessments will be summarized at the Primary Baseline only:

- Baseline Height (cm)
- Baseline Weight (kg)
- Baseline Body Mass Index (BMI) (kg/m^2) for children (2 – 5 years, 6 – 11 years), adolescents (12 – 17 years), and adult (≥ 18 years)
- Baseline DLQI/CDLQI/IDQOL
- Baseline DFI
- Baseline POEM
- Baseline PHQ-8
- Baseline PHQ-A
- Baseline CDI-2

*Subjects who reported more than one race will be summarized as 'Multiple' races in the table. All races selected will be displayed in the listing.

A listing of all demographics and baseline characteristics will be provided.

9 SURGICAL AND MEDICAL HISTORY

Surgical and medical history will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA), Version 25.0 or higher. Medical and surgical history collected in the parent studies for subjects that rolled over into this study will be included and presented.

Surgical and medical history will be summarized by system organ class (SOC), preferred term (PT), and treatment group as defined in Section 5 using the safety population. A subject who experienced the same surgical and medical history event multiple times will be counted only once for the corresponding PT. Similarly, if a subject experienced multiple surgical and medical history events within the same SOC, the subject will be counted only once for that SOC. Surgical and medical history events will be sorted alphabetically by SOC and within each SOC the PT will be presented by descending frequency in the safety population.

10 PRIOR AND CONCOMITANT MEDICATIONS

Medications will be coded according to the World Health Organization Drug Dictionary (WHO-DD), March 2022 B3.

Prior medications are defined as any medication started and discontinued prior to the first study treatment dosing in the OLE. Concomitant medications are defined as any medication taken on or after the first dose of study treatment in OLE, including those who started prior to the first dose of study treatment date and continued past that date in OLE study. See **Appendix 2** for handling of completely or partially missing dates for prior and concomitant medications.

Incidence of concomitant medications will be tabulated by Anatomic Therapeutic Class (ATC) level 3, preferred drug name, and treatment group as defined in Section 5 using the safety population. A subject with the same medication taken multiple times will be counted only once for the corresponding preferred drug name. Similarly, if a subject has taken more than one medication within the same ATC level, then the subject will be counted only once for that ATC. Concomitant medications will be sorted alphabetically by ATC level and within each ATC level, the preferred drug name will be presented by descending order.

Prior medications are not planned to be listed and summarized in this study as prior medications were presented in parent studies.

A listing of concomitant medications will be provided.

11 STUDY TREATMENT EXPOSURE AND TREATMENT COMPLIANCE

A summary of exposure to study treatment will be presented using the safety population by treatment group as defined in Section 5. It will include descriptive statistics on the number of days on OLE study treatment, as well as the number of study treatment applications based on diary, for each treatment group. Since subjects in the roflumilast arm of their parent study will simply be continuing the same treatment, their first treatment date will be the date of their week 4 visit in the parent study, even if study drug application was interrupted on that date. The number of days on study treatment in OLE study, will be calculated as follows:

(Last treatment date in OLE - First treatment date in OLE) + 1

For each subject, compliance (%) with study treatment will be calculated as follows:

$$\frac{\text{Number of study treatment applications}}{\text{Number of expected study treatment applications}} \times 100$$

Applied dose count and missed dose applications count were collected in eCRF study drug interruption module. Number of study treatment applications will be calculated as the sum of applied dose count.

Number of expected study treatment applications will be calculated as the sum of number of investigational product applications and number of doses missed. Of note, number of expected study treatment applications might change during the course of the study (refer to Section 3.1 for more details).

The number and percent of subjects who switch to a maintenance dosing schedule (BIW) at least once after week 4 will be provided by treatment group. The cumulative time and percent of time (Section 5.5) on the maintenance dosing schedule will be summarized only for those subjects who switched to BIW dosing at least once. The percent of time on the maintenance dosing schedule will also be summarized categorically as]0-25%],]25-50%],]50-75%], and >75%. The number of times each subject returned to QD dosing after switching to BIW dosing will be summarized categorically.

Subjects who roll over from the ARQ-151-315 study and turn 6 years old during the study will switch to Roflumilast cream 0.15% at their first scheduled post-birthday visit. The percentage of study days spent on each dose in this study will be summarized.

Descriptive statistics for the number of missed applications as well as the number and percentage of subjects who missed more than 3 consecutive doses while on the QD dosing schedule or more than 2 consecutive doses while on the BIW dosing schedule will be presented by treatment group as defined in Section 5. Descriptive statistics for the compliance as well as number and percentage of subjects with < 80%, [80% - 100%], and >100% compliance will be presented similarly.

Furthermore, the number and percentage of subjects deemed as compliant with the dosing regimen, defined as subjects meeting both of the following requirements will also be provided by treatment group as defined in Section 5 using the safety population:

- Apply at least 80% of the expected applications during the study drug application period
- Do not miss more than 3 consecutive doses while on the QD dosing schedule or more than 2 consecutive dosing while on the BIW dosing schedule.

Total weight of study treatment applied will be summarized by treatment group as defined in Section 5 using descriptive statistics in the safety population. Weight of study treatment used will be documented in source documents and in eCRF. Total weight of study treatment applied is determined by subtracting returned tube weight from the dispensed tube weight for each tube that was dispensed and summing the weights. Of note, if a tube lasts for more than 1 visit, the Dispensed Tube Weight is the last weight measured and not the full tube weight. If a tube is not returned at a specific visit, then weight will be considered as missing at that visit.

A listing of exposure to study treatment, compliance with study treatment, and total weight of study treatment applied will be provided. A listing of drug accountability including the kit number, tube number, dispensed and returned weight will also be provided.

12 EFFICACY ANALYSES

Unless otherwise indicated, missing efficacy data will not be imputed.

Efficacy endpoints will be summarized based on the full analysis set. For change and percent change efficacy endpoints relative to baseline, summaries will be performed relative to both Primary Baseline and OLE Baseline (refer to Section 5.2).

12.1 Primary Efficacy Endpoint

Not applicable.

12.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints are:

- vIGA-AD score of clear (0) or almost clear (1) at each assessment (refer to Section 5.5)
- vIGA-AD success (refer to Section 5.5) over time
- Change and percent change from baseline in WI-NRS over time
- Change and percent change from baseline in EASI total score (refer to Section 5.5) over time

These endpoints are based on the following scales/questionnaire:

- The vIGA-AD scale is a static evaluation of qualitative overall AD severity. This global assessment scale is an ordinal scale with five severity grades (reported only in integers of 0 to 4), where a higher score represents a greater severity. Each grade is defined by a distinct and clinically relevant morphologic description that minimizes inter-observer variability. vIGA-AD is evaluated for the entire body except the scalp, palms, and soles.
- The WI-NRS scale is a single-item assessment of the subject-reported worst itch severity during the previous 24-hour period. The scale is from 0 to 10 ('no itch' or 'worst itch imaginable'). The WI-NRS will be self-reported by subjects for subjects ≥ 6 years of age and reported by parent/guardian for subjects < 6 years of age.
- The EASI questionnaire assessed the severity of erythema, induration/infiltration (papules), excoriation, and lichenification, as seen on the day of the examination, of four anatomic sites separately (head, upper extremities, trunk, and lower extremities). The severity of each sign is assessed using a 4-point scale (half steps are allowed e.g., 0.5): 0 (none), 1 (mild), 2 (moderate), and 3 (severe). The area affected by AD within a given anatomic site is estimated as a percentage of the total area of that anatomic site and assigned a numerical value according to the degree of AD involvement as follows: 0 (no involvement), 1 (1-9%), 2 (10-29%), 3 (30-49%), 4 (50-69%), 5 (70-89%), and 6 (90-100%). The EASI total score is then computed as specified in Section 5.5.

12.2.1 Main Analysis

Missing vIGA-AD secondary efficacy endpoints will be imputed as described in Section 6.2.1. The combined common proportion of subjects achieving each vIGA-AD secondary endpoint will be provided by visit and treatment group (refer to Section 5) along with their combined common two-sided 95% CI. The vIGA-AD success summaries based on the OLE Baseline (refer to Section

5.2) will only include subjects who have an OLE baseline score other than clear (0) or almost clear (1).

The change and percent change from baseline in WI-NRS score will be summarized among subjects 12 years or older at the start of the parent study* using descriptive statistics by visit and treatment group (refer to Section 5) based on observed cases (i.e., no imputation). Two-sided 95% CI of the mean will also be provided. WI-NRS score will be summarized similarly.

* WI-NRS score, change from baseline and percent change from baseline for subjects < 12 years old at the start of the parent study will be summarized in subgroup analysis (refer to Sections 6.6 and 12.2.3).

The change and percent change from baseline in EASI total score will be summarized using descriptive statistics by visit and treatment group (refer to Section 5) based on observed cases (i.e., no imputation). Two-sided 95% CI of the mean will also be provided. EASI total score will be summarized similarly.

12.2.2 Sensitivity Analyses

As sensitivity analyses to the imputation method used for the main analysis of the vIGA-AD endpoints (refer to Section 12.2.1), the main analysis of these endpoints will be repeated, but using the following imputation method:

- Non-responder imputation (refer to Section 6.2.2)
- Observed cases (i.e., no imputation)

No sensitivity analyses will be performed for the secondary efficacy endpoints based on WI-NRS and EASI total score.

12.2.3 Subgroup Analyses

The main analysis of the secondary efficacy endpoints of vIGA-AD score of clear (0) or almost clear (1) and vIGA-AD success (refer to Section 12.2.1) will be repeated for each of the subgroups mentioned in Section 6.6. Subgroup analyses will only be conducted for results relative to the primary baseline only.

12.3 Other Efficacy Analyses

Other efficacy endpoints in this study are:

- Duration (days) of the first BIW dosing interval

- Cumulative time (days) on BIW dosing
- Percent of time on BIW dosing
- Duration (days) of the first disease control interval
- Cumulative time (days) with disease control
- Percent of time with disease control
- Duration (days) of the first interval with vIGA-AD score of Clear (0)
- Cumulative time with vIGA-AD score of Clear (0)
- Percent of time with vIGA-AD of Clear (0)
- Duration (days) of the first interval with vIGA-AD score of Clear (0) or Almost Clear (1)
- Cumulative time with vIGA-AD score of Clear (0) or Almost Clear (1)
- Percent of time with vIGA-AD of Clear (0) or Almost Clear (1)
- Change and percent change from baseline in vIGA-AD score over time
- In subjects ≥ 12 years old at the start of the parent study* with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS over time
- Achievement of at least a 50%, 75%, 90%, 100% reduction in the EASI total score (EASI-50, EASI-75, EASI-90, EASI-100) over time
- Change and percent change from baseline in BSA affected by AD over time
- Change and percent change from baseline in IDQOL, CDLQI, and DLQI over time
- Change and percent change from baseline in DFI over time
- Change and percent change from baseline in SCORAD over time
- Change and percent change from baseline in POEM over time

* WI-NRS score, change from baseline and percent change from baseline for subjects < 12 years old at the start of the parent study will be summarized in subgroup analysis (refer to Sections 6.6 and 12.2.3).

Endpoints that are durations, cumulative times, or percent of times are defined in Section 5.5 while endpoints that are not durations, cumulative times, or percent of times are based on the following scales/questionnaires/assessments/tools:

- The vIGA-AD scale (refer to Section 12.2)
- The WI-NRS scale (refer to Sections 5.5 and 12.2)
- The EASI questionnaire (refer to Section 12.2)
- Percentage of BSA affected by AD will be determined by the subject's hand method, where the subject's hand (including fingers) surface area is assumed to equal 1% of BSA (excluding the scalp, palms, soles).
- The DLQI and CDLQI questionnaires are designed to measure the health-related quality of life of patients suffering from a skin disease. The DLQI/CDLQI consists of 10 questions concerning patients' perception of the impact of skin diseases on different aspects of their health-related quality of life over the last week. The DLQI will be completed by subjects ≥17 years of age (based on age at start of preceding studies ARQ-151-311/312). The CDLQI will be completed for subjects ≥4 years of age and ≤16 years of age (based on age at start of preceding studies ARQ-151-311/312/315). Subjects who were assessed using CDLQI in their preceding study (ARQ-151-311/312/315) will remain on CDLQI in the current study (ARQ-151-313). DLQI/CDLQI total score will be computed as specified in Section 5.5.
- The IDQOL questionnaire is designed to assess the impact of AD on the quality of life of infants below the age of four years (≤4 years). Subjects who were assessed using IDQOL in their preceding study (ARQ-151-315) will remain on IDQOL in the current study (ARQ-151-313). It should be completed by the child's parent(s) or regular caregiver. IDQOL score will be computed as specified in Section 5.5.
- The DFI questionnaire measures how much having a child with AD affects the quality of life of other (adult) members of the family. It will be completed by parents/guardians/caregivers for subjects ≤17 years of age. DFI is not to be conducted for subjects who were ≤17 years of age in the ARQ-151-311/312 studies and turned 18 years of age prior to enrollment into ARQ-151-313. DFI total score and subscale scores will be computed as specified in Section 5.5.
- The SCORAD clinical tool assesses the severity (i.e., extent, intensity) of AD as objectively as possible. First, the overall %BSA affected by AD is evaluated (from 0% to

100%, where a subject's palm represents 1% of his/her total BSA). Secondly, the AD severity is evaluated based on 6 items (erythema, edema/papulation, oozing/crusts, excoriation, lichenification, and dryness) graded using a 4-point scale (half steps are not allowed): 0 (absence), 1 (mild), 2 (moderate), and 3 (severe). Lastly, 2 subjective items (loss of sleep and intensity of pruritus) are evaluated by having the subject indicates on a 10.0 cm visual analog scale (VAS) the point corresponding to the average value over the last 3 days (0 cm = none to 10 cm= maximum). SCORAD total score will be computed as specified in Section 5.5.

- The POEM tool is used to monitor AD severity. It focuses on the illness as experienced by the subject. It consists of a 5-point scale measuring the frequency of each of 7 AD symptoms (i.e., dryness, itching, flaking, cracking, sleep loss, bleeding, and weeping) over the past week scored from 0 (no days), 1 (1 to 2 days), 2 (3 to 4 days), 3 (5 to 6 days), and 4 (every day). POEM total score will be computed as specified in Section 5.5.

12.3.1 Main Analysis

Missing other efficacy endpoints will not be imputed.

Duration, cumulative time, and percent of time endpoints will be summarized using descriptive statistics by treatment group (refer to Section 5) and two-sided 95% CI of the mean will also be provided.

For duration endpoints, the number and percentage of subjects experiencing a status change or censored will be provided overall and by treatment group, along with reason for censoring. Additionally, these endpoints will also be analyzed using Kaplan–Meier (KM) method. The KM estimates of the time-to-event will be provided overall and by treatment group for the 25th, 50th (median), and 75th percentiles along with their corresponding two-sided 95% CI. The estimates of the SEs will be computed using the Greenwood's formula. The two-sided 95% CI for the survival function at the 25th, 50th, and 75th percentiles will be calculated according to Brookmeyer and Crowley (1982)[2]. KM curves will be provided overall and by treatment group.

Dichotomic, change from baseline, and percent change from baseline endpoints will be summarized similarly as the secondary efficacy endpoints (refer to Section 12.2).

12.3.2 Sensitivity Analyses

No sensitivity analyses will be performed for the other efficacy endpoints.

12.3.3 Subgroup Analyses

The main analysis of the other efficacy endpoints of achievement of at least a 4-point reduction on the WI-NRS, EASI-50, and EASI-75 will be repeated for the subgroups mentioned in Section 6.6 (including subgroup analysis for achievement of at least a 4-point reduction on the WI-NRS among subjects < 12 years old at entry in parent study). Additional subgroup analyses for duration, cumulative time and percent of time endpoints are also described in Section 6.6. Subgroup analyses will only be conducted for results relative to the primary baseline unless otherwise indicated in this SAP.

Subgroup analyses will not be performed for any other efficacy endpoints.

13 SAFETY ANALYSES

Safety analyses will be conducted using the safety population. No p-values or confidence intervals will be provided for safety endpoints, with the exception of EAIRs (refer to Section 5.5) which will be accompanied by 95% CIs.

13.1 Adverse Events

AEs will be coded according to the latest available version of MedDRA (Version 25.0 or higher).

A treatment emergent adverse event (TEAE) is defined as an AE that started post application of study medication at the Day 1 visit of study ARQ-151-313 or was present at treatment initiation but worsened during treatment, through study completion. See **Appendix 2** for handling of completely or partially missing dates for AEs. In the case where it is not possible to define an AE as treatment emergent or not, the AE will be classified as treatment emergent.

An overall summary of TEAEs will be presented by treatment group as defined in Section 5, which will include the total number of events, and the number and percentage of subjects who experienced at least one TEAE, TEAE by the maximum severity, TEAE by relationship and TEAE by the strongest relationship, treatment-emergent SAE, treatment-emergent Non-SAE, TEAE leading to study treatment discontinuation, TEAE leading to study discontinuation, TEAE on an application site, and TEAE leading to death. In addition, similar summary as described above will be presented by age group (refer to Section 6.6) for rollovers from study ARQ-151-311/312.

The number and percentage of subjects who experience at least one TEAE will be summarized by SOC, PT, and treatment group as defined in Section 5. A subject experiencing the same TEAE multiple times within the same PT will be counted only once for that PT. Similarly, a subject experiencing the same TEAE multiple time within the same SOC will be counted only once for

that SOC. TEAEs will be sorted alphabetically by SOC and within each SOC, the PT will be presented by descending frequency. In addition, similar summary as described above will be presented by age group (refer to Section 6.6) for rollovers from study ARQ-151-311/312.

The number and percentage of subjects who experience at least one TEAE will also be summarized by SOC, PT, strongest relationship with study treatment (related or not related) and treatment group. A treatment-related TEAE is defined as any TEAE that is assessed by the Investigator as likely, probably, or possibly related to study treatment. TEAE that is assessed as unrelated or unlikely will be defined as not treatment-related. If a subject experiences more than one TEAE within different relationship categories within the same SOC/PT, only the worst case (the strongest relationship) will be reported. TEAE with an unknown relationship will be considered as treatment-related.

The number and percentage of subjects who experience TEAE will also be summarized by SOC, PT, maximum severity (mild/moderate/severe/life threatening/death related to AE), and treatment group as defined in Section 5. If a subject experiences more than one TEAE within different severity categories within the same SOC/PT, only the worst case (the maximum severity) will be reported. TEAE with an unknown severity will be considered as severe.

The number and percentage of subjects who had treatment-emergent SAE will be summarized by SOC, PT, and treatment group as defined in Section 5. TEAEs leading to study drug discontinuation and TEAEs leading to study discontinuation will be summarized similarly.

Frequency and percentage of subjects who experience TEAE on an application site will be summarized by SOC, PT, and treatment group as defined in Section 5.

A plot of most frequent TEAE (≥ 1%) will be provided by PT and treatment group as defined in Section 5 (overall TEAE and overall treatment-emergent SAE will be included in the same plot).

The EAIR and the corresponding 95% CI will be reported for all tables of TEAEs by SOC and PT described earlier in this section. The EAIR per 100-pt year, defined in Section 5.5, will be calculated in each arm and the 95% CI will be calculated by using a Poisson regression model with number of TEAEs as dependent variable and log of exposure time as offset. Additionally, the difference in EAIR between the 'Roflumilast cream 0.15%' and 'Vehicle / Roflumilast cream 0.15%' treatment groups (refer to Section 5) will be provided along with the associated two-sided 95% CI.

All the AEs will be listed. TEAEs with an outcome of death, treatment-emergent SAEs, TEAEs leading to study discontinuation, and TEAE leading to study treatment discontinuation will be listed. The TEAE related to application site will be flagged in the AE listing.

13.2 Clinical Laboratory

For continuous hematology, chemistry, and urinalysis tests, observed values, changes from primary baseline, and percent changes from primary baseline will be summarized descriptively by visit and treatment group as defined is Section 5 . Categorical urinalysis tests will be summarized using frequencies by visit and treatment group. Of note, hematology, chemistry, and urinalysis tests will only be collected in subjects ≥ 12 years of age at time of enrollment in the parent study.

Shift tables from primary baseline to each post-baseline visit describing shifts to out-of-normal range will be provided for chemistry, hematology, quantitative and qualitative urinalysis tests. Only subjects with a baseline result and a result at the specified post-baseline visit for the test will be considered.

Listings of abnormal laboratory will be provided for each test where a subject had at least one abnormal result.

Laboratory data will be presented in SI units.

13.3 Vital Signs

For each vital sign (systolic blood pressure, diastolic blood pressure, heart rate, body temperature, weight, and height for subjects <18 years), observed values, changes from primary baseline and percent changes from primary baseline will be summarized descriptively by visit and treatment group as defined is Section 5.

The number and percentage of subjects with a gain or lose >5% in body weight from primary baseline to each OLE post-baseline visit will be summarized by visit and treatment group as defined in Section 5. A similar summary further broken down by age group (refer to Section 6.6) will also be provided.

For adults (≥ 18 years), the BMI observed values will be summarized using descriptive statistics by treatment group as defined in Section 5. For children (2 to 11 years) and adolescents (12 to 17 years), the BMI percentile rather than BMI observed value will be summarized using descriptive statistics by treatment group as defined in Section 5.

A listing of all vital sign assessments, including weight and BMI, will be provided.

13.4 Local Tolerability Assessments

The investigator's dermal response assessment of the application site reaction (0= no evidence of irritation to 7= strong reaction spreading beyond application site) will be summarized by visit and

treatment group as defined in Section 5 using both categorical methods (number and percentage of subject with each score) as well as continuous methods (descriptive statistics).

The investigator's assessment of the application site reaction other effects (slight glazed appearance, marked glazing, glazing with peeling and cracking, glazing with fissures, film of dried serous exudates, small petechial erosions and/or scabs, and no other effects) will be summarized using number and percentage of subjects in each category by visit and treatment group as defined in Section 5.

Local tolerability (burning/stinging scored as 0= none to 3= severe) as assessed by the subject (parents/guardian for subjects < 6 years of age) will be summarized similarly.

13.5 Patient Health Questionnaire Depression Scale (PHQ-8) and Patient Health Questionnaire Depression Scale (Modified PHQ-A)

The PHQ-8 assessment will be performed in adult subjects (≥18 years of age) while the modified PHQ-A assessment (question 9 has been removed since it is better evaluated by use of the C-SSRS tool) will be performed in adolescent subjects (12-17 years of age).

The PHQ-8/Modified PHQ-A score is the sum of the responses for the 8 questions, each question ranging from 0 (Not at all) to 3 (Nearly every day). The score will be computed as defined in Section 5.5.

Then, five severity categories of depression are defined based on the PHQ-8/Modified PHQ-A score as follows:

- 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)

The number and percentage of subjects in each category of depression will be summarized by visit and treatment group as defined in Section 5.

13.6 Children's Depression Inventory 2 (CDI-2)

The CDI-2 assessment will be performed in children subjects (6-11 years old, inclusive).

The observed values and changes from both primary and OLE baseline will be calculated for CDI-2 total score and the 2 subscales (refer to Section 5.5), and will be summarized descriptively by visit and previous treatment group in parent studies as defined in Section 5.

A CDI-2 total score ≥21 for females and ≥22 for males is considered as meeting clinically significant depression. The proportion of subjects meeting the clinically significant criteria will also be summarized by visit and previous treatment group in parent studies as defined in Section 5.

13.7 Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS questionnaire prospectively assesses Suicidal Ideation and Suicidal Behavior and will be performed for subjects 12 years of age and older. The "Since Last Visit" version will be used at all visits. The C-SSRS will be analyzed as per the Scoring and Data Analysis Guide from the Columbia Lighthouse Project (https://cssrs.columbia.edu/wp-content/uploads/ScoringandDataAnalysisGuide-for-Clinical-Trials-1.pdf).

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 as defined in Section 5.

13.8 Physical Examination

The number and percentage of subjects with normal and abnormal findings in the physical examination will be presented by body system, visit, and treatment group as defined in Section 5.

14 PHARMACOKINETICS ANALYSIS

A single PK draw will be performed at Week 24 (Cohort-Week 24) or Week 52 (Cohort-Week 52) for subjects ≥ 12 years of age at entry in parent study who originally enrolled into ARQ-151-313 under protocol amendment 1.

Concentration data will be summarized using descriptive and geometric statistics (refer to Section 5.6) by treatment group as defined in Section 5. For computation of mean plasma concentrations, data that are below the limit of quantification (BLQ) will be set to zero. The PK population will be used for these analyses.

PK data will be presented in the listing.

15 REFERENCES

Lott A, Reiter J.P. (2020) Wilson Confidence Intervals for Binomial Proportions with Multiple Imputation for Missing Data. The American Statistician 74:2, 109-115, DOI: 10.1080/00031305.2018.1473796

Brookmeyer R, Crowley J. A. (1982) Confidence Interval for the Median Survival Time. Biometrics 38:29-41.

16 APPENDICES

Appendix 1

Output Conventions

TLF will be generated using SAS® and will be displayed on letter size paper with landscape orientation, 1-inch margins and 9 pt. Courier New font.

The header section will comprise the sponsor's name, the protocol number, the delivery description, the data cut-off date (if applicable), the TLF number, the TLF title, the analysis set, and the page number (Page X of Y). The footer section will include the TLF footnotes, the CRO's name, the date and time of the execution of the program, and the name of the program.

The mean, median, geometric mean will be displayed to one more decimal place than the original value; Q1, Q3, minimum and maximum will keep the same number of decimal places as the original value; standard deviation, standard error, CV and CI will be displayed to two more decimal places than the original value. If derived parameters are to be summarized, the number of decimals of the derived values is to be chosen on a case-by-case basis, but the rule above applies.

For categorical summary tables, percentages will be reported to one decimal place. Percentages between 0 and 0.1 (both exclusive) will be displayed as "<0.1". The denominator for each percentage will be the number of subjects within the population per treatment group unless otherwise specified.

Listings will be ordered by treatment received in parent study and this study, subject number, date and visit (where applicable). Imputed dates will not be presented in the listings.

Dates & Times Format

Date and time (if available) will be presented in the format yyyy-mm-dd/hh:mm.

Presentation of Treatment Groups

When applicable, study treatments will be represented as follows in the different outputs:

For summary tables:

Parent Study Treatment Full Name	Study Treatment Output Name	Notes:
Roflumilast cream 0.15%	Roflumilast cream 0.15%	ARQ-151-311/312 roll-overs
Vehicle	Vehicle / Roflumilast cream 0.15%	ARQ-151-311/312 roll-overs
n/a	Pooled Roflumilast cream 0.15%	ARQ-151-311/312 roll-overs
Roflumilast cream 0.05%	Roflumilast Cream 0.05% or 0.15%	ARQ-151-315 roll-overs
Vehicle	Vehicle / Roflumilast cream 0.05% or 0.15%	ARQ-151-315 roll-overs
n/a	Pooled Roflumilast cream 0.05% or 0.15%	ARQ-151-315 roll-overs

For listings and figures:

Parent Study Treatment Full Name	Study Treatment Output Name	Notes:
Roflumilast cream 0.15%	Roflumilast cream 0.15%	ARQ-151-311/312 roll-overs
Vehicle	Vehicle / Roflumilast cream 0.15%	ARQ-151-311/312 roll-overs
Roflumilast cream 0.05%	Roflumilast Cream 0.05% or 0.15%	ARQ-151-315 roll-overs
Vehicle	Vehicle / Roflumilast cream 0.05% or 0.15%	ARQ-151-315 roll-overs

Appendix 2

Algorithm for Imputation of Start/End Date and Time of Adverse Events and Prior/Concomitant Medications

Event Start Date Imputation

- Imputation of event end date should be done before imputation of event start date.
- Completely missing: Impute to the first study treatment date.
- Missing day and month: Impute to January 1st, unless year is the same as year of first study treatment dose date then impute to the first study treatment date.
- Missing day: Impute to the 1st of the month, unless month and year are the same as month and year of first study treatment dose date then impute to the first study treatment date.
- If imputed event start date is after event end date (imputed or not), set the event start date to the imputed event end date.

Event Start Time Imputation (for Adverse Events only)

- Imputation of event end time should be done before imputation of event start date.
- If the event start date is not the same as the first dose date or time part of the first dose date is missing, impute to 00:00.
- If the event start date is the same as the first dose date, impute to time part of first dose date.
- If the event start date is equal to event end date and imputed event start time is after event end time (imputed or not), set the event start time to the imputed event end time.

Event End Date Imputation

- Completely missing (and not flagged as “ongoing”): Impute to the last contact date.
- Missing day and month: Impute to December 31st, unless year is the same as last contact date then impute to the last contact date.
- Missing day: Impute to the last day of the month, unless year and month are the same as year and month of last contact date then impute to the last contact date.

Event End Time Imputation (for Adverse Events only)

- If the event end date is not the same as the last contact date or time part of the last contact date is missing, impute to 23:59.
- If the event end date is the same as the last contact date, impute to time part of last contact date.