Protocol ARQ-151-311

A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis

Sponsor:	Arcutis Biotherapeutics, Inc. 3027 Townsgate Road, Suite 300 Westlake Village, CA 91361
Sponsor Representative:	[redacted]
Medical Monitor:	[redacted]
IND Number:	[redacted]
Protocol Version:	Amendment 4
Date:	28 September 2022

GCP Statement

This study is to be performed in full compliance with the protocol, International Conference on Harmonisation Good Clinical Practices (ICH GCP), and applicable regulatory requirements. All required study documentation will be archived as required by regulatory authorities.

Confidentiality Statement

SITE INVESTIGATOR SIGNATURE PAGE

ARQ-151-311

A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis

SPONSOR: Arcutis Biotherapeutics, Inc.
3027 Townsgate Road, Suite 300
Westlake Village, CA 91361

ISSUE DATE: 28 September 2022

I have read this protocol, including all appendices, and I agree that it contains all necessary details for me and my staff to conduct this study as described. I will conduct this study as outlined herein, in accordance with the current International Conference on Harmonisation Good Clinical Practices (ICH GCPs) and applicable local and regional regulations.

I will provide all study personnel under my supervision copies of the protocol and access to all information provided by Arcutis Biotherapeutics, Inc. I will discuss the material with them to ensure that they are fully informed about ARQ-151 and the study.

I agree that I or my designee will completely inform all subjects in this study concerning the pertinent details and purpose of the study prior to their agreement to participate in the study in accordance with cGCPs and regulatory authority requirements. I will be responsible for maintaining each subject's consent form in the study file and providing each subject with a signed copy of the consent form.

I agree to maintain the confidentiality of all information received or developed in connection with this protocol.

Investigational Site Name: [redacted]

Print Investigator Name: [redacted]

Investigator Signature: [redacted] Date: [redacted]

SUMMARY OF CHANGES

The following sections have been changed in Amendment 4 of the ARQ-151-311 protocol:

Version/Date	Description
31 October 2020	Original Protocol
Amendment 1 08 June 2021	• Added Summary of Changes section. • Revised the proportion of adult subjects in the study from up to 25% to up to 50%. This change was made to the synopsis, study schema, and Section 4.2 "Number of Sites and Subjects".
Amendment 2 18 July 2021	• Updated Arcutis Biotherapeutics, Inc. address. • Updated Medical Monitor information. • Updated Protocol Synopsis to reflect protocol changes. • Clarification added that if an unscheduled visit is required for reasons other than safety, the following assessments are not required: – vIGA-AD and EASI – BSA affected with AD – Local tolerability assessment (by Investigator) • WI-NRS: added statement that subjects will be given instructions on how to complete this questionnaire for subjects 6 years of age and older. • Added clarification that only subjects 18 years and older will be enrolled at sites located in the province of Québec in Canada. • Updated Figure 1 to note that subjects who have undergone a bilateral tubal ligation/occlusion are considered females of childbearing potential using a highly effective method of contraception. • Updated subject ages for hematology, serum chemistries, and urine analysis: all subjects will be evaluated at Screening but only subjects ≥12 years old will be evaluated at Baseline/ Day 1 and at Week 4/Day 29/ET and have a PK sample collected at Week 4/Day 29/ET. • Updated that all AEs occurring after the first application of IP through the end of the study should be collected, and all SAEs should be collected starting at Screening. • Added estimand language in the primary analysis section. • Updated "key secondary endpoints" to "secondary endpoints" and updated the language to be consistent throughout the document. • Updated the multiple testing procedure to hierarchical testing for secondary endpoint family 2. • Added per protocol population.

Version/Date	Description
	• Removed the language for continuous endpoint analysis using ANCOVA model and will be added in SAP. • Removed time to event endpoints. • Added new endpoints of vIGA-AD success at Week 1 and Week 2, vIGA-AD of 'clear' or 'almost clear' at Week 1, Week 2 and Week 4 and WI-NRS success at Week 1 and Week 2. • Updated wording of endpoints to make it consistent throughout the documents.
Amendment 3 23 May 2022	• Updated Medical Monitor information. • Revised the wording for the Per protocol population with the addition of the bolded text: – Per protocol (PP) population will include all subjects in the ITT population, who are at least 80% compliant with study medication application, **have a vIGA-AD assessment within the Week 4 visit window,** and show no major deviations from the study protocol that would affect the interpretation of efficacy. • Revised the wording for the WI-NRS population with the addition of the bolded text: – WI-NRS population will be a subset of the ITT population **who are ≥12 years old at Baseline and have a** Baseline WI-NRS score ≥4. • Revised the secondary endpoints to be tested using the ITT population, with the exception of the WI-NRS endpoints, which will use the WI-NRS population. • Minor grammatical/editorial changes throughout the document.
Amendment 4 28 September 2022	• Updated Sponsor Representative. • Updated the method of handling the intercurrent events of discontinuation due to adverse event or lack of efficacy to identify these subjects as non-responders at any visit that occurred or would have occurred after the date of last dose of study treatment. • Minor grammatical/editorial changes throughout the document.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Definition
α	Alpha Level (significance level)
AE	Adverse Event
AMP	Adenosine Monophosphate
AD	Atopic Dermatitis
AUC	Area Under the Curve
BSA	Body Surface Area
CDI	Children's Depression Inventory
CDLQI	Children's Dermatology Life Quality Index
C_{max}	Maximum Concentration
cm	Centimeter
COPD	Chronic Obstructive Pulmonary Disease
CRF	Case Report Form
C-SSRS	Columbia-Suicide Severity Rating Scale
CTCAE	Common Terminology Criteria for Adverse Events
DFI	Dermatitis Family Impact
DNA	Deoxyribonucleic Acid
DLQI	Dermatology Life Quality Index
EASI	Eczema Area and Severity Index
FDA	U.S. Food and Drug Administration
FOCBP	Female of Child Bearing Potential
FSH	Follicle-Stimulating Hormone
GCP	Good Clinical Practices
GLP	Good Laboratory Practices
hr	Hour
IB	Investigational Brochure
IC_{50}	Half Maximal Inhibitory Concentration
ICF	Informed Consent Form
ICH	International Conference on Harmonisation
ID	Identification
IGA	Investigator Global Assessment

Abbreviation	Definition
IL	Interleukin
IND	Investigational New Drug
IP	Investigational Product
IRB	Institutional Review Board
ITT	Intent to Treat
IWRS	Interactive Web Response System
Kg	Kilogram
LED	Light Emitting Device
µg	Microgram
MedDRA	Medical Dictionary for Regulatory Activities
Mg	Milligram
Min	Minute
mL	Milliliter
MTD	Maximum Tolerated Dose
NCI	National Cancer Institute
NIH	National Institutes of Health
NOAEL	No Observed Adverse Effect Level
Ng	Nanogram
P-450	Cytochrome P450
PDE-4	Phosphodiesterase 4
PDMP	Protocol Deviation Management Plan
PHQ-8	Patient Health Questionnaire-8
PI	Principal Investigator
PK	Pharmacokinetics
POEM	Patient-Oriented Eczema Measure
QD	Once Daily ("quaque die")
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SCORAD	Scoring Atopic Dermatitis
SD	Standard Deviation
SUSAR	Suspected Unexpected Serious Adverse Reaction

Abbreviation	Definition
TCPS	Tri-Council Policy Statement
TEAE	Treatment Emergent Adverse Event
T_{max}	Time to Reach Maximum Concentration
TPA	Target Plaque Area
TPSS	Target Plaque Severity Score
US	United States
UVR	Ultraviolet Radiation
V79	Chinese Hamster Cell Line
vIGA-AD	Validated Investigator Global Assessment - Atopic Dermatitis
WI-NRS	Worst Itch - Numeric Rating Score

1. PROTOCOL SUMMARY

1.1. Synopsis

Protocol Title:	A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis
Clinical Indication:	Atopic Dermatitis
Investigational Product:	• ARQ-151 will be supplied as an emollient cream at 0.15% strength • Matching vehicle cream will contain only excipients of ARQ-151
Study Design:	This is a Phase 3, parallel group, double blind, vehicle-controlled study in which ARQ-151 cream 0.15% or vehicle is applied QD for 4 weeks to subjects 6 years of age and older with mild to moderate atopic dermatitis. At entry, subjects will have ≥3% BSA involvement (excluding the scalp, palms, soles) and mild or moderate atopic dermatitis (AD) based on vIGA-AD assessment. Upon determination of eligibility, subjects will be randomized 2:1 to either ARQ-151 cream 0.15% cream or matching vehicle cream. The randomization will be stratified by vIGA-AD score at Baseline/Day 1 ('Mild' vs. 'Moderate') and by study site. Subjects/caregivers will apply ARQ-151 cream 0.15% or vehicle cream QD to all AD affected areas and any newly appearing AD lesions that arise during the study, except on the scalp. Subjects/caregivers should maintain treatment of these areas with study drug for the duration of the study regardless of whether treatable areas of AD clear prior to Week 4/Day 29. At the Week 4 visit, subjects may be eligible to enroll in a 12-month, open label extension study (ARQ-151-313) evaluating ARQ-151 cream 0.15% QD.
Study Objective:	To assess the safety and efficacy of ARQ-151 cream 0.15% vs vehicle administered QD x 4 weeks to individuals with atopic dermatitis.
Study Sites:	Approximately 50 sites in the US and Canada. During the conduct of the study, additional countries and/or sites may be added if necessary.
Study Population:	Subjects will be male and female children and adolescents (6-17 years), and adults (≥18 years). Subjects will have mild to moderate atopic dermatitis involvement with a vIGA-AD score of '2' (Mild) or '3' (Moderate) for study entry. Up to ~50% of the subjects will be ≥18 years old. Approximately 650 subjects are planned to be randomized in this study.

Duration of Participation for Subjects:	Screening (up to 30 days) + Treatment phase (4 weeks) for a total of about 8 weeks. Upon completion of the treatment phase of the study (Week 4/Day 29) subjects may have the opportunity, subject to regulatory approval and enrollment has not been completed, to participate in an open-label extension study (ARQ-151-313) of up to 12 months.
Inclusion Criteria:	1. For adult subjects: Participants legally competent to sign and give informed consent. For pediatric and adolescent subjects: Informed consent of parent(s) or legal guardian, and, if age appropriate, assent by the subjects, as required by local laws. 2. Males and females, ages 6 years and older at time of signing Informed Consent (Screening). Only subjects 18 years and older will be enrolled at sites located in the province of Québec in Canada. 3. Diagnosed with mild to moderate atopic dermatitis according to the criteria of Hanifin and Rajka (1980) prior to or at the screening visit. Subjects must have at least 3 of the 4 basic features per Hanifin and Rajka (1. Pruritus; 2. Typical morphology and distribution [flexural lichenification in adults and facial and extensor eruptions in infants and children]; 3. Chronic or chronically relapsing dermatitis; or 4. Personal or family history of atopy), in addition to 3 or more minor criteria. 4. History of AD for at least 3 months in subjects 6-17 years of age or 6 months in subjects ≥18 years of age, as determined by the Investigator using information from the subject's medical chart, from the subject's physician, or through subject/parent/caregiver interview. Stable disease for the past 4 weeks with no significant flares in atopic dermatitis before screening. 5. EASI Score ≥5 at Baseline. EASI is evaluated for the entire body except the scalp, palms, and soles. 6. vIGA-AD score of 'Mild' ('2') or 'Moderate' ('3') at Baseline. The vIGA-AD is evaluated for the entire body except the scalp, palms, and soles. 7. Has AD involvement of ≥3% BSA (excluding the scalp, palms, soles) at Baseline. 8. Females of childbearing potential (FOCBP) must have a negative serum pregnancy test at Screening and a negative urine pregnancy test at Baseline/Day 1. In addition, sexually active FOCBP must agree to use at least one form of a highly effective or barrier method of contraception throughout the trial. The use of abstinence as a contraceptive measure is acceptable as long as this is a consistent part of a lifestyle choice and an acceptable backup method has been identified if the subject becomes sexually active.

	9. Females of non-childbearing potential should either be pre-menarchal, or post-menopausal with spontaneous amenorrhea for at least 12 months (post-menopausal status should be confirmed with FSH testing) or have undergone surgical sterilization (permanent sterilization methods include hysterectomy, bilateral oophorectomy, or bilateral salpingectomy). 10. In good health as judged by the Investigator, based on medical history, physical examination, vital signs, serum chemistry labs, hematology values, and urinalysis. 11. Subjects and parent(s)/legal guardian(s) are considered reliable and capable of adhering to the Protocol and visit schedule, according to the judgment of the Investigator.
Exclusion Criteria:	1. Subjects with any serious medical condition or clinically significant laboratory, vital signs, or physical examination abnormality that would prevent study participation or place the subject at significant risk, as judged by the Investigator 2. Liver function tests excursions that exceed: • AST or ALT >2X ULN • Total bilirubin: – >1.5 x ULN or – >ULN and ≤1.5 x ULN AND direct bilirubin is >35% of total bilirubin • ALP ≥2x ULN 3. Subjects who cannot discontinue medications and treatments prior to the Baseline visit and during the study according to Excluded Medications and Treatments (Table 2). 4. Has unstable AD or any consistent requirement for high potency topical steroids to manage AD signs or symptoms. 5. Subjects who have significant active systemic or localized infection (eg, molluscum contagiosum), including known actively infected AD, or have had any infection that required oral or intravenous administration of antibiotics, antifungal or antiviral agents within 14 days prior to Baseline/Day 1. 6. Subjects who are unwilling to refrain from prolonged sun exposure and from using a tanning bed or other artificial light emitting devices (LEDs) for 4 weeks prior to Baseline/Day 1 and during the study. 7. Subjects with skin conditions other than AD that would interfere with evaluations of the effect of the study medication on AD, as determined by the Investigator. Subjects with any condition on the treatment area which, in the opinion of the Investigator, could confound efficacy measurements, eg, molluscum contagiosum. 8. Subjects with known genetic dermatological conditions that overlap with AD, such as Netherton syndrome.

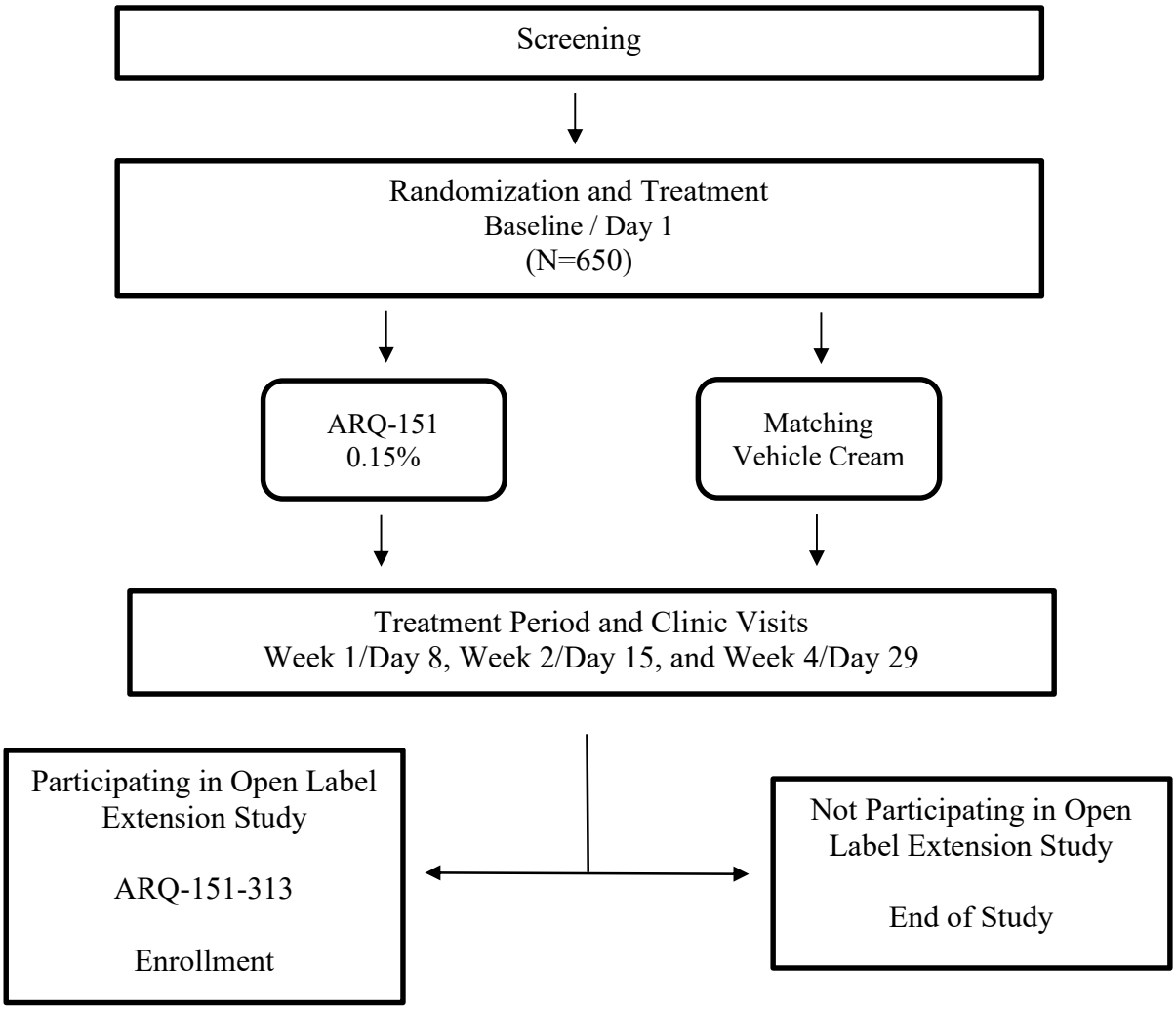
	9. Known allergies to excipients in ARQ-151 cream [redacted] 10. Subjects who cannot discontinue the use of strong cytochrome P-450 CYP3A4 inhibitors eg, indinavir, nelfinavir, ritonavir, clarithromycin, itraconazole, ketoconazole, nefazodone, saquinavir, suboxone and telithromycin for 2 weeks prior to Baseline/Day 1 and during the study period. 11. Subjects who cannot discontinue the use of strong cytochrome P-450 CYP3A4 inducers eg, efavirenz, nevirapine, glucocorticoids, barbiturates (including phenobarbital), phenytoin, rifampin, and carbamazepine for 2 weeks prior to Baseline/ Day 1 and during the study period. 12. Subjects who have received oral roflumilast (Daxas®, Daliresp®) within 4 weeks prior to Baseline/ Day 1. 13. Known or suspected: • Severe renal insufficiency – Severe renal insufficiency is defined as calculated creatinine clearance <30 mL/min. • Moderate to severe hepatic disorders (Child-Pugh B or C) 14. History of severe depression, suicidal ideation or behavior, Baseline/Screening C-SSRS (for adolescents and adults 12 years old and older) indicative of suicidal ideation or behavior, whether lifetime or recent/current. 15. Subjects with a PHQ-8 (adults) or modified PHQ-A (adolescents, 12-17 years old inclusive) score ≥10 at Screening or Baseline/Day 1 visits. 16. Subjects (6 to 11 years old, inclusive) with a CDI-2 (parent report) raw score ≥17 for females and ≥18 for males at Screening or Baseline/ Day 1 visits. 17. Females who are pregnant, wishing to become pregnant during the study, or are breast-feeding. 18. Previous treatment with ARQ-151. 19. Subjects currently undergoing allergy testing (eg, food allergy testing or skin prick testing), patch testing, food challenges, or allergy desensitization, or plan to do so during the study. 20. Subjects with any serious medical condition (eg, uncontrolled hypo- or hyper-thyroidism) or clinically significant laboratory abnormality that would prevent study participation or place the subject at significant risk, as determined by the Investigator. 21. Subjects with a history of a major surgery within 4 weeks prior to Baseline/Day 1 or subjects who have a major surgery planned during the study.

	22. Subjects with a history of chronic alcohol or drug abuse within 6 months prior to Screening. 23. Current or a history of cancer within 5 years with the exception of fully treated skin basal cell carcinoma, cutaneous squamous cell carcinoma or carcinoma in situ of the cervix. 24. Parent(s)/legal guardian(s) who are unable to communicate, read, or understand the local language(s). Subjects who are unable to communicate, read or understand the local language, or who display another condition, which in the Investigator's opinion, makes them unsuitable for clinical study participation. 25. Subjects who are family members of the clinical study site, clinical study staff, or sponsor, or family members of enrolled subjects (subjects enrolled in other studies of ARQ-151) living in the same house.
Key Assessments:	**Safety Assessments:** • Safety will be monitored through local tolerability assessments, vital signs, physical examination, safety labs, Children's Depression Inventory 2 (CDI-2, parent report for children 6-11 years old, inclusive), modified PHQ-A (for adolescents 12-17 years old, inclusive), PHQ-8 (for adults), C-SSRS (for adolescents and adults 12 years old and older), and AEs. • All AEs that occur after the first application of IP through the end of the study should be collected. All SAEs should be collected starting at Screening. • The investigator or a properly trained and designated subinvestigator will perform local tolerability assessments at Baseline/Day 1, and Weeks 1, 2, and 4 (Days 8, 15, and 29). Subjects will have vital signs measured at each study visit. Height will be collected at Visit 1 (Screening) only. • A limited physical exam (skin, lungs, and heart only) will be performed at Screening, Baseline/Day 1 and Week 4/Day 29. Blood and urine samples for routine safety laboratory tests (hematology, serum chemistry, and urinalysis) for all subjects will be obtained at Screening, and for subjects ≥12 years of age at Baseline/Day 1 and Week 4/Day 29/ET along with a PK sample collection at Week 4/Day 29/ET. For all female subjects of childbearing potential, a urine pregnancy test will be administered at all clinic visits except for Screening where a serum pregnancy test will be performed. A negative pregnancy result is required for continued participation in the study, and results (of the urine pregnancy test) must be available prior to dispensing of study drug at study visits. **Efficacy Assessments:** • Efficacy assessments will include vIGA-AD (Appendix 7), EASI, WI-NRS, BSA, DLQI/CDLQI (Appendix 8 and Appendix 9), DFI (Appendix 10), SCORAD (Appendix 11), and POEM (Appendix 12).

	Pharmacokinetic Assessment: • A single PK assessment (trough) will be performed in all subjects with a blood sample collected at Week 4/Day 29.
Study Endpoints:	1. Primary Efficacy Endpoint: • IGA Success, defined as a vIGA-AD score of 'clear' or 'almost clear' PLUS a 2-grade improvement from Baseline at Week 4 2. Secondary Efficacy Endpoints: • In subjects with a vIGA-AD score of 'Moderate' at randomization, vIGA-AD Success at Week 4 • In subjects ≥12 years old with Baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 4 • In subjects ≥12 years old with Baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 2 • In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 1 • Achievement of at least a 75% reduction in the Eczema Area and Severity Index (EASI-75) at Week 4 • vIGA-AD of 'clear' or 'almost clear' at Week 4 • vIGA-AD Success at Week 2 • vIGA-AD Success at Week 1 • vIGA-AD of 'clear' or 'almost clear' at Week 2 • vIGA-AD of 'clear' or 'almost clear' at Week 1 3. Pharmacokinetic endpoints include concentrations of roflumilast and its N-oxide metabolite.
Power and Sample Size:	Approximately 650 subjects are planned to be randomized in this study. To test the secondary endpoint of IGA success in subjects with a vIGA-AD score of 'Moderate' at randomization, approximately 490 of the subjects to be accrued will have vIGA-AD score of 'Moderate' at randomization. Randomization will be stratified by vIGA-AD score ('Mild' vs. 'Moderate') and by study site. This sample size provides approximately 95% power to detect an overall 15% difference between treatment groups on vIGA-AD success at Week 4 at α=0.05 using a 2-sided stratified Cochran-Mantel-Haenszel test. The results from a recent phase 2 study (ARQ-151-212) of ARQ-151 cream 0.15% compared to vehicle treatment were used to estimate the treatment difference. Specifically, in the phase 2 trial, approximately 37% of subjects demonstrated vIGA-AD success at Week 4 in the ARQ-151 0.15% group compared to 22% in the vehicle group. This sample size also provides approximately 90% power to detect an overall 17% difference between treatment groups on IGA success at Week 4 among subjects with vIGA-AD score 'Moderate' at randomization. The same testing method, the stratified Cochran-Mantel-Haenszel test, will be used as for the primary endpoint.

Statistical Analysis:	The analysis populations are defined as follows: • Intent-to-Treat (ITT) population will include all subjects who are randomized. • Per protocol (PP) population will include all subjects in the ITT population, who are at least 80% compliant with study medication application, have a vIGA-AD assessment within the Week 4 visit window, and show no major deviations from the study protocol that would affect the interpretation of efficacy. In addition, subjects who miss the Week 4 vIGA-AD assessment specifically due to novel coronavirus disease-19 (COVID-19) disruptions will be excluded from per protocol population. • vIGA-AD Moderate ITT population will be a subset of the ITT population with vIGA-AD score 'moderate' at randomization. • vIGA-AD Moderate PP population will be a subset of the PP population with vIGA-AD score 'Moderate' at randomization. • WI-NRS population will be a subset of the ITT population who are ≥12 years old at Baseline and have a Baseline WI-NRS score ≥4. • Safety population will include all subjects who are enrolled and received at least one confirmed dose of study medication. • PK population will include all subjects receiving the active drug with quantifiable plasma concentrations of roflumilast. To control for familywise type I error at level of 0.05, the secondary endpoint of IGA success at Week 4 for subjects with vIGA-AD score of 'Moderate' at randomization will only be tested if the primary endpoint demonstrates statistical significance. In addition, the remaining secondary endpoints will be inferentially tested only if the primary and above secondary endpoint (IGA success at Week 4 for subjects with vIGA-AD score of 'Moderate' at randomization) comparisons are statistically significant using hierarchical testing procedure by partitioning of the alpha and use of the Fallback Method. Descriptive statistics for continuous variables will include mean, median, standard deviation, Q1, Q3, min, max. Descriptive statistics for categorical variables will include frequencies and percentages. For missing data, the primary imputation method and sensitivity methods will be detailed in the SAP. The primary endpoint will be analyzed with a Cochran-Mantel-Haenszel test stratified by the randomization factors (disease severity determined by vIGA-AD and by study site). Categorical secondary efficacy analysis will be analyzed in the same manner as the primary endpoint. The incidence of adverse events will be summarized as well as laboratory parameters and vital signs.

1.2. Study Schema



A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis

Approximately 650 subjects with atopic dermatitis will be randomized 2:1 to receive either:

- ARQ-151 cream 0.15% or Vehicle cream

Subjects will have ≥3% BSA involvement (excluding the scalp, palms, soles) with a vIGA-AD score of '2' (mild) or '3' (moderate) for study entry

Up to ~50% of the subjects will be ≥18 years old

1.3. Schedule of Visits and Assessments

Study Procedure	Screen	Baseline Day 1	Wk 1 Day 8	Wk 2 Day 15	Wk 4 Day 29 / ET
Visit	**1**	**2**	**3**	**4**	**5**
Visit Window	**-30 days**	**N/A**	**+/- 3 days**	**+/- 3 days**	**+/- 3 days**
Informed consent/assent	X				
Demographics	X				
Medical and surgical history	X				
Physical examination[a]	X	X			X
I/E criteria	X	X			
Hematology, Serum Chemistries, and Urine Analysis[b]	X[b]	X[b]			X[b]
Vital signs, height, weight[c]	X	X	X	X	X
vIGA-AD, EASI, BSA, SCORAD[d]	X	X	X	X	X
WI-NRS pruritus[e]	X	X	X	X	X
POEM[f]	X	X	X	X	X
Local Tolerability Assessment[g]		X	X	X	X
CDI-2, PHQ-8, PHQ-A, C-SSRS[h]	X	X	X	X	X
DLQI, CDLQI, DFI[i]	X	X	X	X	X
Medical Photography[j]		X	X		X
Serum pregnancy test (FOCBP only)	X				
Urine pregnancy test[k]		X	X	X	X
PK draws[l]					X
Drug/vehicle application in clinic[m]		X	X	X	
Dispense/Re-dispense study medication kit[n]		X	X[o]	X[o]	X[o]
Dispense/review diary	X	X	X	X	X
Weigh study medication kit[p]		X	X	X	X
Compliance determination[q]			X	X	X
Adverse event assessment[r]	X	X	X	X	X
Concomitant medications	X	X	X	X	X
Study Exit[s]					X

Footnotes from table above:

[a] Limited physical examination: skin (including assessment of Fitzpatrick skin type at Screening only), lungs, and heart only

[b] For all subjects entering this study under Amendment 2, to be collected at Screening, but subsequent samples will be collected only for subjects ≥12 years old (Baseline/Day 1 and Week 4/Day 29/ET). For subjects 12 to 18 years of age, if Baseline/Day 1 is within 3 weeks of Screening, the Screening results may be used.

[c] Height will be collected at Screening only. 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 other 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. **The vIGA-AD assessment should be completed prior to other physician assessments**. SCORAD total score will range between 0 and 103.

[e] Subjects will self-assess their pruritus at home on a daily basis starting 7 days prior to the Baseline/Day 1 visit, and then every day thereafter. WI-NRS score will be determined by the subject assessing worst itch over the past 24 hours. The scale is from 0 (no itch) to 10 (worst itch) and this value will be recorded by the subject each day. Subjects will be trained at the Screening visit in the accurate completion of the WI-NRS. In addition, parents/ caregivers of children and adolescent subjects will be trained at the Screening visit by study staff on how to assist the subject, if needed, in completing the WI-NRS.

[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] Local tolerability assessments should be recorded prior to study drug application for the Investigator assessment of skin irritation (Berger and Bowman skin irritation score). **Note for investigator tolerability assessments: reactions at the site of product application, which may occur post-Baseline, should be differentiated from the preexisting inflammation associated with the subject's atopic dermatitis.** The subject will assess burning/stinging (0-3 score) 10-15 minutes post drug application. **Note subject burning stinging assessment: at Day 29, subjects will provide a recall assessment of burning/stinging experienced post drug application on the previous day (Day 28).**

[h] Adolescents and adults will complete the C-SSRS (12 years of age and older). Adults will complete the PHQ-8. Adolescents (ages 12 to 17, inclusive) will complete the PHQ-A (PHQ-9 modified). Parents/caregivers will complete CDI-2 (parent report) for children 6-11 years of age, inclusive.

[i] The DLQI will be completed by subjects ≥17 years of age. The CDLQI will be completed for subjects 6 to 16 years old, inclusive. The Dermatitis Family Impact Questionnaire (DFI) will be completed by parents/caregivers for all subjects 6 to ≤17 years of age.

[j] Photography of AD lesion(s) selected by the Investigator will be performed at all investigational sites. All efforts will be made to de-identify the subjects. Canfield equipment will be used to capture photographs. Subjects who are unwilling to participate in the medical photography will be allowed to opt out of this procedure, as documented on the Informed Consent Form.

[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 entering this study under Amendment 2, a single PK trough draw will be collected at Day 29 only for subjects ≥12 years old (Baseline/Day 1 and Week 4/Day 29/ET). Ensure study medication was not applied in the area where PK will be drawn.

Footnotes from table above:

[m] Subjects to apply assigned IP during clinic visits, except for the Day 29/ET visit.

[n] It is expected that kits will be dispensed based on %BSA affected. See IP Handling Manual for details.

[o] On Day 8 and 15, dispensing of IP is optional. Site should review IP kit to ensure sufficient IP is available until the next visit and only dispense additional IP if needed. On Day 29, if the subject is unable to perform the Day 29 clinic visit due to COVID-19 restrictions (isolation, quarantine, etc.) then additional IP may need to be dispensed so IP can continue to be applied at home until the subject is able to return to the clinic to complete the Day 29 assessments (see IP Handling Manual for the process to dispense additional IP at or after Day 29).

[p] Every tube should be weighed and recorded when dispensed and returned. See IP Handling Manual for details.

[q] Compliance determination is described in the IP Handling Manual

[r] All AEs should be collected starting after the first application of the investigational product through the end of the study. All SAEs should be collected starting after the signing of the informed consent through 30 days after the last day of the application of the investigational product or the end of the study (whichever is later). Any AEs (whether serious or non-serious) and clinically abnormal laboratory test values(s) will be evaluated by the PI and treated and/or followed up for up to 30 days after end of treatment or until symptoms or value(s) return to normal, or acceptable level, as judged by the PI (if the subject is continuing into the ARQ-151-313 OLE study, then AEs from this study (ARQ-151-311) will only be followed until exit from this study).

[s] Subjects who enroll into the open label extension study (ARQ-151-313) must complete the ARQ-151-311 visit requirements at Week 4.

2. INTRODUCTION

2.1. Background

Roflumilast is a phosphodiesterase 4 (PDE-4) inhibitor approved globally to reduce the risk of exacerbations in patients with severe chronic obstructive pulmonary disease (COPD) associated with chronic bronchitis. Roflumilast and its active metabolite, roflumilast N-oxide, are high affinity selective inhibitors of PDE-4 (a major cyclic-3′,5′-adenosine monophosphate (cyclic AMP)-metabolizing enzyme), whose activity leads to accumulation of intracellular cyclic AMP. There are four different subtypes of PDE-4: PDE-4a, PDE-4b, PDE-4c, and PDE-4d, each with several isoforms (splicing variants). IC_{50} values of both roflumilast and roflumilast N-oxide for the different PDE-4 isoforms and subtypes are mostly sub-nanomolar and single digit nanomolar (Hatzelmann 2010). The PDE-4 family of enzymes are the most prevalent phosphodiesterases in immune cells and inhibition of PDE-4 subtypes has been associated with anti-inflammatory effects in many biological systems.

Roflumilast was initially developed as a 500 µg tablet for oral therapy in patients with COPD, and as such has been thoroughly evaluated in nonclinical studies. The safety profile is well-established. Oral roflumilast (500 µg tablet) was approved by Health Canada as DAXAS® in December 2010 and by the US FDA as DALIRESP® in February 2011 for the treatment of COPD. The study sponsor has conducted nonclinical studies in which roflumilast is applied dermally.

The dermal nonclinical program for ARQ-151 cream followed current International Conference on Harmonisation (ICH) guidelines and includes a 13-week dermal toxicity study in minipigs, a 13-week dermal toxicity study in mice, a 39-week dermal toxicity study in minipigs, a skin

sensitization study in guinea pigs, a phototoxicity study, an eye irritation study and a 104-week carcinogenicity study in mice, the in life portion of which is complete.

Refer to the current ARQ-151 Investigator's Brochure (IB) for the most current PDE-4 dermal and oral/systemic nonclinical and clinical information.

Atopic dermatitis is a chronic inflammatory skin disorder affecting children and adults, with the majority presenting with disease of mild to moderate severity. The use of topical corticosteroids and/or topical calcineurin inhibitors, in combination with emollients has been the mainstay for treating atopic dermatitis. In 2016, Eucrisa® (crisaborole), a PDE-4 inhibitor was approved for the topical treatment of mild to moderate atopic dermatitis in patients 2 years of age and older.

Topical calcineurin inhibitors block the activation of T-lymphocytes and diminish inflammation, but are accompanied by a boxed warning for the development of lymphomas and other lymphoproliferative diseases. Topical corticosteroids can cause skin atrophy and hypothalamic-pituitary axis suppression, and their use is often accompanied by poor adherence due to corticophobia (fear of using corticosteroids in patients or doctors). Eucrisa provides precedence for the effectiveness of topical PDE-4 inhibitors in atopic dermatitis, but it is a twice-daily ointment, its efficacy may be modest, and its use may be accompanied by burning, stinging, and local skin reactions. As a result, there is a need for the development of new topical products for the treatment of atopic dermatitis (Nygaard 2017).

The therapeutic use of PDE-4 inhibitors in AD is based on the recognized intracellular role of PDE-4 in keratinocytes (Dastidar 2007, Hanifin 1996). Circulating leukocytes in AD patients have PDE-4 activity, which has been associated with higher production of proinflammatory mediators and lower production of the anti-inflammatory mediator IL-10, in part due to hydrolyzation of cyclic adenosine monophosphate (cAMP) (Grewe 1982, Furue 2014, Baumer 2007). This consequently diminishes levels of cAMP, which leads to increased transcription of numerous cytokines, accelerating a number of intracellular functions involved in acute and chronic inflammation (Grewe 1982). Thus, targeting PDE-4 has been shown to directly attenuate inflammation due to inhibition of the breakdown of cAMP, consequently reducing the levels of tumour necrosis factor-α, IL-12, IL-23, and other signaling effectors (Murrell 2015, Nazarian 2009).

2.2. Conclusions on Toxicity Findings

The safety profile of oral roflumilast is well-established. An extensive systemic toxicity program that evaluated both roflumilast and its active N-oxide metabolite in multiple species via the oral route of administration was conducted to support registration of the 500 µg tablet for COPD.

The previously-conducted systemic toxicity program included studies to evaluate reproductive toxicity, genotoxicity and carcinogenicity, and the results of those studies are included in the labeling for oral roflumilast.

To support the development of ARQ-151 topical cream a GLP-compliant dermal toxicity program is being conducted. To date, no new risks have been identified through the dermal toxicity program. In 13-week dermal toxicity studies in minipigs and mice, and a 39-week dermal toxicity study in minipigs, no evidence of systemic toxicity was observed.

Histopathological evaluation of skin in the minipig study included very slight to slight erythema at all treatment levels, and minor degrees of irritation such as hyperplasia. The NOAEL in both studies was the 1% concentration of ARQ-151 (20 mg/kg), the highest dose administered and the maximum feasible concentration.

Across the dermal and systemic toxicology programs, the exposure to parent drug and N-oxide metabolite differs by route and species. While exposure to roflumilast and its active metabolite are likely to be higher following topical administration of ARQ-151 relative to oral administration, when the margins from the toxicity studies are considered as a whole, the NOAELs across routes and species provide assurance that the anticipated exposures with ARQ-151 cream will be safe.

2.3. Clinical Studies

2.3.1. Topical Roflumilast Cream

The formulation of topical roflumilast, ARQ-151 cream, has been evaluated in both plaque psoriasis (currently in Phase 3) and atopic dermatitis (through Phase 2).

2.3.2. Psoriasis Phase 2a (ARQ-151-101)

ARQ-151-101 (NCT03392168) was a Phase 2a study of two active doses of ARQ-151, 0.5% and 0.15% vs vehicle in the topical treatment of adult patients with chronic plaque psoriasis of up to 5% BSA involvement.

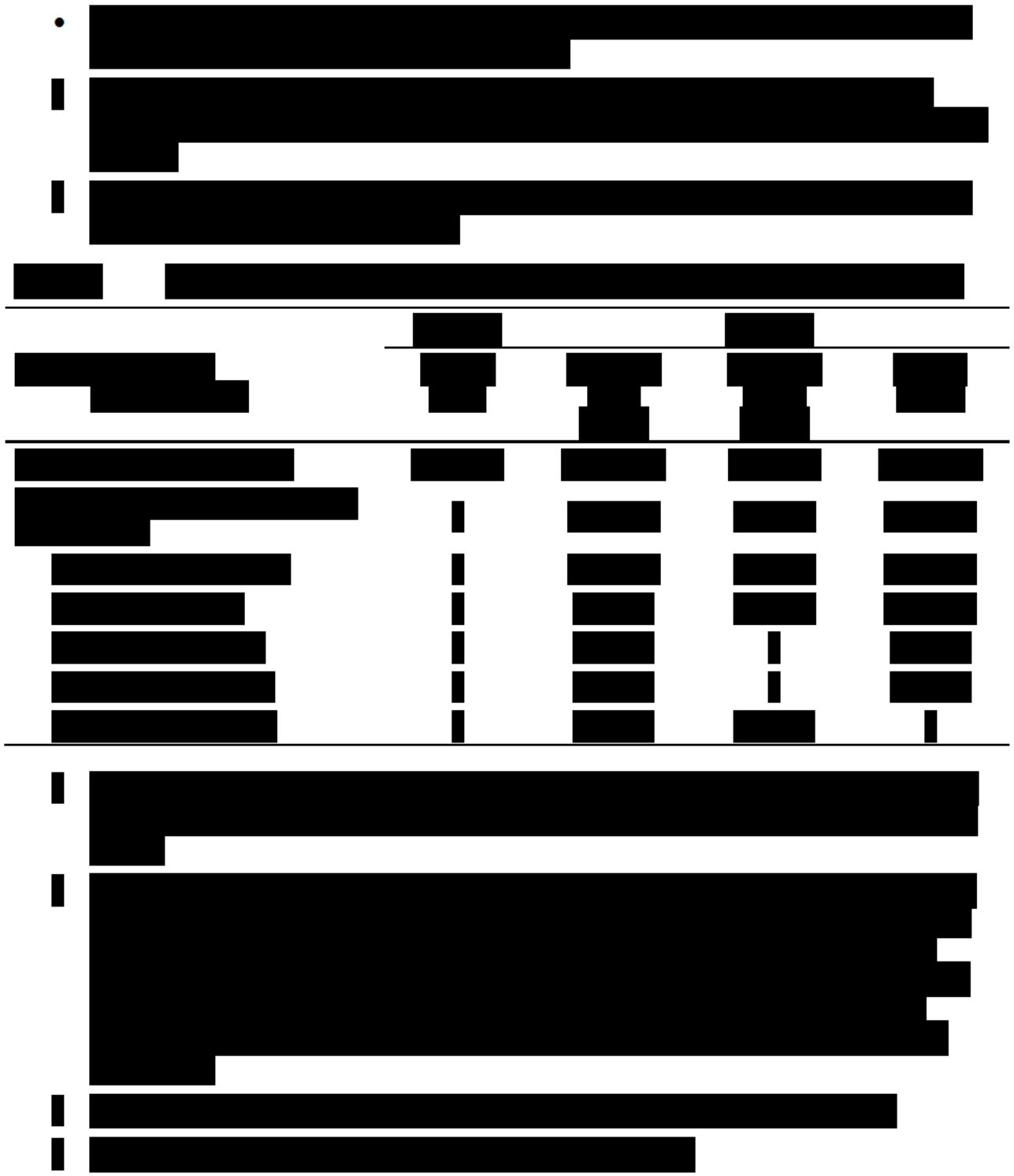
An initial cohort (Cohort 1) of 8 adult psoriasis subjects (no vehicle subjects) was treated as follows with the results indicated:

- Single dose application of ARQ-151 cream 0.5% to a 25 cm^2 area of psoriatic plaque on the trunk or extremities (not on the face, genital area, palms or soles)
- Local tolerability and systemic safety labs monitored
- PK assessments at baseline (pre-dose), 1, 2, 4, 6 and 24 hours
- Skin permeation of topically applied drug was ~0.4%
- Local tolerability and systemic safety labs were unremarkable

Six Cohort 1 subjects plus 83 additional psoriasis subjects were then enrolled into Cohort 2, an inter-individual, parallel group, randomized and blinded assessment of two concentrations of ARQ-151 drug product (0.15% and 0.5%) versus vehicle applied QD x 28 days, analyzing target psoriatic plaques for efficacy. Subjects were randomized 1:1:1 to receive 0.5% drug product, 0.15% drug product or vehicle to psoriatic plaques up to 5.0% of BSA. In each subject, up to 3 target plaques were identified for efficacy analysis.

PK assessments conducted on Day 1: 1, 2, 4 and 6 hours; Day 14: pre-dose (trough) and 1-hour post-dose; and Day 28: pre-dose (trough), 1, 2, 4, 6 and 24 hours.

Safety results follow:



Day 28 pharmacokinetic results of ARQ-151-101 are as follows:

- Following 28 days of ARQ-151 cream 0.5% and 0.15% application to a mean BSA% [redacted]

2.3.3. Psoriasis Phase 2b (ARQ-151-201)

ARQ-151-201 (NCT03638258) was a parallel group, double blind, vehicle-controlled study in which ARQ-151 cream 0.15%, ARQ-151 cream 0.3%, or vehicle cream was applied QD for 12 weeks to over 300 adult subjects with 2% to 20% BSA of chronic plaque psoriasis and baseline IGA of Mild or greater. In this study, both ARQ-151 cream 0.3% and ARQ-151 cream 0.15% were safe and well tolerated, demonstrating similar safety and tolerability profiles compared to each other and compared to vehicle. The safety data are summarized below:

- [redacted]

- [redacted]

Pharmacokinetic results of ARQ-151-201 are as follows:

- [redacted]

2.3.4. Atopic Dermatitis Phase 1 PK Study in Adults (ARQ-151-102)

ARQ-151-102 was an open label, Phase 1, pharmacokinetics and safety study of ARQ-151 Cream 0.15% and ARQ-151 Cream 0.05% administered QD in adult subjects with mild to moderate AD.

[redacted]

2.3.5. Phase 1 Study in Adolescents and Pediatrics (ARQ-151-105)

ARQ-151-105 (NCT04156191) is an ongoing open-label, Phase 1, pharmacokinetics, maximal usage PK, safety, and efficacy study of ARQ-151 cream 0.15% administered QD in adolescent and pediatric subjects with mild to moderate atopic dermatitis.

The study is being conducted in three parts, the first two of which are completed. The first part consisted of three cohorts in which subjects aged 2 to 17 years old had 1.5-35% BSA involvement (excluding the scalp, palms, soles) and mild or moderate atopic dermatitis based on vIGA-AD.

The second part of the study consisted of three cohorts in which subjects were evaluated under maximal use conditions (MUSE) and had BSA involvement (excluding the scalp, palms, soles) of ≥35% in subjects 2 to 11 years old (inclusive) or ≥25% in subjects 12 to <17 years old with mild or moderate atopic dermatitis based on vIGA-AD. At least 60% of the enrolled subjects had moderate atopic dermatitis.

The third part of the study will consist of one cohort (Cohort 7) in which subjects 2 to 5 years of age (inclusive) will be administered a lower concentration of ARQ-151 cream (0.05%) and evaluated under maximal use conditions (MUSE). Subjects will have BSA involvement (excluding the scalp, palms, soles) of ≥35% with mild or moderate atopic dermatitis based on vIGA-AD. At least 60% of the enrolled subjects had moderate atopic dermatitis.

Preliminary Study Results

[redacted]

2.3.6. Atopic Dermatitis Phase 2 Dose Ranging Study (ARQ-151-212)

ARQ-151-212 (NCT03916081) was a parallel group, double blind, vehicle-controlled, Phase 2 study that evaluated ARQ-151 cream 0.05% and 0.15% in the treatment of mild to moderate atopic dermatitis in 136 adolescent and adult subjects with 1.5 to 35% BSA of involvement.

Ninety-three female (68.4%) and 43 male (31.6%) subjects with mild to moderate AD participated in the study. Overall, the demographic and baseline disease characteristics were similar across all study groups. The mean age for all 136 study subjects was 41.6 years, including 8 adolescent subjects (between 12-17 years). The mean EASI score at Baseline for all study subjects was 9.04. The majority of subjects were in the moderate vIGA-AD category (77.9%). The mean BSA involvement was 9.5% for all study subjects.

[redacted]

2.3.7. Oral Roflumilast Tablet

Oral roflumilast (DALIRESP®) has been approved globally for the treatment of COPD and has been evaluated in nine Phase III/IV randomized double-blind clinical trials (Wedzicha 2016). Overall, the safety of oral roflumilast has been well established in its targeted population of mostly middle- and upper-aged individuals who currently smoke cigarettes or have smoked them extensively in the past. Adverse events (AEs) reported with roflumilast tablets have been consistent with those expected for oral PDE-4 inhibitors. In a pooled analysis of safety data from 6-month and 1-year clinical trials (N=8630), the most common AEs were diarrhea, weight loss and nausea. Other AEs reported more frequently with roflumilast treatment than with placebo were back pain, influenza, insomnia and decreased appetite (Michalski 2012, Wedzicha 2016).

In addition to the self-reported cases of weight loss in the 6-month and 1-year oral trials, clinically significant weight loss was also reported in two prospective studies that evaluated weight (Michalski 2012).

Psychiatric-related AEs were also greater in patients treated with roflumilast tablets (5.9%) compared to those treated with placebo (3.3%). The most common psychiatric-related AEs were insomnia, depression and anxiety. A small number of cases of completed suicide and suicide ideation have been reported in patients taking oral roflumilast in clinicals trials and also during post-marketing experience (Michalski 2012).

The only contraindication to oral roflumilast is use in patients with moderate to severe liver impairment (Child-Pugh B or C), where systemic levels of roflumilast may become elevated.

2.4. Rationale for Development

Atopic dermatitis is currently treated with topical calcineurin inhibitors and/or topical corticosteroids in combination with emollients. In 2016, Eucrisa® (crisaborole), a less potent PDE-4 inhibitor than roflumilast, was approved for the topical treatment of atopic dermatitis. Topical calcineurin inhibitors block the activation of T-lymphocytes and diminish inflammation, but are accompanied by a 'black box' warning for the development of lymphomas and other lymphoproliferative diseases. Topical corticosteroids can cause skin atrophy and hypothalamic-pituitary axis suppression, and their use is often accompanied by poor adherence due to corticophobia (fear of using corticosteroids in patients or doctors). Eucrisa provides precedence for the effectiveness of topical PDE-4 inhibitors in atopic dermatitis, but it is a twice-daily

ointment, its efficacy is modest, and its use is often accompanied by burning, stinging, and local skin reactions. In our Phase 2 AD study (ARQ-151-212), we observed excellent local toleration of ARQ-151 cream formulations. Since roflumilast is a more potent PDE-4 inhibitor than crisaborole (Hatzelmann 2010), we believe that ARQ-151 has potential to provide greater efficacy with better local toleration than Eucrisa.

This study will evaluate the safety and efficacy of ARQ-151 cream in children, adolescent, and adult subjects with mild to moderate atopic dermatitis.

2.4.1. Dose Selection

In ARQ-151-212, results for the primary efficacy endpoint, mean absolute change from baseline in EASI score at Week 4, were numerically higher in the ARQ-151 cream 0.05% and ARQ-151 cream 0.15% (p= 0.097) groups than in the vehicle group. Furthermore, the result of the sensitivity analysis of the primary endpoint at Week 4 was statistically significant (ARQ-151 cream 0.15%, p= 0.027). Statistical significance was reached for numerous other clinically important efficacy endpoints including % change from baseline in EASI score, EASI-75 responders, and patients achieving vIGA-AD score of clear or almost clear. Both doses of topical roflumilast (0.15% and 0.05%) had a similar and favorable safety and tolerability profile, with generally more favorable efficacy observed at the ARQ-151 cream 0.15% dose. Results of this study support the use of ARQ-151 cream 0.15% in studies of adult and adolescent subjects with mild to moderate AD. PK data from ARQ-151-105 support the use of the same concentration (0.15%) in subjects 6 to 11 years old and the use of a lower concentration in subjects 2 to 5 years old.

2.4.2 Risks and/or Benefits to Subjects

A favorable local and systemic benefit-risk profile has been observed in prior studies of ARQ-151. Subjects 6 years of age and older, included in this study, randomized to active treatment group may see an improvement in their atopic dermatitis with ARQ-151 0.15% cream, based on the activity of doses tested in atopic dermatitis (0.05%-0.15%) and psoriasis (0.15%-0.5%), and approval of a less potent topical PDE-4 inhibitor (crisaborole) for atopic dermatitis. Subjects may also see some benefit as the cream formulation of ARQ-151 may have a moisturizing effect.

Oral roflumilast has now been used for almost a decade in the treatment of COPD exacerbations and its safety record has been well-documented. The known adverse effects of oral treatment in the COPD population (nausea, vomiting, diarrhea, weight loss, psychiatric AEs (see Section 2.3.7) may be readily monitored as specified in this protocol. The profile that is emerging from studies of topical roflumilast appears different from the safety and tolerability profile of oral roflumilast. While oral PDE-4 inhibitors (DALIRESP, OTEZLA) have been associated with, in particular, a moderate incidence of GI AEs, these AEs, and perhaps others, appear to be reported far less frequently with topical PDE-4 inhibitors, including EUCRISA®, and ARQ-151 cream to date in clinical trials. For ARQ-151 cream, this may be related to the lack of 'peak to trough' C_{max} variation, lower C_{max} values than observed following oral administration, or bypassing of the gastrointestinal tract with topical administration.

This study has been designed with adequate safety monitoring practices (i.e., physical examinations, vital signs/weight, local skin toleration assessments, hematology, serum chemistry, urinalysis, PHQ-A/PHQ-8, CDI-2, C-SSRS and AE reporting).

3. STUDY OBJECTIVES AND ENDPOINTS

3.1. Study Objectives

3.1.1. Primary Objective

The primary objective of this study is to assess the safety and efficacy of ARQ-151 cream 0.15% vs vehicle administered QD x 4 weeks to individuals 6 years of age and older with atopic dermatitis.

3.2. Study Endpoints

3.2.1. Primary Efficacy Endpoint

The primary efficacy endpoint of this study is:

- IGA Success, defined as a vIGA-AD score of 'clear' or 'almost clear' PLUS a 2-grade improvement from Baseline at Week 4

3.2.2. Secondary Efficacy Endpoints

The secondary efficacy endpoints of this study are:

- In subjects with a vIGA-AD score of 'Moderate' at randomization, vIGA-AD Success at Week 4
- In subjects ≥12 years old with baseline WINRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 4
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 2
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 1
- Achievement of at least a 75% reduction in the Eczema Area and Severity Index (EASI-75) at Week 4
- vIGA-AD of 'clear' or 'almost clear' at Week 4
- vIGA-AD Success at Week 2
- vIGA-AD Success at Week 1
- vIGA-AD of 'clear' or 'almost clear' at Week 2
- vIGA-AD of 'clear' or 'almost clear' at Week 1

3.2.3. Pharmacokinetic Endpoint

- Pharmacokinetic endpoints include concentrations of roflumilast and its N-oxide metabolite.

4. INVESTIGATIONAL PLAN

4.1. Overall Study Design and Plan

This is a Phase 3, parallel group, double blind, vehicle-controlled study in which ARQ-151 cream 0.15% or vehicle is applied QD x 4 weeks to subjects with mild to moderate atopic dermatitis.

- At entry, subjects will have ≥3% BSA involvement (excluding the scalp, palms, soles) and mild or moderate atopic dermatitis (AD) based on vIGA-AD assessment.
- Upon determination of eligibility, subjects will be randomized 2:1 to either ARQ-151 cream 0.15% or matching vehicle cream. The randomization will be stratified by vIGA-AD score at baseline ('Mild' vs. 'Moderate') and by study site.
- Subjects/caregivers will apply ARQ-151 cream 0.15% or vehicle cream QD to all AD affected areas and any newly appearing AD lesions that arise during the study, except on the scalp. Subjects/caregivers should maintain treatment of these areas with study drug for the duration of the study regardless of whether treatable areas of AD clear prior to Week 4.
- At the Week 4 visit, subjects may be eligible to enroll in a 12-month, open label extension study (ARQ-151-313) in which they will receive ARQ-151 cream 0.15% QD.

4.2. Number of Sites and Subjects

A total of up to approximately 650 subjects will be randomized at approximately 50 study sites in the United States and Canada. Only subjects 18 years and older will be enrolled at sites located in the province of Québec in Canada. During the conduct of the study, additional countries and/or sites may be added if necessary. Subjects will be male and female children and adolescents (6-17 years), and adults (≥18 years). Subjects will have mild to moderate atopic dermatitis involvement with a vIGA-AD score of '2' (mild) or '3' (moderate) for study entry. Up to ~50% of the subjects will be ≥18 years old and up to 25% of subjects will have a vIGA-AD score of '2'.

4.3. Subject Participation

Subject participation involves a minimum of 5 clinic visits including Screening, Baseline, Week 1, Week 2, and Week 4. The interval between the Screening and Baseline visits could be up to 30 days, therefore the anticipated maximum duration of subject participation is approximately 8 weeks.

Upon completion of the treatment phase of the study (Week 4) subjects may have the opportunity, subject to regulatory approval and enrollment has not been completed, to participate in an open-label extension study (ARQ-151-313) of up to 12 months.

4.4. Numbering of Subjects

All subjects who sign an informed consent form will be assigned a unique 6-digit subject identification (ID) number by the IWRS system.

The subject identifier number is 6-digits (SXX-YYY) and will contain the study number-site number (where S = 1 and XX is this study site number such as 01, 02, etc.) and the subject number (YYY). It will be assigned in numerical order at the screening visit based on chronological order of screening dates.

Individuals who do not meet the criteria for participation in this trial (screen failure) may be rescreened once, if deemed acceptable by the investigator. Rescreened subjects should be assigned a different subject number than the initial screening. All procedures planned at the screening visit, including signature of a new consent form, will be performed.

In the case of a screening laboratory value abnormality, the test can be repeated once within the original screening time window, if the investigator believes there is a reasonable possibility that the subject would be eligible if re-tested. This would not be considered a screen failure and a new subject number would not be assigned.

The clinical site is responsible for maintaining a current log of subject ID number assigned to each subject. The subject ID number will be used to identify the subject throughout the study and is required to be entered on all clinical study documentation (eg, case report forms, labeling of clinical materials and sample containers, investigational product accountability logs, etc.).

4.5. Selection of Study Population

4.5.1. Inclusion Criteria

Subjects must fulfill all of the following inclusion criteria to be eligible for participation in the study:

1. For adult subjects: Participants legally competent to sign and give informed consent. For pediatric and adolescent subjects: Informed consent of parent(s) or legal guardian, and, if age appropriate, assent by the subjects, as required by local laws.
2. Males and females, ages 6 years and older at time of signing Informed Consent (Screening). Only subjects 18 years and older will be enrolled at sites located in the province of Québec in Canada.
3. Diagnosed with mild to moderate atopic dermatitis according to the criteria of Hanifin and Rajka (1980) prior to or at the screening visit. Subjects must have at least 3 of the 4 basic features per Hanifin and Rajka (1. Pruritus; 2. Typical morphology and distribution [flexural lichenification in adults and facial and extensor eruptions in infants and children]; 3. Chronic or chronically relapsing dermatitis; or 4. Personal or family history of atopy), in addition to 3 or more minor criteria.

4. History of AD for at least 3 months in subjects 6-17 years of age or 6 months in subjects ≥18 years of age, as determined by the Investigator using information from the subject's medical chart, from the subject's physician, or through subject/parent/caregiver interview. Stable disease for the past 4 weeks with no significant flares in atopic dermatitis before screening.
5. EASI Score ≥5 at Baseline. EASI is evaluated for the entire body except the scalp, palms, and soles.
6. vIGA-AD score of 'Mild' ('2') or 'Moderate' ('3') at Baseline/Day 1. The vIGA-AD is evaluated for the entire body except the scalp, palms, and soles.
7. Has AD involvement of ≥3% BSA (excluding the scalp, palms, soles) at Baseline.
8. Females of childbearing potential (FOCBP) must have a negative serum pregnancy test at Screening (Visit 1) and a negative urine pregnancy test at Baseline/Day 1. In addition, sexually active FOCBP must agree to use at least one form of a highly effective or barrier method of contraception throughout the trial. The use of abstinence as a contraceptive measure is acceptable as long as this is a consistent part of a lifestyle choice and an acceptable backup method has been identified if the subject becomes sexually active.
9. Females of non-childbearing potential should either be pre-menarchal, or post-menopausal with spontaneous amenorrhea for at least 12 months (post-menopausal status should be confirmed with FSH testing) or have undergone surgical sterilization (permanent sterilization methods include hysterectomy, bilateral oophorectomy, or bilateral salpingectomy).
10. In good health as judged by the Investigator, based on medical history, physical examination, vital signs, serum chemistry labs, hematology values, and urinalysis.
11. Subjects and parent(s)/legal guardian(s) are considered reliable and capable of adhering to the Protocol and visit schedule, according to the judgment of the Investigator.

4.5.2. Exclusion Criteria

Subjects who meet any of the following exclusion criteria will be excluded from participation in this study:

1. Subjects with any serious medical condition or clinically significant laboratory, vital signs, or physical examination abnormality that would prevent study participation or place the subject at significant risk, as judged by the Investigator
2. Liver function tests excursions that exceed:
   - AST or ALT >2X ULN
   - Total bilirubin:
     - >1.5 x ULN or
     - > ULN and ≤1.5 x ULN AND direct bilirubin is >35% of total bilirubin
   - ALP ≥2x ULN

3. Subjects who cannot discontinue medications and treatments prior to the Baseline visit and during the study according to Excluded Medications and Treatments (Table 2).
4. Has unstable AD or any consistent requirement for high potency topical steroids to manage AD signs or symptoms.
5. Subjects who have significant active systemic or localized infection (eg, molluscum contagiosum), including known actively infected AD, or have had any infection that required oral or intravenous administration of antibiotics, antifungal or antiviral agents within 14 days prior to Baseline/Day 1.
6. Subjects who are unwilling to refrain from prolonged sun exposure and from using a tanning bed or other artificial light emitting devices (LEDs) for 4 weeks prior to Baseline/Day 1 and during the study.
7. Subjects with skin conditions other than AD that would interfere with evaluations of the effect of the study medication on AD, as determined by the Investigator. Subjects with any condition on the treatment area which, in the opinion of the Investigator, could confound efficacy measurements, eg, molluscum contagiosum.
8. Subjects with known genetic dermatological conditions that overlap with AD, such as Netherton syndrome.
9. Known allergies to excipients in ARQ-151 cream [redacted]
10. Subjects who cannot discontinue the use of strong cytochrome P-450 CYP3A4 inhibitors eg, indinavir, nelfinavir, ritonavir, clarithromycin, itraconazole, ketoconazole, nefazodone, saquinavir, suboxone and telithromycin for 2 weeks prior to the Baseline/Day 1 and during the study period.
11. Subjects who cannot discontinue the use of strong cytochrome P-450 CYP3A4 inducers eg, efavirenz, nevirapine, glucocorticoids, barbiturates (including phenobarbital), phenytoin, rifampin, and carbamazepine for 2 weeks prior to the Baseline/Day 1 and during the study period.
12. Subjects who have received oral roflumilast (Daxas®, Daliresp®) within the past 4 weeks.
13. Known or suspected:
    - Severe renal insufficiency
      - Severe renal insufficiency is defined as calculated creatinine clearance <30 mL/min.
    - Moderate to severe hepatic disorders (Child-Pugh B or C)
14. History of severe depression, suicidal ideation or behavior, Baseline/Screening C-SSRS (for adolescents and adults 12 years old and older) indicative of suicidal ideation or behavior, whether lifetime or recent/current.

15. Subjects with a PHQ-8 (adults) or modified PHQ-A (adolescents, 12-17 years old inclusive) score ≥10 at Screening or Baseline visits
16. Subjects (6 to 11 years old, inclusive) with a CDI-2 (parent report) raw score ≥17 for females and ≥18 for males at Screening or Baseline/Day 1 visits
17. Females who are pregnant, wishing to become pregnant during the study, or are breast-feeding.
18. Previous treatment with ARQ-151.
19. Subjects currently undergoing allergy testing (eg, food allergy testing or skin prick testing), patch testing, food challenges, or allergy desensitization, or plan to do so during the study.
20. Subjects with any serious medical condition (eg, uncontrolled hypo- or hyper-thyroidism) or clinically significant laboratory abnormality that would prevent study participation or place the subject at significant risk, as determined by the Investigator.
21. Subjects with a history of a major surgery within 4 weeks prior to Baseline/Day 1 or subjects who have a major surgery planned during the study.
22. Subjects with a history of chronic alcohol or drug abuse within 6 months prior to Screening.
23. Current or a history of cancer within 5 years with the exception of fully treated skin basal cell carcinoma, cutaneous squamous cell carcinoma or carcinoma in situ of the cervix.
24. Parent(s)/legal guardian(s) who are unable to communicate, read, or understand the local language. Subjects who are unable to communicate, read or understand the local language(s), or who display another condition, which in the Investigator's opinion, makes them unsuitable for clinical study participation.
25. Subjects who are family members of the clinical study site, clinical study staff, or sponsor, or family members of enrolled subjects (subjects enrolled in other studies of ARQ-151) living in the same house.

4.6. Randomization

Randomization will take place at the Baseline visit after the Investigator confirms that the subject meets all eligibility criteria listed in Section 4.5.

Subjects will be randomly assigned to apply ARQ-151 cream 0.15% QD, or matching vehicle QD. Assignment of drug or vehicle will be made at a 2:1 ratio (drug:vehicle) and stratified by vIGA-AD score ('Mild' vs. 'Moderate'), and by study site according to a computer-generated randomization list. Kits containing tubes of study medication will be assigned to each subject using an internet-based response system (IWRS). A subject may receive more than one kit for the treatment period. The kits and tubes are blinded and each kit is numbered with a unique kit number.

4.7. Study Restrictions

4.7.1. Prohibitions and Concomitant Therapy

Prohibited medications and products are detailed in Table 2 (Excluded Medications and Treatments).

Generally, the addition of new medications, including nonprescription medications, during the course of the study is discouraged. However, the short-term use of a medication may be authorized by the Investigator. The Investigator must make the decision to authorize the use of any such a medication only after consideration of the clinical situation, the potential for masking symptoms of a more significant underlying event, and whether the use of the medication will compromise the outcome or validity of the clinical investigation. Other medications may be authorized by the Investigator for conditions other than AD. If medication is required, the name, strength, frequency, duration of use, and reason for use will be recorded in source documents and entered into the CRFs. Medications which have been used chronically by subjects, in particular statins and anti-hypertensives, are allowed for use during the study, except as prohibited in Table 2. No rescue medication for AD is allowed during this study up to Week 4.

Table 2: Excluded Medications and Treatments

Excluded Medications and Treatments	Washout Period Prior to Day 1
Approved biologics such as dupilumab	6 months
Investigational biologics	6 months
• Systemic treatments that could affect AD; eg, corticosteroids, retinoids, calcineurin inhibitors, hydroxycarbamide (hydroxyurea), methotrexate, cyclosporine, azathioprine, hydroxychloroquine, mycophenolate mofetil, or other immunosuppressive therapies, or systemic treatment with nonsedating antihistamines in a nonstable regimen. • Systemic treatments with nonsedating antihistamines (eg, cetirizine, desloratadine, loratadine) in a stable regimen is allowed.	4 weeks or 5 half-lives, whichever is longer
PUVA or NBUVB phototherapy	4 weeks
Topical products containing urea	1 week
Sedating antihistamines and other over the counter remedies containing sedating antihistamine, such as sleep aids (eg, ZzzQuil™ LIQUICAPS® SLEEP-AID), and cough/cold remedies (eg, Theraflu® night time, NyQuil™ Cold & Flu Night time)	1 week
Topical corticosteroids, calcineurin inhibitors, or Eucrisa®. Topical antibacterial medications or products, including soaps, dilute bleach baths, or sodium hypochlorite-based products anywhere on the body.	2 weeks
Strong cytochrome P-450 CYP3A4 inhibitors eg, indinavir, nelfinavir, ritonavir, clarithromycin, itraconazole, ketoconazole, nefazodone, saquinavir, suboxone and telithromycin	2 weeks

Table 2: Excluded Medications and Treatments (Continued)

Excluded Medications and Treatments	Washout Period Prior to Day 1
Strong cytochrome P-450 CYP3A4 inducers eg, efavirenz, nevirapine, glucocorticoids, barbiturates (including phenobarbital), phenytoin, rifampin, and carbamazepine	2 weeks
Systemic antibiotics	2 weeks
Tanning beds, other light emitting devices	4 weeks
Oral roflumilast (Daxas®, Daliresp®)	4 weeks
All other investigational drugs	4 weeks or 5 half-lives, whichever is longer

- Eye / ear drop and nasal corticosteroid preparations are allowed. Inhaled corticosteroid preparations are allowed if used for a stable condition and at a stable dose for >28 days before screening and are continued at the same dose throughout the study.
- All subjects should apply medication each evening, except on clinic visit days when the study product will be applied at the clinical site. If the subject takes an evening shower/bath, the ARQ-151 cream or vehicle can be applied as soon as the skin is nearly dry, but no later than 20 minutes before going to bed. Subjects are not to wash areas where ARQ-151 cream or vehicle has been applied until at least 4 hours after study drug application. Non-medicated emollients or moisturizers will be allowed once daily in a stable regimen as normally used by the subjects. For subjects that apply a non-medicated emollient or moisturizer after an evening shower/bath, the study drug must be applied first to the treatment areas. The non-medicated emollient or moisturizer can then be applied but only to other untreated areas of the subject's skin.
- Sunscreens will be allowed daily, as needed by the subjects when applied at least 2 hours after application of randomized study drug.
- Concomitant other medications for chronic conditions (eg, NSAIDs, statins, anti-hypertensives) are permitted unless specifically prohibited in the Protocol.
- Topical antibiotics, topical antihistamines, or any other topical agents are not allowed to be applied to treated areas.

4.8. Treatment

4.8.1. Drug Supplies, Packaging and Labeling

ARQ-151 cream or matching vehicle will be supplied in 45 gram tubes. The tubes will be packaged in kits, containing multiple tubes of investigational product. The number of kits dispensed to a subject will be based on the BSA involvement of atopic dermatitis. The kits and tubes will be labeled in a blinded manner. The kit(s) dispensed to a subject will be labeled with a unique number.

The Sponsor will supply sufficient quantities of the study drug and matching vehicle to each site to allow for completion of this study.

Records will be made of the receipt and dispensing of the study drugs supplied. At the conclusion of the study, any used/unused study drugs will be returned to the Sponsor or designee, or destroyed, as per Sponsor instructions.

Refer to the most current version of the IP Handling Plan for details on accountability, storage, and management of ARQ-151.

4.8.2. Blinding

This is a double-blind study, therefore neither the subjects nor the Investigator and clinical personnel will be aware of which treatment an individual subject receives.

4.8.3. Breaking Treatment Codes

In the event of a medical emergency where breaking the blind is required to provide medical care to the subject, the Investigator may obtain treatment assignment directly from the IWRS system for that subject. Refer to the current version of the ARQ-151-311 IWRS User Manual for details on unblinding. Treatment assignment should, however, remain blinded unless the assignment knowledge is necessary to determine subject emergency medical care. The rationale for unblinding must be clearly explained in source documentation and on the CRF, along with the date on which the treatment assignment was obtained. The investigator is requested to contact the Medical Monitor promptly in the event of any treatment unblinding.

Blinding of study treatment is critical to the integrity of this clinical trial and therefore, if a subject's treatment assignment is disclosed to the Investigator, the subject will have the study treatment discontinued. All subjects will be followed until study completion unless consent to do so is specifically withdrawn by the subject.

4.8.4. Treatment Administration

Initial treatment with the IP will occur on Day 1. ARQ-151 cream 0.15% is administered once daily as a topical product to cover the skin surface at an application rate of approximately 2 mg/cm^2.

At Baseline visit, the study staff will demonstrate to the subject/caregiver(s) how to apply ARQ-151 cream using the first tube from the kit that is assigned to the subject. Study site staff will be trained to ensure a unit dose (a pea size unit of ARQ-151 cream will cover approximately 1% BSA) is properly squeezed from the tube and applied to atopic dermatitis lesion(s) as a thin film and rubbed in using the index and middle finger, rubbing in thoroughly but gently, until the 'white' has disappeared. The subject/caregiver will then practice squeezing a similar amount onto their index and middle finger and apply a thin film to other areas to be treated. At Baseline/Day 1, the study staff will ensure that the subject/caregiver's application technique is correct and that a thin layer is applied as instructed (which represents an application rate of approximately 2 mg/cm^2).

Re-training will be conducted at subsequent visits as needed (i.e., if the returned tube(s) weighs substantially different than the expected weight).

Subjects/caregivers will be instructed to apply investigational product once daily to all treatment areas identified by the Investigator at Baseline using a Body Diagram (see Appendix 1).

Note:

- All subjects should apply medication each evening (except on clinic visit days when the investigational product will be applied at the clinic). If the subject takes an evening shower/bath, the ARQ-151 cream or vehicle can be applied as soon as the skin is nearly dry, but no later than 20 minutes before going to bed. Subjects are not to wash areas where ARQ-151 cream or vehicle has been applied until at least 4 hours after study drug application.

Caregivers should wash their hands with soap and water after applying IP to a child. Parents/guardians/caregivers who are pregnant, or women of childbearing potential who are trying to become pregnant, or who are breastfeeding, or planning to breastfeed during the study should avoid accidental exposure by either avoiding applying investigational product or by wearing gloves during its application.

- Subjects should maintain treatment of areas with study drug for the duration of the study regardless of whether treatable areas of AD clear prior to Week 4 visit.
- New lesions that develop during the study should be treated (except scalp). An unscheduled visit is not required for starting treatment of new lesions.

Each investigational product tube will be weighed prior to dispensing at the Baseline visit and at each subsequent visit. Investigational product tubes must be returned by subjects at each study visit, both empty and full, and will be weighed. If the subject's actual use is substantially different than the expected use for the subject's BSA (see IP Handling Plan), the subject/caregiver will be retrained on the study drug application technique.

4.8.5. Treatment Compliance

Investigational product tubes will be weighed at each clinic visit.

Subjects/caregivers will complete a daily diary recording the date and time of each dose applied, any missed doses, and a comment section should the subjects have a comment, eg, record potential AEs. Site personnel will review the diaries and use the information to question the subject regarding compliance and AEs and then record appropriate information in source documents and complete Case Report Forms (CRFs). If a subject misses a dose, they should be instructed to return to the protocol investigational product administration schedule (i.e. if subject forgets a dose they should wait until that evening and apply as usual).

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

Compliance will be assessed by review of the dosing diary. Weight of investigational product applied (via dispensed and returned tube weights) will be measured for reporting purposes.

If the diary shows less than 80% of expected daily applications (but not more than 3 consecutive missed doses), the subject is using too little study drug and retraining must be conducted and documented.

Compliance will be documented in source and in eCRF.

4.8.6. Removal of Subjects from Study Treatment

Subject treatment with study drug in this trial may be discontinued for any of the following reasons:

1. Occurrence of any medical condition or circumstance that, in the opinion of the Investigator does not allow the subject to adhere to the requirements of the Protocol.
2. Adverse Events as described in Section 5.1.12 and Section 5.9. The Investigator must follow the subject until the AE resolves or satisfactorily stabilizes.
3. Treatment must be discontinued immediately in the event of a female subject's pregnancy.
4. Subject's decision to discontinue treatment with study drug.
5. C-SSRS (Section 5.1.10) indicative of suicidal ideation.
6. PHQ-8 (Section 5.1.7) or modified PHQ-A (Section 5.1.8) score ≥15 if determined by Investigator in consultation with mental health professional.
7. CDI-2 (Section 5.1.9) raw total score of ≥32 if determined by Investigator in consultation with mental health professional.
8. Requirement for use of prohibited concomitant medication after consultation with the Sponsor and Medical Monitor.
9. Subject's repeated failure to comply with protocol requirements or study related procedures.
10. The subject interrupts trial study drug application for more than 50% of scheduled doses.
11. Termination of the study by the Sponsor, FDA, or other regulatory authorities.

4.8.7. Removal of Subjects from the Study

A subject may be removed from study participation for any of the following reasons:

- Subject death.
- Subject's decision to withdraw from the study.
- Subject is lost to follow-up. A subject will be considered lost to follow-up after three phone and three email attempts and documentation of a certified letter sent to the subject's address.
- Termination of the study by the Sponsor, FDA, or other regulatory authorities.

5. STUDY PROCEDURES

The Schedule of Visits and Assessments (Section 1.3) summarizes the clinical procedures to be performed at each visit. Individual clinical procedures are described in detail below.

5.1. Safety Assessments

This study assesses the safety and efficacy of ARQ-151 cream. Safety will be determined by evaluating physical examinations, local tolerability assessments, vital signs/weight, clinical laboratory parameters, either PHQ-8 (adults, ≥18 years old) or modified PHQ-A (adolescents, 12-17 years old) or Children's Depression Inventory 2 (CDI-2, parent report for children 6-11 years old, inclusive), C-SSRS (12 years and older) and AEs as outlined in the Schedule of Visits and Assessments (Section 1.3).

Additional evaluations/testing may be deemed necessary by the PI and/or the Sponsor for reasons related to subject safety.

5.1.1. Screening

Before a subject's participation in the clinical study, the Investigator is responsible for obtaining written informed consent from the subject or written assent from adolescent subjects and consent from their parent(s) or legal guardian(s) for children and/or adolescents after adequate explanation of the study design, anticipated benefits, and the potential risks. A subject is considered a participant of the trial once the ICF or written assent for adolescent subjects is completely signed.

Subjects must provide informed consent/assent as per their age group at screening. During the study, if a subject changes age group, the subject must provide informed consent/assent relative to his/her current age group. Subjects will continue with the assessments specific to their age group at the time of consent/assent at Screening.

The following procedures/assessments will be performed at the Screening Visit (within 4 weeks after signing the informed consent):

- Review of medical and surgical history
- Collection of demographic data including sex, age, race, ethnicity
- Vital signs including temperature, heart rate, and blood pressure
- Collection of body weight (kg), and height (cm)
- Atopic dermatitis assessments (eg, vIGA-AD, BSA, EASI, SCORAD)
- Limited physical examination of skin (including assessment of Fitzpatrick skin type at Screening only), lungs, and heart
  - Fitzpatrick skin phototype will be rated as follows:
    - I: Always burns easily; never tans (sensitive)
    - II: Always burns easily; tans minimally (sensitive)

III: Burns moderately; tans gradually (light brown) (normal)

IV: Burns minimally; always tans well (moderate brown) (normal)

V: Rarely burns; tans profusely (dark brown) (insensitive)

VI: Never burns, deeply pigmented (insensitive)

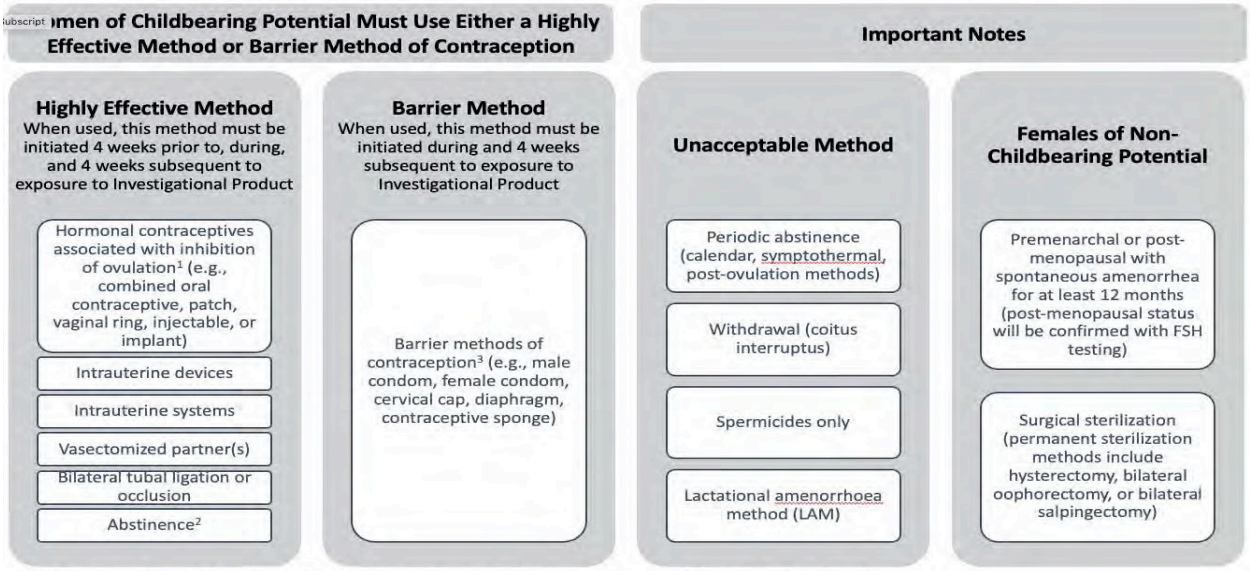
- Laboratory tests: hematology, chemistry, urinalysis, serum pregnancy test (for female subjects of child bearing potential)
- Completion of WI-NRS, CDI-2, DLQI, CDLQI, DFI, C-SSRS, POEM, and PHQ (-8 or -A)
- Collection of concomitant medications and adverse events

Subjects may be re-screened one time and the subject will be assigned a new Subject ID.

5.1.2. Contraception Requirements

Females of childbearing potential (FOCBP) must have a negative serum pregnancy test at Screening and a negative urine pregnancy test at Baseline (Day 1). In addition, sexually active FOCBP must agree to use at least one form of a highly effective or barrier method of contraception throughout the trial according to Contraception Requirements (Figure 1).

Figure 1: Contraception Requirements for Female Subjects



[1]Subjects using hormonal contraceptives must have been on a stable dose for at least 4 weeks before Day 1.
[2]The above list of contraceptive methods does not apply to subjects who are abstinent for at least 4 weeks before Day 1 and will continue to be abstinent from penile-vaginal intercourse throughout the study. The reliability of sexual abstinence needs to be evaluated in relation to the duration of the clinical trial and the preferred and usual lifestyle of the participant.
[3]Female condom and male condom should not be used together.

5.1.3. Baseline (Day 1)

Randomization will take place at the Baseline visit after the subject has been found to be fully eligible for participation. A subject will be considered enrolled into the study upon the first IP application.

If the Baseline visit occurs within 21 days of Screening for subjects 12 to 18 years old, the Screening laboratory test results may be used.

5.1.4. Physical Examination

Physical examinations will be performed according to the Schedule of Visits and Assessments (Section 1.3). The physical exam will be limited to skin, lungs and heart only.

5.1.5. Vital Signs, Height and Weight

Vital signs will be performed according to the Schedule of Visits and Assessments (Section 1.3). Blood pressure, heart rate, and temperature will be collected in seated position after 5 mins of rest. For weight measurement, subjects will be instructed to void prior to weight being taken and to remove any objects of significant weight (eg, jackets, outerwear, shoes, cell phones, wallet, key chains, etc.). Weight should be obtained using a calibrated weight scale and the same scale, whenever possible, should be used for a subject throughout the duration of the study. 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 (eg, 55.5 pounds or 25.1 kilograms). For subjects <18 years of age, measure the weight in triplicate and report the average weight in EDC. An unexplained, clinically significant weight loss should be reported to the Medical Monitor.

Height will be measured at Screening only.

5.1.6. Laboratory Tests

All tests listed in Table 3 below will be performed according to the Schedule of Visits and Assessments (Section 1.3) unless otherwise noted. Hematology, serum chemistries, and urine analysis will be collected for all subjects at Screening, but subsequent samples will be collected only for subjects ≥12 years old (Baseline/Day 1 and Week 4/Day 29/ET). No food restrictions are required for the collection of specimens. In addition, laboratory safety tests may be performed at various unscheduled time points, if deemed necessary by the Investigator.

Table 3: Laboratory Tests

Hematology	**Serum Chemistry**
Hemoglobin Hematocrit Total and differential leukocyte count Red blood cell count with indices and morphology Platelet count	Blood Urea Nitrogen Bilirubin (total and direct) Alkaline phosphatase Aspartate aminotransferase Alanine aminotransferase Albumin Sodium Potassium Chloride Glucose Creatinine
Urinalysis	**Additional Tests**
pH Specific gravity Protein* Glucose Ketones Bilirubin Blood* Nitrite* Urobilinogen Leukocyte esterase*	Urine pregnancy test** (for females of child bearing potential only) Serum pregnancy test (hCG)*** FSH test, (post menopausal) ***

* If urinalysis is positive for protein, blood, nitrite and/or leukocyte esterase, a microscopic examination (for red blood cells, white blood cells, bacteria, casts, and epithelial cells) will be performed.

** At Baseline and Weeks 1, 2, and 4 for FOCBP only

*** At screening only

5.1.7. Patient Health Questionnaire Depression Scale (PHQ-8)

The PHQ-8 Assessment (Appendix 2) will be performed in adult subjects according to the Schedule of Visits and Assessments (Section 1.3).

PHQ-8 score is the sum of the responses for the 8 questions.

Five severity categories of depression are defined 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)

A subject with a PHQ-8 score of 10-14 should be referred to a mental health professional for evaluation, regardless of the reason for their change in score.

A subject with a PHQ-8 score ≥15 should be immediately referred to a mental health professional, and Investigators in consultation with the mental health professional should consider interruption or discontinuation of the study drug.

5.1.8. Patient Health Questionnaire Depression Scale (Modified PHQ-A)

The Modified PHQ-A Assessment (Appendix 3) will be performed in adolescent subjects (12-17 years old, inclusive).

Modified PHQ-A score is the sum of the responses for five severity categories of depression defined 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)

A subject with a modified PHQ-A score of 10-14 should be referred to a mental health professional for evaluation, regardless of the reason for their change in score.

A subject with a modified PHQ-A score ≥15 should be immediately referred to a mental health professional, and Investigators in consultation with the mental health professional should consider interruption or discontinuation of the study drug.

5.1.9. Children's Depression Inventory 2 (CDI-2)

The CDI-2 Assessment will be performed according to the Schedule of Visits and Assessments (Section 1.3) for subjects 6 to 11 years old, inclusive.

The CDI-2 quantifies depressive symptomatology using reports from children/adolescents, teachers, and parents or caregivers. It is recommended for use in initial evaluation and is appropriate when there is a need for an assessment and robust description of a child's depressive symptoms.

This study will use the CDI Parent Report Form. An example of the Parent report form is presented in Appendix 4.

A subject with a CDI-2 raw score of ≥21 for females and ≥22 for males should be referred to a mental health professional for evaluation.

A subject with a CDI-2 raw total score of ≥32 should be referred to a mental health professional, and Investigators in consultation with the mental health professional should consider interruption or discontinuation of the study drug.

5.1.10. Columbia-Suicide Severity Rating Scale (C-SSRS)

C-SSRS Assessments will be performed according to the Schedule of Visits and Assessments (Section 1.3) for subjects 12-years old and older.

The administration schedule of the C-SSRS will be:

- The "Baseline/Screening" version (Appendix 5) will be used at Screening to provide a pre-treatment assessment.
- On all subsequent visits, the "Since Last Visit" version (Appendix 6) will be used (Baseline/Day 1, Week 1/Day 8, Week 2/Day 15 or Week 4/Day 29).
- A score greater than 0 at the Screening or Baseline visit in suicidal ideation may indicate the need for mental health intervention. The investigator should not enroll the subject in the study.
- Any score greater than 0 in the suicidal ideation score may indicate the need for mental health intervention. The Investigator should give consideration for the subject to discontinue from the study drug and prompt referral to an identified mental health professional and/or an appropriate emergency room. The Medical Monitor should be contacted.

The C-SSRS administer will be trained via the C-SSRS training video. A training certificate for the administer(s) will be on file in the trial master file at the site.

The Investigator must review the completed C-SSRS. A qualified mental health care provider must be available, immediately if needed, to refer the subject for further evaluation.

5.1.11. Local Tolerability Assessment

The Investigator Local Tolerability Assessment will be performed according to the Schedule of Visits and Assessments (Section 1.3).

Application site reactions will be graded at each timepoint. Irritation reactions are graded using the scale detailed in the following section (Berger 1982).

Reactions at the site of product application, which may occur post-Baseline, should be differentiated from the preexisting inflammation associated with the subject's atopic dermatitis.

The investigator assessments will be conducted by the investigator or a properly trained and designated subinvestigator prior to study drug application in the clinic.

Dermal Response

0. no evidence of irritation
1. minimal erythema, barely perceptible
2. definite erythema, readily visible; minimal edema or minimal papular response
3. erythema and papules
4. definite edema

5. erythema, edema and papules
6. vesicular eruption
7. strong reaction spreading beyond application site

Other Effects

A. = slight glazed appearance
B. = marked glazing
C. = glazing with peeling and cracking
D. = glazing with fissures
E. = film of dried serous exudates
F. = small petechial erosions and/or scabs
G. = no other effects

The Subject Local Tolerability Assessment will be performed according to the Schedule of Visits and Assessments (Section 1.3).

The subject will assess burning/stinging (0-3 score):

Grade	**Sensation Following Drug Application**
0 (none)	No sensation
1 (mild)	Slight warm, tingling sensation; not really bothersome
2 (moderate)	Definite warm, tingling sensation that is somewhat bothersome
3 (severe)	Hot, tingling/stinging sensation that has caused definite discomfort

This assessment will be administered by the site 10 to 15 minutes after study drug application in the clinic at Baseline and at every clinic visit.

- Note: for subject burning stinging assessment at Day 29, subjects will provide a recall assessment of burning/stinging experienced post drug application on the previous day (Day 28).

5.1.12. Adverse Events

Adverse events (AEs) will be collected and assessed throughout the study according to the Schedule of Visits and Assessments (Section 1.3). The Investigator is responsible for ensuring that all adverse events observed by the clinical staff or reported by the subject that occur after the first application of investigational product through the end of the study are recorded in the subject's medical record and the eCRF.

The Investigator is responsible for ensuring that all serious adverse events observed by the clinical staff or reported by the subject that occur after signing of the informed consent through 30 days after the last day of the application of the investigational product or the end of study (whichever is later) are recorded in the subject's medical record and are submitted per SAE reporting requirements (Section 5.7.5).

AEs (whether serious or non-serious) and clinically significant abnormal laboratory test value(s) will be evaluated by the PI and treated and/or followed up for up to 30 days after end of treatment or until the symptoms or value(s) return to normal, or acceptable levels, as judged by the PI. Refer Section 5.7 for further details on Adverse Events.

5.2. Efficacy Evaluations

For efficacy evaluation subjects will have ≥3% BSA of AD involvement (excluding the scalp, palms, soles). Palms and soles may be treated with investigational product in this study, but will not be counted towards vIGA-AD, EASI, or BSA assessments. EASI is evaluated for the entire body except the scalp, palms, and soles.

5.2.1. Validated Investigator Global Assessment Scale for Atopic Dermatitis

Validated Investigator Global Assessment scale for Atopic Dermatitis (vIGA-AD) assessments should be completed prior to other physician assessments.

vIGA-AD assessment will be performed according to the Schedule of Visits and Assessments (Section 1.3). The vIGA-AD 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). Each grade is defined by a distinct and clinically relevant morphologic description that minimizes inter-observer variability (see Appendix 7). vIGA-AD is evaluated for the entire body except the scalp, palms, and soles.

Note: All atopic dermatitis lesions on a subject will be treated including the face, trunk, genitals/skin folds, or limbs (excluding the scalp). The palms and soles will be treated but will not be counted towards any measurements of efficacy (EASI, vIGA-AD, BSA).

Every effort must be made for the same Evaluator to complete the IGA for the subject at every study visit.

IGA will be assessed at clinic visits prior to the subject applying Investigational Product at the site.

5.2.2. Eczema Area and Severity Index (EASI)

EASI scores (Hanifin 2001) will be performed according to the Schedule of Visits and Assessments (Section 1.3)

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; eg, 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:

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

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

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

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

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.

5.2.3. Worst Itch Numerical Rating Scale (WI-NRS)

The WI-NRS has been developed as a simple, single item to assess the patient-reported severity of this symptom at its highest intensity during the previous 24-hour period. (Newton 2019). The WI-NRS will be determined by the subject's recording of daily assessment of worst itch over the past 24 hours. The scale is from '0 to 10' ("no itch" to "worst itch imaginable" or "worst imaginable itch").

Date (DD/MMM/YYYY): _____/_____/_____ Time (HH:MM): _____:_____ ☐ AM ☐ PM

Please rate your itching severity by circling the number that best describes your worst level of itching in the past 24 hours:

0 1 2 3 4 5 6 7 8 9 10

0 = No itch **10 = Worst itch imaginable**

The WI-NRS will be **self-reported by the subjects.** Subjects will be reminded not to review responses from the previous assessment when completing the WI-NRS. Subjects will be trained at the Screening visit in the accurate completion of the WI-NRS. In addition, parents/caregivers of children and adolescent subjects will be trained at the Screening visit by study staff on how to assist the subject, if needed, in completing the WI-NRS. If subjects, despite the assistance of a parent/caregiver, are unable to complete the WI-NRS no entry should be made in the subject diary.

WI-NRS Assessments will be performed according to the Schedule of Visits and Assessments (Section 1.3) starting 7 days prior to the Baseline/Day 1 clinic visit (during the 7 days prior to Baseline/Day 1 the subject will record the WI-NRS value every day) and until Week 4/ET.

5.2.4. Dermatology Life Quality Index (DLQI) and Children's Dermatology Life Quality Index (CDLQI)

The DLQI (ages 17+ years) and CDLQI (ages 6-16 years, inclusive) will be performed according to the Schedule of Visits and Assessments (Section 1.3). The DLQI/CDLQI is a simple, self-administered and user-friendly validated questionnaire. The DLQI/CDLQI is designed to measure the health-related quality of life of adult 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. Subjects/caregivers will complete the CDLQI/DLQI. Refer to Appendix 8 for the DLQI and Appendix 9 for the CDLQI.

5.2.5. Dermatitis Family Impact Questionnaire (DFI)

This questionnaire measures how much having a child with atopic dermatitis affects the quality of life of other (adult) members of the family. To be completed by parents/guardians/caregivers of subjects ≤17 years of age (Appendix 10).

5.3. Other Evaluations

5.3.1. Body Surface Area (BSA)

BSA assessments will be performed according to the Schedule of Visits and Assessments (Section 1.3).

The BSA affected for atopic dermatitis will be determined by the subject's hand method, where the subject's hand (including fingers) surface area is assumed to equal 1% of body surface area (excluding the scalp, palms, soles).

5.3.2. SCORAD ("SCORing Atopic Dermatitis")

SCORAD is a clinical tool for assessing the severity (i.e. extent, intensity) of atopic dermatitis as objectively as possible. It gives approximate weights of 60% to intensity and 20% each to spread (extent) and subjective signs (insomnia, etc.).

See Appendix 11.

5.3.3. Patient-Oriented Eczema Measure (POEM)

The Patient-Oriented Eczema Measure (POEM) is a tool used for monitoring atopic eczema severity. It focuses on the illness as experienced by the patient.

POEM is a 5-point scale measuring the frequency of each of seven AD symptoms (dryness, itching, flaking, cracking, sleep loss, bleeding, and weeping) over the past week scored as occurring "no days" (0), "1 to 2 days" (1), "3 to 4 days" (2), "5 to 6 days" (3) or "every day" (4). Total score ranges from 0–28, with higher score indicating greater symptom impact.
See Appendix 12. The self/proxy report questionnaire will be used in this study (for children unable to read and/or understand the POEM questionnaire, the parent/guardian/caregiver will complete the questionnaire).

5.3.4. Pharmacokinetics Assessment

PK draws will be performed according to the Schedule of Visits and Assessments (Section 1.3) for all subjects at all sites under Amendment 2:

- A single PK assessment (trough) will be performed in subjects ≥12 years old with a blood sample collected at Week 4/Day 29/ET.
- No PK sample will be collected Week 4/Day 29 for subjects 6-11 years old.

Ensure study medication is not applied in the area where PK will be drawn.

5.3.5. Medical Photography

Photography of AD lesion(s) selected by the Investigator will be performed by all sites at all investigational visits, except Week 2/Day 15. All efforts will be made to de-identify the subjects. Canfield equipment will be used to capture photographs.

Photography should be focused on single lesions or specific body sections (eg, arm). Body or half body photos should only be taken if necessary. 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.

Refer to the current Photography Manual for instructions regarding photography.

5.4. Final Study Visit – End of Study

The approximate final study visit will occur at Week 4/Day 29. The procedures performed during this visit are as described in the Schedule of Visits and Assessments (Section 1.3). A 3-day scheduling window is allowed for this visit. Adverse events will be recorded as reported by the participant or and followed to resolution or stabilization (as necessary).

5.5. Early Termination Visit

If a subject is withdrawn or wishes to exit the study, a termination visit will be scheduled. This visit should include the procedures and assessments that would be performed at the Week 4/ Day 29 visit.

5.6. Unscheduled Visit

Unscheduled visits may be necessary to repeat testing following abnormal laboratory results, for follow-up of AEs, or for any other reason, as warranted in the judgement of the Investigator.

The following information will be collected for all subjects:

- Concomitant medications/procedures
- AEs

The following information also will be collected:

- vIGA-AD and EASI
- BSA affected with AD
- Local tolerability assessment (by Investigator)

However, if an unscheduled visit is required for reasons other than safety (eg, procedures such as labs or images that were either missed at the regular subject visit or need to be repeated), the vIGA-AD and EASI, BSA affected with AD, and Local Tolerability assessment (by Investigator) are not required.

The rules for how to tally vIGA-AD, BSA or other proportions of categorical responses will be described in the Statistical Analysis Plan.

5.7. Adverse Events

5.7.1. Adverse Event Definition

An AE is defined as any untoward medical occurrence associated with the use of a drug in humans, whether or not considered drug related. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not related to the medicinal (investigational) product.

A treatment emergent adverse event (TEAE) is defined as an AE that started post application of study medication at the Baseline visit or was present at treatment initiation but worsened during treatment, through study completion.

The definition of AEs includes worsening of a pre-existing medical condition. Worsening indicates that the pre-existing medical condition has increased in severity, frequency, and/or duration or has an association with a worse outcome. When recording such events, provide descriptions that the pre-existing condition has changed (eg, worsening hypertension for a subject with pre-existing hypertension). A pre-existing condition that has not worsened during the study or involves an intervention, such as elective cosmetic surgery or a medical procedure while on study, is not considered an AE.

Progression of atopic dermatitis including new or worsening of anticipated clinical signs or symptoms, which are collected as clinical efficacy variables and assessed as unequivocally associated with the disease progression and /or lack of efficacy, should NOT be reported as adverse events unless the disease progression is greater than anticipated in the natural course of the disease.

In general, abnormal laboratory findings without clinical significance (based on the investigator's judgment) are not to be recorded as AEs. However, abnormal laboratory findings that result in new or worsening clinical sequelae, or that require therapy or adjustment in current therapy, are considered AEs. Where applicable, clinical sequelae (not the laboratory abnormality) are to be recorded as the AE.

5.7.2. Serious Adverse Event Definition

The definitions and reporting requirements of the Food and Drug Administration (FDA)/ICH Guidelines for Clinical Safety Data Management, Definitions and Standards for Expedited Reporting, Topic E2A will be adhered to. Statements regarding mandatory reporting of all serious unexpected adverse drug reactions (SUSARs) to Health Canada [as per C.05.014 (1) of the FDR] will be adhered. If any AEs are serious, as defined by ICH Guidelines for Clinical Safety, required procedures will be followed.

An SAE is defined as any AE that, in the view of either the PI or Sponsor, meets at least 1 of the following serious criteria:

- Fatal
- Life-threatening (places the subject at immediate risk of death)

- Requires inpatient hospitalization or prolongation of existing hospitalization
- Results in a persistent or significant incapacity/disability
- Congenital anomaly/birth defect.
- Other important medical events that, based upon appropriate medical judgment, may jeopardize the subject and may require medical or surgical intervention to prevent one of the outcomes listed in the above definition.

Hospitalization does not include the following:

- Rehabilitation facilities, hospice facilities or respite care (eg, caregiver relief)
- Nursing homes or skilled nursing facilities
- Emergency room visits
- Same day surgeries (as outpatient/same day/ambulatory procedures)
- <24 hour admissions for observation or evaluation

Hospitalization or prolongation of hospitalization in the absence of a precipitating, clinical AE is not in itself an SAE. Examples include:

- Admission for treatment of a preexisting condition that did not worsen
- Hospitalizations for cosmetic elective surgery, social, and/or convenience admissions
- Pre-planned treatments or surgical procedures should be noted in the baseline documentation for the individual subject.
- Diagnostic and therapeutic procedures, such as surgery, should not be reported as AEs; however, the medical condition for which the procedure was performed should be reported if it occurs during the reporting period and meets the definition of an AE. For example, an acute appendicitis that begins during the AE reporting period should be reported as an AE, and the resulting appendectomy should be recorded as treatment of the AE.

Unexpected is defined as an AE that is not listed in the IB or is not listed at the specificity or severity that has been observed; or is not consistent with the risk information described in the general investigational plan or study document.

If a SAE occurs to a subject on this study, contact the Medical Monitor within one business day of knowledge of event.

5.7.3. Suspected Unexpected Serious Adverse Reaction (SUSAR)

SUSAR is defined as a serious adverse reaction, the nature or severity of which is not consistent with the known study treatment information. A serious event or reaction is not defined as a SUSAR when: ‘it is serious but expected’ or it does not fit the definition of an SAE, whether expected or not.

5.7.4. Safety Review with Subject

At each subsequent clinic visit after the screening visit, subjects will be queried with an open-ended question such as: 'How have you been feeling since your last visit?' Additionally, the study staff will review subject diaries and, if it appears that a potential AE was recorded, study staff will query the subject and determine if an AE occurred.

AEs (whether serious or non-serious) and clinically significant abnormal laboratory test value(s) will be evaluated by the PI and treated and/or followed for up to one month after end of treatment until the symptoms or value(s) return to normal, or acceptable levels, as judged by the Investigator.

Where appropriate, medical test(s) and/or examination(s) will be performed to document resolution of event(s). Outcome may be classified as resolved, improved, unchanged, worse, fatal or unknown (lost to follow-up).

5.7.5. Adverse Event Reporting

The Investigator is responsible for recording all adverse events, observed by the clinic staff or reported by the subject that occur after the first application of investigational product through the end of the study. All SAEs should be reported starting after the signing of the informed consent through 30 days after the last day of the application of the investigational product or the end of the study (whichever is later).

Any AEs (whether serious or non-serious) and clinically abnormal laboratory test value(s) will be evaluated by the PI and treated and/or followed up for up to 30 days after end of treatment or until symptoms or value(s) return to normal, or acceptable level, as judged by the PI (if the subject is continuing into the ARQ-151-313 OLE study, then AEs from this study (ARQ-151-311) will only be followed until they exit from this study).

All adverse events that meet the criteria for "serious" (i.e., SAEs) will be reported to the Sponsor via fax or e-mail within 24 hours of becoming aware of the event, whether or not the serious events are deemed drug-related. Reporting should be done by sending the completed SAE form to the following e-mail address (faxing can also be done as a second option in case e-mailing is not possible).

Safety Contact Information: [redacted]
E-mail: [redacted]

All serious event reporting will adhere to ICH E6: Guideline for Good Clinical Practice and ICH E2A: Clinical Safety Data Management: Definitions and Standards for Expedited Reporting. The IRB will be notified of the Alert Reports as per FDA, ICH and the IRB's policies and procedures. The sponsor, or delegate, will be responsible for reporting SAEs to health authorities per local reporting requirements. The Investigator will be responsible for reporting events to their respective IRBs in accordance to the IRB requirements.

The Investigator will review each adverse event and assess its relationship to Investigational Product (unrelated, unlikely, possibly, probably, likely). Each sign or symptom reported will be graded on the NIH NCI CTCAE toxicity grading scale 5-point severity scale (Grade 1, 2, 3, 4

and 5). The date and time of onset, time relationship to drug dosing, duration, and outcome (resolved, improved, unchanged, worse, fatal, or unknown/lost to follow-up) of each event will be noted.

The relationship of each AE to the Investigational Product will be assessed using the following definitions:

Unrelated	The AE must clearly be caused by the subject's clinical state, or the study procedure/conditions. Definitely not related to drug. Temporal sequence of an AE onset relative to administration of drug not reasonable. Another obvious cause of an AE.
Unlikely	Time sequence is unreasonable. There is another more likely cause for an AE.
Possibly	Corresponds to what is known about the drug. Time sequence is reasonable. Could have been due to another equally, likely cause.
Probably	Is a known effect of the drug. Time sequence from taking drug is reasonable. Ceases on stopping the drug. Cannot be reasonably explained by the known characteristics of the subject's clinical state.
Likely	Is a known effect of the drug (eg, listed in Physicians' Desk Reference, Compendium of Pharmaceuticals and Specialties, IB). Time sequence from taking drug is reasonable. Event stops upon stopping drug, event returns upon restarting drug.

The following CTCAE toxicity grading scale 5-point severity scale definitions for rating maximum severity will be used:

Grade 1	Mild; asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated.
Grade 2	Moderate; minimal, local or noninvasive intervention indicated; limiting age-appropriate instrumental activities of daily living.*
Grade 3	Severe or medically significant but not immediately life-threatening; Hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care activities of daily living.** Note**:** An experience may be severe but may not be serious, eg, severe headache).
Grade 4	Life-threatening consequences; urgent intervention indicated.
Grade 5	Death related to AE.

Note: A semi-colon indicates 'or' within the description of the grade.

* Instrumental activities of daily living refer to preparing meals, shopping for groceries or clothes, using the telephone, managing money, etc.

** Self-care activities of daily living refer to bathing, dressing and undressing, feeding self, using the toilet, taking medications, and not bedridden.

AEs will be coded using the most current MedDRA® version available at the start of the study.

5.8. Reporting Pregnancy

During study participation, all subjects should be instructed to contact the investigator immediately if they suspect they might be pregnant (eg, missed or late menstrual period). If pregnancy is confirmed, Investigational Product must be discontinued immediately, the subject should be referred to an obstetrician experienced in reproductive toxicity for evaluation and counseling, and the subject should be followed until conclusion of the pregnancy.

The investigator is responsible for reporting all available pregnancy information on the pregnancy report and submitting to the Medical Monitor within 24 hours of becoming aware of the event, although pregnancy itself is not considered an AE. Follow-up information detailing the outcome of the pregnancy and the health of the newborn should be reported as it becomes available. Any pregnancy complication must be reported as a SAE. In addition, any pregnancy resulting in a congenital abnormality or birth defect of the newborn, or neonatal death occurring within 30 days of the birth must be reported as a SAE, regardless of causality. Any infant death that occurs after the 30-day reporting period that the investigator suspects is related to Investigational Product must also be reported as an SAE.

Partner pregnancies of a male subject do not need to be reported.

5.9. Treatment Stopping Rules

If a subject has non-cutaneous adverse events of concern, clinically significant laboratory values or any condition that the investigator determines could possibly be related to the study drug, the Investigator should immediately contact the Medical Monitor to discuss if the subject should be discontinued from treatment with investigational product.

Treatment for any individual subject will be discontinued if the subject experiences:

- A serious adverse event (SAE) or a clinically significant non-serious AE which in the opinion of the Principal Investigator or Medical Monitor warrants discontinuation from the study for that subject's well-being.
- A treatment-emergent severe (Grade 3) laboratory abnormality (confirmed by repeat sample; see Appendix 13 for more information).

A subject with a PHQ-8 or modified PHQ-A score ≥15 should receive immediate referral to a mental health professional, and Investigators in consultation with the mental health professional should consider interruption or discontinuation of the study drug.

- Subjects with a PHQ-8 or modified PHQ-A score of 10-14 should be referred to a mental health professional for evaluation, regardless of the reason for their change in score

Subjects with a CDI-2 raw total score of ≥32 should be referred to a mental health professional, and Investigators in consultation with the mental health professional should consider interruption or discontinuation of the study drug.

- Subjects with a CDI-2 raw score of ≥21 for females and ≥22 for males should be referred to a mental health professional for evaluation

A subject that is experiencing suicidal ideation and behavior should be referred immediately to a qualified mental health care provider and consideration given to discontinuation from study drug.

As noted above, study treatment must be discontinued immediately in the event of a female subject's pregnancy.

Treatment should be interrupted:

- If a subject develops an application site reaction with the clinical appearance of an 'irritation reaction', and with a severity of a Dermal Response Score of 5 (erythema, edema and papules) or greater on the scale of Berger and Bowman, treatment should be interrupted for up to one week and may then be resumed if the reaction has, in the opinion of the Investigator, adequately resolved.

Treatment should be discontinued:

- If the application site reaction reoccurs, treatment should be discontinued permanently, and the subject followed until the reaction resolves.

For cases of suspected allergic contact dermatitis, the medical monitor and sponsor should be notified and there should be discussion about performing patch testing to further evaluate. Patch testing is encouraged in such cases.

In the event of a medical emergency where unblinding is required to provide medical care to the subject, refer to the most current IWRS User Manual (Section 4.8.3). Contact the Medical Monitor and the Sponsor promptly.

6. DATA ANALYSIS

Data will be handled and processed according to the Contract Research Organization's Standard Operating Procedures, which are written based on the principles of GCP.

6.1. Statistical Methods

The methodology presented below is a summary of the more detailed analysis plan that will be presented in the Statistical Analysis Plan (SAP). The SAP will be finalized before the database is locked and unblinded. Any changes to the methods described in the final SAP will be described and justified in the clinical study report.

All statistical processing will be performed using SAS® (Version 9.4) unless otherwise stated.

6.2. Determination of Sample Size

There are approximately 650 subjects planned for this study. In order to test the secondary endpoint of IGA success in subjects with a vIGA-AD score of 'Moderate' at randomization, approximately 490 of the subjects to be accrued will have vIGA-AD score of 'Moderate' at randomization. Randomization will be stratified by vIGA-AD score ('Mild' vs. 'Moderate') and by study site.

This sample size provides approximately 95% power to detect an overall 15% difference between treatment groups on vIGA-AD success at Week 4 at α=0.05 using a 2-sided stratified Cochran-Mantel-Haenszel test. The results from a recent phase 2 study (ARQ-151-212) of ARQ-151 cream 0.15% compared to vehicle treatment were used to estimate the treatment difference. Specifically, in the phase 2 trial, approximately 37% of subjects demonstrated vIGA-AD success at Week 4 in the ARQ-151 0.15% group compared to 22% in the vehicle group. This sample size also provides approximately 90% power to detect an overall 17% difference between treatment groups on IGA success at Week 4 among subjects with vIGA-AD score 'moderate' at randomization. The same testing method, the stratified Cochran-Mantel-Haenszel test, will be used as for the primary endpoint.

To control for familywise type I error at level of 0.05, the secondary endpoint of vIGA-AD success at Week 4 in subjects with vIGA-AD of 'Moderate' at randomization will only be tested if the primary endpoint demonstrates statistical significance. In addition, the remaining secondary endpoints will be inferentially tested only if the primary and above secondary endpoint (vIGA-AD success at week 4 for subjects with vIGA-AD score of 'Moderate' at randomization) comparisons are statistically significant using the hierarchical testing procedure by partitioning of the alpha and the Fallback Method.

6.3. Subjects to Analyze

The analysis populations are defined as follows:

- Intent-to-Treat (ITT) population will include all subjects who are randomized.
- Per protocol (PP) population will include all subjects in the ITT population, who are at least 80% compliant with study medication application, have a vIGA-AD assessment within the Week 4 visit window, and show no major deviations from the study protocol that would affect the interpretation of efficacy. In addition, subjects who miss the Week 4 vIGA-AD assessment specifically due to novel coronavirus disease-19 (COVID-19) disruptions will be excluded from per protocol population.
- vIGA-AD Moderate ITT population will be a subset of the ITT population with vIGA-AD score 'moderate' at randomization.
- vIGA-AD Moderate PP population will be a subset of the PP population with vIGA-AD score 'Moderate' at randomization
- WI-NRS population will be a subset of the ITT population who are ≥12 years old at Baseline and have a Baseline WI-NRS score ≥ 4.

- Safety population will include all subjects who are enrolled and received at least one confirmed dose of study medication.
- Pharmacokinetic (PK) population will include all subjects receiving the active drug with quantifiable plasma concentrations of roflumilast.

6.4. Background and Demographic Characteristics

Descriptive statistics will be used to summarize demographic characteristics (age, sex, ethnicity, and race) and background characteristics for the randomized subjects.

6.5. Study Medication Compliance

The number of study drug applications by each subject based on diary data will be summarized using summary statistics (mean, standard deviation [SD], median, minimum, and maximum), and categorically.

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

The amount of investigational product used by each subject based on tube weight will be summarized by treatment using descriptive statistics, and categorically.

Investigational product application compliance will be calculated based on number of applications divided by the expected number (amount) of investigational product applications for each subject. Compliance will be summarized descriptively by treatment group.

6.6. Safety Analysis

Descriptive statistics will be calculated for quantitative safety data and frequency counts will be compiled for classification of qualitative safety data. The safety population will be used for these analyses.

6.7. Efficacy Analysis

The Primary Efficacy Endpoint will be tested in all randomized subjects and defined as:

- IGA Success, defined as a vIGA-AD score of ‘clear’ or ‘almost clear’ PLUS a 2-grade improvement from Baseline at Week 4

The primary estimand is the ratio of the odds of achieving vIGA-AD success after 4 weeks of using ARQ-151 (roflumilast cream 0.15%), relative to the odds of success after 4 weeks of using a matching vehicle cream. In the course of the 4-week randomized treatment period, subjects may be exposed to possible known or unknown inter-current events that could possibly impact the estimates of the estimand, such as treatment discontinuation due to a specific adverse effect or perhaps a lack of effect. A composite strategy will be implemented that handles subjects who discontinue due to lack of efficacy or adverse event as missing not at random, differently than all other subjects. The “Treatment Policy Strategy” has been adopted for handling all other known or unknown intercurrent events in this study other than discontinuation due to lack of efficacy or adverse event. Subjects who discontinue due to lack of efficacy or adverse event will be treated

as non-responders at all analysis visits that occurred or would have occurred on or after the date of last dose of treatment application. Odds ratio of achieving vIGA-AD success for ARQ-151 (roflumilast cream 0.15%) relative to vehicle after 4 weeks will be evaluated accordingly. This estimand shall be estimated using the CMH approach. The primary efficacy endpoint is success in IGA of disease severity, defined as an vIGA-AD score of 'clear' or 'almost clear' PLUS a 2-grade improvement from Baseline at Week 4.

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

To control for the familywise type I error at level of 0.05, a hierarchical testing scheme will be used to test the following secondary endpoint:

- IGA success at Week 4 in subjects with a vIGA-AD score of 'Moderate' at randomization

Upon successful demonstration of statistical significance for the primary and above secondary endpoint, the remaining endpoints will be grouped into secondary endpoint family 1, comprised of the 4-point reduction on the WI-NRS endpoint, at Week 4, Week 2 and Week 1, and secondary endpoint family 2, comprised of the endpoints of EASI-75 at Week 4, vIGA-AD of 'clear' or 'almost clear' at Week 4, vIGA-AD of success at Week 2 and Week 1, vIGA-AD of 'clear' or 'almost clear' at Week 2 and Week 1. An alpha level of 0.03 will be used to test the endpoints in the secondary endpoint family 1 sequentially. An alpha level of 0.02 will be used to test the endpoints in secondary endpoint family 2 sequentially. See Figure 2 for details.

The endpoints listed below will be tested using the ITT population, with the exception of the WI-NRS endpoints, which will use the WI-NRS population.

In addition to the partitioning of the overall 0.05 alpha into two families, the Fallback Method will be applied. The fallback method is a modification of the fixed-sequence method, providing opportunity to test an endpoint later in the sequence even if an endpoint tested early in the sequence has failed to show statistical significance. The order of the endpoints remains important. The appeal of the fallback method is that if an endpoint later in the sequence has a robust treatment effect while the preceding endpoint is unsuccessful, there is a modest amount of alpha retained as a fallback to allow interpretation of that endpoint without inflating the Type I error rate. Applying the fallback method begins by dividing the total alpha (not necessarily equally) among the endpoints and maintains a fixed sequence for the testing. In this study, the Fallback Method will be applied to the fixed sequence of testing Family 1, and then Family 2.

As the testing sequence progresses, a successful test preserves its assigned alpha as "saved" ("unused" or "accumulated") alpha that is passed along to the next test in the sequence, as is the case for the sequential method. This accumulated alpha is added to the prospectively assigned alpha (if any) of that next endpoint and the summed alpha is used for testing that endpoint. Thus, as sequential tests are successful, the alpha accumulates for the endpoints later in the sequence; these endpoints are then tested with progressively larger alphas.

In this study, the Fallback Method will be applied following this sequence:

Family 1: Testing will proceed at the 0.03 level sequentially within Family 1. Should all 3 endpoints in Family 1 be statistically significant at the 0.03 level, then the full 0.03 alpha will be carried to Family 2. Family 2 would then be tested at the full (α=0.02+0.03=0.05).

Should, anywhere during the sequential testing of Family 1, there be a p-value >0.03, the testing within Family 1 will stop, and no additional alpha can be carried over to Family 2.

Secondary Endpoint Family 1 (α=0.03, hierarchical testing within Family 1)

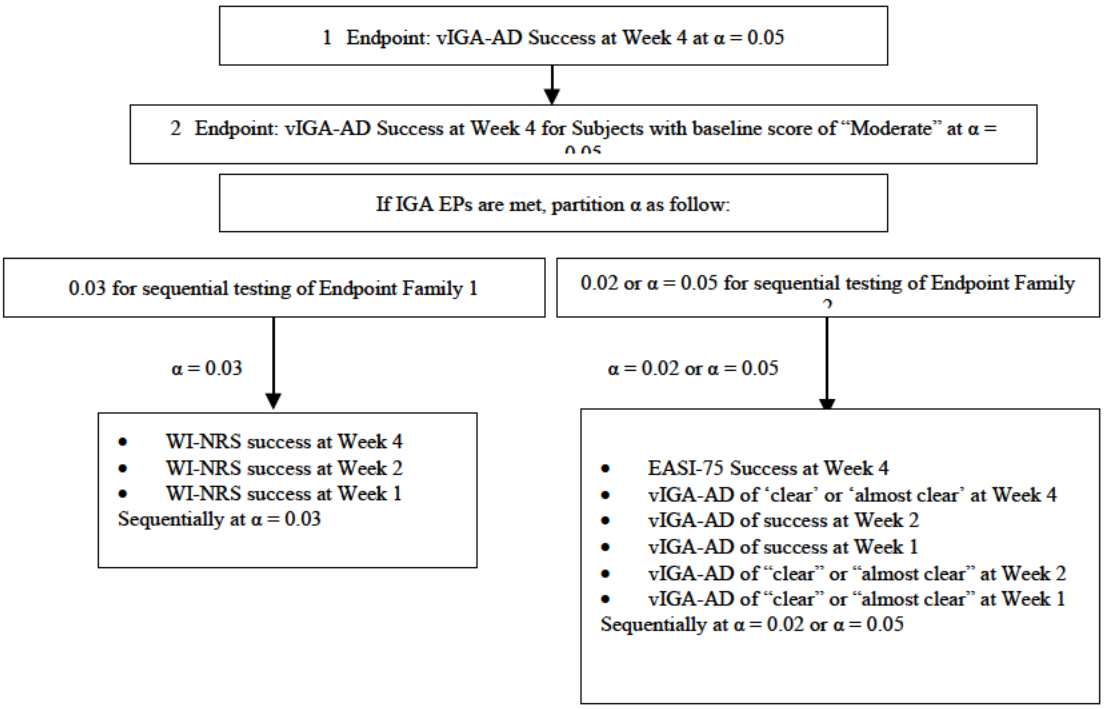
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the WI-NRS at Week 4
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the WI-NRS at Week 2
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the WI-NRS at Week 1

Secondary Endpoint Family 2 (α = 0.02 or 0.05 pending Family 1 results, hierarchical testing within Family 2)

- Achievement of at least a 75% reduction in the Eczema Area and Severity Index at Week 4 (EASI-75)
- vIGA-AD score of ‘clear’ or ‘almost clear’ at Week 4
- vIGA-AD of ‘clear’ or ‘almost clear’ at Week 4
- vIGA-AD Success at Week 2
- vIGA-AD Success at Week 1
- vIGA-AD of ‘clear’ or ‘almost clear’ at Week 2
- vIGA-AD of ‘clear’ or ‘almost clear’ at Week 1

A supplemental analysis of the primary endpoint, v-IGA Success at Week 4, and the first secondary endpoint, vIGA-AD Success at Week 4, for subjects with baseline score of “Moderate”, will be performed using the PP population and the vIGA-AD Moderate PP population, respectively.

Figure 2: Multiple Testing Scheme



Achievement of vIGA-AD success is a score of "clear" or "almost clear" plus a 2-grade improvement from baseline.
WI-WRS Success is a 4-point reduction in WI-NRS among subjects ≥12 years old with WI-NRS ≥4 at baseline.
EASI-75: achievement of at least a 75% reduction in the Eczema Area and Severity Index

6.8. Adverse Events

All TEAEs occurring during the study will be recorded and classified on the basis of MedDRA terminology for the safety population. TEAEs are defined as those AEs with an onset on or after the time of first study drug application. All reported TEAEs will be summarized by treatment group.

Descriptive statistics will be calculated for quantitative safety data and frequency counts will be compiled for classification of qualitative safety data.

Adverse Events:

The subject incidence of treatment-emergent adverse events (TEAE) will be summarized overall, by severity, and by attribution.

Clinical Laboratory Results:

Shifts in clinical laboratory parameters from baseline to worst post-baseline grade will be provided.

Vital Signs:

The subject incidence of >5% weight loss or gain on study will be provided, as well as whether weight loss was explained or unexplained.

6.9. Body Surface Area

Body surface area (BSA) affected by AD will be analyzed descriptively.

6.10. Local Tolerance Assessment

For Investigator's assessment, the numeric application site reaction scores will be summarized individually by using number and percentage of subjects by visit, as well as mean/median scores.

6.11. Medical History and Physical Examinations

Medical history for all subjects will be presented descriptively by parameter.

Physical examination for all subjects will be presented descriptively by parameter. Changes in physical examinations will be described in the text of the final report.

6.12. Clinical Laboratory Results and Vital Signs

All clinical laboratory results and vital signs measurements will be summarized descriptively by parameter, visit, and treatment group along with time point of collection.

A shift from baseline table describing out-of-normal range shifts will be provided for clinical laboratory results.

A shift from baseline table will identify subjects who gain or lose >5% body weight over the course of the study.

6.13. Prior and Concomitant Medications

Prior and concomitant medication information for all randomized subjects will be presented in a by-subject listing. Summary tables will be presented by World Health Organization-Anatomical Therapeutic Chemical Classification System (WHO-ATC) therapeutic category and product.

6.14. Subject Reported Outcomes Analyses

Weekly average WI-NRS scale will be summarized by treatment group and by visits.

6.14.1. Dermatology Life Quality Indexes, Children's Dermatology Life Quality Index, SCORAD and POEM

Dermatology Life Quality Index (DLQI), Children's Dermatology Life Quality Index (CDLQI), the Dermatitis Family Impact (DFI), the Scoring Atopic Dermatitis (SCORAD), and the Patient-oriented Eczema Measure (POEM) will be analyzed by evaluation of the reduction in total score at Week 4. These efficacy endpoints will be analyzed descriptively.

6.15. Pharmacokinetic Analysis

Plasma drug concentrations will be summarized using descriptive statistics, reporting n, mean, standard deviation, median, minimum, and maximum. The PK population will be used for these analyses.

7. STUDY ADMINISTRATION

7.1. Ethics

7.1.1. Ethics Review Board

Before screening of subjects into the study, the current protocol, ICF, and any accompanying material to be provided to the subjects will be reviewed and approved by an appropriate IRB, as required by FDA (21 CFR § 56) and ICH GCP regulations. A letter documenting the IRB approval must be received by the Sponsor before the initiation of the study at a clinical site. Amendments to the protocol will be subject to the same requirements as the original protocol.

The Investigator, Sponsor, or designee will submit a progress report at least once yearly to the IRB. However, the frequency of these reports will depend on IRB requirements. As soon as possible after completion or termination of the study, the Investigator will submit a final report to the IRB per the IRB requirements, and in compliance with FDA and ICH GCP guidelines.

The Investigator, the Sponsor, or designee shall promptly notify the IRB of any SAEs, SUSARs, or any other information that may affect the safe use of the study drug(s) during the study, per the IRB local requirements, and in compliance with FDA regulations and ICH GCP guidelines.

7.1.2. Ethical Conduct of the Study

This research will be carried out in accordance with the protocol, the principles of the Tri-Council Policy Statement (TCPS), the ethical principles set forth in the Declaration of Helsinki, and the ICH harmonized tripartite guideline regarding GCP (E6 Consolidated Guidance, April 1996).

7.1.3. Subject Information and Consent/Assent

The investigator is responsible for obtaining written informed consent from each individual participating in this study after adequate explanation (in non-technical terms) of the aims, methods, objectives, and potential hazards of the study and before undertaking any study-related procedures. The investigator must use the most current IRB- or IEC-approved consent form for documenting written informed consent. Subjects will be assured that they may withdraw from the study at any time without jeopardizing their medical care. Each informed consent (or assent as applicable) will be read, appropriately signed and dated by the subject or the subject's legally authorized representative and the person conducting the consent discussion, and also by an impartial witness if required by IRB or IEC or local requirements.

Subjects will be given a signed copy of their ICF/assent.

7.2. Study Completion and Termination

7.2.1. Study Completion

The study is considered completed with the last visit of the last subject participating in the study. The final data from the investigational site will be sent to the Sponsor (or designee) after completion of the final subject visit at that site, in the time frame specified in the Clinical Trial Agreement.

7.2.2. Study Termination

The Sponsor reserves the right to close the investigational site or terminate the study at any time for any reason at the sole discretion of the Sponsor. Investigational sites will be closed upon study completion. An investigational site is considered closed when all required documents and study supplies have been collected and a site closure visit has been performed. The investigator may initiate site closure at any time, provided there is reasonable cause and sufficient notice is given in advance of the intended termination.

Reasons for the early closure of an investigational site by the Sponsor or investigator may include but are not limited to:

- Failure of the investigator to comply with the protocol, the Sponsor's procedures, or GCP guidelines.
- Inadequate recruitment of subjects by the investigator.
- Discontinuation of further drug development.

7.3. Study Monitoring

Prior to the initiation of the clinical investigation, Sponsor representatives or designees may visit the clinical site where the investigation is to be conducted. Sponsor representatives or designees shall ensure that the Investigator understands the investigational status of the investigational product, all requirements of the investigation to be undertaken, and all of his/her responsibilities as an Investigator. Sponsor representatives or designees will also visit the clinical site at appropriate intervals as required to ensure compliance with the protocol and to verify the accuracy and completeness of data reported on the CRFs. The Study Director or designees shall be available for consultation with the Investigator and serve as liaisons between the clinical site and the Sponsor.

The Sponsor or authorized designees may inspect all documents and records required to be maintained by the Investigator, including but not limited to medical records (office, clinic, or hospital) and investigational product dispensation logs for the subjects in this clinical investigation. The Investigator must permit access to such records. The Investigator must obtain, as part of informed consent, permission for an authorized representative of the Sponsor, or regulatory authorities, to review, in confidence, any records identifying subjects.

7.4. Data Quality Assurance

In order to ensure the collection of accurate, consistent, complete, and reliable data during this clinical investigation, Sponsor representatives or designees may conduct audits of participating sites at appropriate intervals throughout the study. The results of these periodic site audits may be subject to review by independent auditors at completion of the clinical investigation.

All clinical data will undergo a quality control check prior to clinical database lock. Edit checks are performed for appropriate databases as a validation routine using SAS version 9.4 to check for missing data, data inconsistencies, data ranges etc. Corrections are made prior to database lock.

7.5. Data Handling and Record Keeping

During the clinical study, the Investigator will maintain adequate source records, including medical records, records detailing the progress of the study for each subject, laboratory reports, signed informed consent forms, IP disposition records, correspondence with the IRB and Study Monitor/Sponsor, AE reports, and information regarding subject discontinuation and completion of the clinical investigation.

All required study data will be recorded on eCRFs. Any change of data will be recorded on the audit trail and a reason for the change will be entered.

The principal investigator must retain all documentation relating to the study for a period of at least 2 years after the last marketing application approval or, if not approved, 2 years following the discontinuance of the test article for investigation. If this requirement differs from any local regulations, the local regulations will take precedence unless the local retention policy is less than 2 years.

7.6. Protocol Amendments and Deviations

No change or amendment to this protocol may be made by the investigator or Sponsor after the protocol has been agreed to and signed by all parties unless such change(s) or amendment(s) has (have) been agreed upon by the investigator and Sponsor. Any change agreed upon will be recorded in writing, and the written amendment will be signed by the investigator and Sponsor. Institutional review board approval is required prior to the implementation of an amendment, unless overriding safety reasons warrant immediate action, in which case the IRB(s) will be promptly notified.

No deviation from the protocol will be made except to protect the life or physical well-being of a subject in an emergency. Except in such emergencies, prior approval of the Sponsor, and the IRB, is required before deviations from the planned protocol. All protocol deviations that occur during the study will be documented and reported to Sponsor and to the IRB(s), if applicable, according to regulations. Further details about the documentation, evaluation, and follow-up of protocol deviations are detailed in this study's clinical monitoring plan and Protocol Deviation Management Plan (PDMP).

No waivers to inclusion/exclusion criteria will be granted; subjects need to meet all criteria, exactly as specified, to be enrolled. Additionally, prospective deviations from the protocol or investigational plan are not permitted except to protect the life or physical well-being of a subject in an emergency. Deviations that occur unintentionally or are the result of action by the subject must be documented and reported to the IRB(s), if applicable, according to regulations. Further details about the documentation, evaluation, and follow-up of protocol deviations are detailed in this study's clinical monitoring plan and PDMP.

7.7. Confidentiality and Privacy

The investigator must assure that subjects' anonymity will be strictly maintained and that their identities are protected from unauthorized parties. Only an identification code and any other unique identifier(s) as allowed by local law (such as date of birth) will be recorded on any form or biological sample submitted to the Sponsor.The investigator agrees that all information received from Arcutis Biotherapeutics, Inc., including but not limited to the investigator brochure, this protocol, CRF/eCRF, the IP, and any other study information, remain the sole and exclusive property of Arcutis Biotherapeutics, Inc. during the conduct of the study and thereafter. This information is not to be disclosed to any third party (except employees or agents directly involved in the conduct of the study or as required by law) without prior written consent from Arcutis Biotherapeutics, Inc. The investigator further agrees to take all reasonable precautions to prevent the disclosure by any employee or agent of the study site to any third party or otherwise into the public domain.

7.8. Conflict of Interest

All study investigators will provide documentation of their financial interest or arrangements with Arcutis Biotherapeutics, Inc., or proprietary interests in the investigational drug under study. This documentation must be provided prior to the investigator's participation in the study. All investigators with reported conflict of interest will be required to have such conflicts managed in a way that is appropriate to their participation in the design and conduct of this study.

7.9. Report Format

According to the ICH Harmonized Tripartite Guideline (Organization of the Common Technical Document for the Registration of Pharmaceuticals for Human Use M4 and the ICH M2 Expert Working Group), the final report will be written according to the ICH E3 Guideline (Structure and Content of Clinical Study Reports).

7.10. Publication Policy

The Sponsor is supportive of publishing clinical trial findings. Any form of publication that is derived from this study must be submitted to Arcutis Biotherapeutics Inc. for review and approval. The process of coordinating publication efforts is detailed in the Clinical Trial Agreement.

8. REFERENCES

Baumer W, Hoppmann J, Rundfeldt C, et al. Highly selective phosphodiesterase 4 inhibitors for the treatment of allergic skin diseases and psoriasis. *Inflamm Allergy Drug Targets.* 2007;6: 17-26.

Berger RS, Bowman JP. A reappraisal of the 21-day Cumulative Irritation Test in man. *J. Toxicol Ot & Ocular Toxicol.* 1982:1(2);109-115.

Bethke TD, Lahu G. High absolute bioavailability of the new oral phosphodiesterase-4 inhibitor roflumilast. *Int. J Clin Pharmacol Ther.* 2011;49:51-57.

DALIRESP (roflumilast) tablets [package insert]. AstraZeneca Pharmaceuticals LP, Wilmington, DE USA; 2018.

Dastidar SG, Rajagopal D, Ray A. Therapeutic benefit of PDE4 inhibitors in inflammatory diseases. *Curr Opin Investig Drugs*. 2007;8:364–372.

Furue M, Kitahara Y, Akama H, et al. Safety and efficacy of topical E6005, a phosphodiesterase 4 inhibitor, in Japanese adult patients with atopic dermatitis: results of a randomized, vehicle-controlled, multicenter clinical trial. *J Dermatol.* 2014;41:577-585.

Grewe SR, Chan SC, Hanifin JM. Elevated leukocyte cyclic AMP-phosphodiesterase in atopic disease: a possible mechanism for cy- clic AMP-agonist hyporesponsiveness. *J Allergy Clin Immunol.* 1982;70:452–457.

Hanifin JM, Chan SC, Cheng JB, et al. Type 4 phosphodiesterase inhibitors have clinical and in vitro anti-inflammatory effects in atopic dermatitis. *J Invest Dermatol.* 1996;107:51–56.

Hanifin JM, Thurston M, Omoto M, et al. The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. *Exp Dermatol.* 2001;10: 11-18.

Hanifin J.M. and Rajka G. Diagnostic features of atopic dermatitis. *Acta Derm Venereol Suppl (Stockh).* 1980;92: 44–47.

Hatzelmann A, Morcillo EJ, Lungarella G, et al. The preclinical pharmacology of roflumilast – a selective, oral phosphodiesterase 4 inhibitor in development for chronic obstructive pulmonary disease. *Pulm Pharmacol Ther.* 2010;23:235–256.

Investigator's Brochure, Edition 4.0, 22 January 2021.

Michalski JM, Golden G, Ikari J, et al. PDE4: A novel target in the treatment of chronic obstructive pulmonary disease. *Clin Pharmacol Ther.* 2012;91:134–142.

Murrell DF, Gebauer K, Spelman L, et al. Crisaborole topical ointment, 2% in adults with atopic dermatitis: a phase 2a, vehicle-controlled, proof-of-concept study. *J Drugs Dermatol.* 2015;14: 1108–1112.

Nazarian R, Weinberg JM. AN-2728, a PDE4 inhibitor for the potential topical treatment of psoriasis and atopic dermatitis. *Curr Opin Investig Drugs.* 2009;10:1236–1242.

Newton et al., Exploring Content and Psychometric Validity of Newly Developed Assessment Tools for Itch and Skin Pain in Atopic Dermatitis. *Journal of Patient-Reported Outcomes*. 2019; 3:42.

Nygaard U, Deleuran M, Vestergaard C. Emerging treatment options in atopic dermatitis: topical therapies. *Dermatology.* 2017;233:333-343.

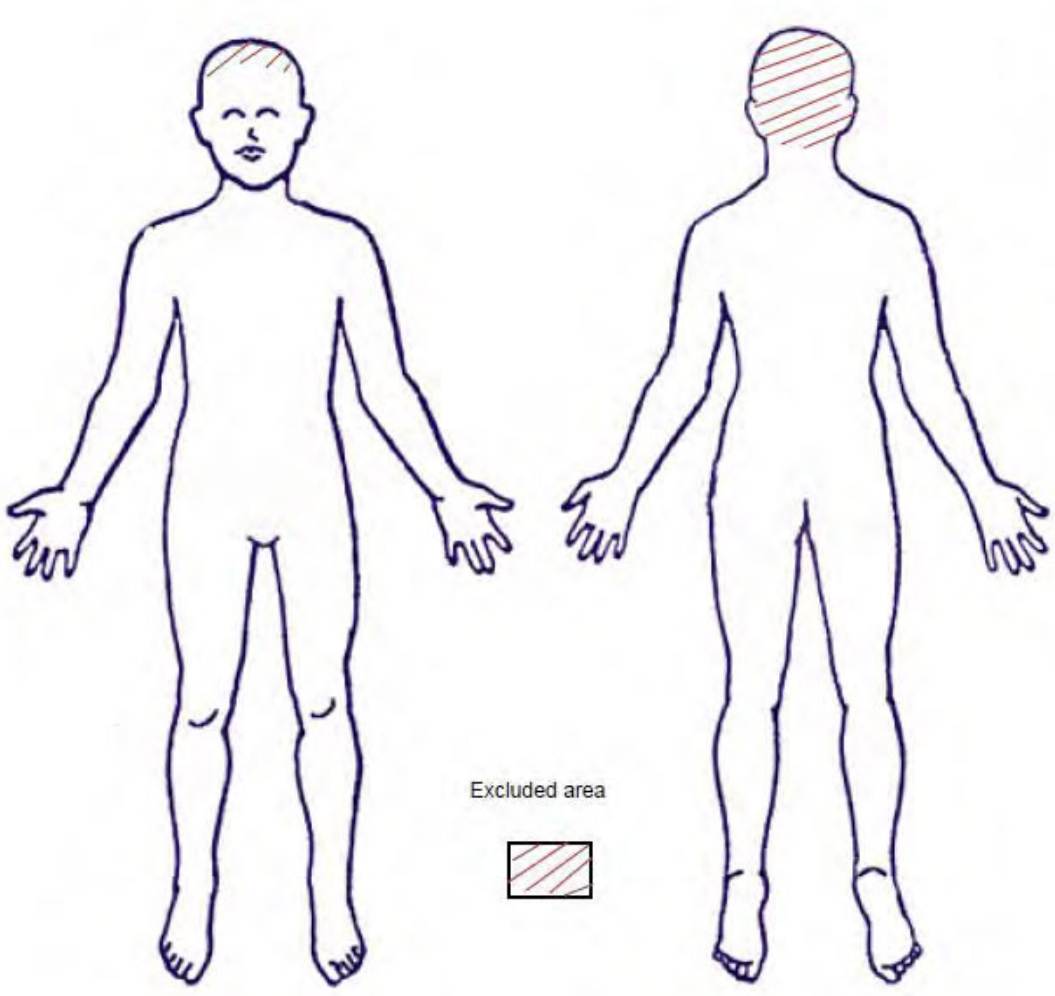
Wedzicha JA, Calverley PMA, Klaus FR. Roflumilast: a review of its use in the treatment of COPD. *Int J Chron Obstruct Pulmon Dis*. 2016;11:81-90.

9. APPENDICES

APPENDIX 1. BODY DIAGRAM

Site personnel to mark treatable areas identified by the Investigator.

(Reminder: Application will be all areas affected (except for the scalp). Continue to apply even if area(s) clears and treat new lesions (except scalp).



Study Personnel Completing Body Diagram (initial/date)
Baseline

Site to photocopy this page after updating at the Baseline and retain the original in source. Provide the copy to the subject to refer to for study application at home.

APPENDIX 2. PATIENT HEALTH QUESTIONAIRE-8 (PHQ-8)

SMRC
Self-Management
Resource Center

Personal Health Questionnaire Depression Scale (PHQ-8)

Over the **last 2 weeks**, how often have you been bothered by any of the following problems?
*(circle **one** number on each line)*

How often during the past 2 weeks were you bothered by...	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself, or that you are a failure, or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3

Scoring

If two consecutive numbers are circled, score the higher (more distress) number. If the numbers are not consecutive, do not score the item. Score is the sum of the 8 items. If more than 1 item missing, set the value of the scale to missing. A score of 10 or greater is considered major depression, 20 or more is severe major depression.

APPENDIX 3. PATIENT HEALTH QUESTIONNAIRE DEPRESSION SCALE (MODIFIED PHQ-A)

Instructions: How often have you been bothered by each of the following symptoms during the past **two weeks**? For each symptom put an **"X"** in the box beneath the answer that best describes how you have been feeling.				
	(0) Not at all	**(1) Several days**	**(2) More than half the days**	**(3) Nearly every day**
1. Feeling down, depressed, irritable, or hopeless?				
2. Little interest or pleasure in doing things?				
3. Trouble falling asleep, staying asleep, or sleeping too much?				
4. Poor appetite, weight loss, or overeating?				
5. Feeling tired, or having little energy?				
6. Feeling bad about yourself – or feeling that you are a failure, or that you have let yourself or your family down?				
7. Trouble concentrating on things like school work, reading, or watching TV?				
8. Moving or speaking so slowly that other people could have noticed? Or the opposite – being so fidgety or restless that you were moving around a lot more than usual?				

APPENDIX 4. CHILDREN'S DEPRESSION INVENTORY 2 (PARENT REPORT)

By Maria Kovacs, Ph.D.

CDI 2
PARENT

Child's Name/ID: _____ Child's Sex: Male Female (Circle One)
Parent's Name/ID: _____ Date of Birth: ____/____/____ (Year / Month / Day)
Relationship to Child: _____ Today's Date: ____/____/____ (Year / Month / Day)
Child's Age: _____ Child's Grade: _____

Instructions:

For each of the statements below, select one response that best describes your observations of your child in the **past two weeks**.

Indicate your response for each item by **circling** the number that best corresponds to your choice. You may change an item response by drawing an X through your original choice and selecting a new response.

Remember, for each statement, pick **one** answer that best describes your observations of your child in the PAST TWO WEEKS.

My child	Not at all	Some of the time	Often	Much or most of the time
1. looks sad	0	1	2	3
2. has fun.	0	1	2	3
3. does not like himself or herself.	0	1	2	3
4. blames himself or herself for things.	0	1	2	3
5. cries or looks tearful.	0	1	2	3
6. is cranky or irritable.	0	1	2	3
7. enjoys being with people.	0	1	2	3
8. thinks that he or she is ugly.	0	1	2	3
9. has to push himself or herself to do schoolwork	0	1	2	3
10. has trouble sleeping at night	0	1	2	3
11. looks tired or fatigued.	0	1	2	3
12. seems lonely.	0	1	2	3
13. enjoys school.	0	1	2	3
14. spends time with friends.	0	1	2	3
15. is showing worse school performance than before.	0	1	2	3
16. does what he or she is told.	0	1	2	3
17. has disagreements and conflicts with others.	0	1	2	3

APPENDIX 5. COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS) "BASELINE/SCREENING" VERSION

COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS)

Baseline/Screening Version

Version 1/14/09

Posner, K.; Brent, D.; Lucas, C.; Gould, M.; Stanley, B.; Brown, G.; Fisher, P.; Zelazny, J.; Burke, A.; Oquendo, M.; Mann, J.

Disclaimer:

This scale is intended to be used by individuals who have received training in its administration. The questions contained in the Columbia-Suicide Severity Rating Scale are suggested probes. Ultimately, the determination of the presence of suicidal ideation or behavior depends on the judgment of the individual administering the scale.

Definitions of behavioral suicidal events in this scale are based on those used in ***The Columbia Suicide History Form****, developed by John Mann, MD and Maria Oquendo, MD, Conte Center for the Neuroscience of Mental Disorders (CCNMD), New York State Psychiatric Institute, 1051 Riverside Drive, New York, NY, 10032. (Oquendo M. A., Halberstam B. & Mann J. J., Risk factors for suicidal behavior: utility and limitations of research instruments. In M.B. First [Ed.]* Standardized Evaluation in Clinical Practice*, pp. 103 -130, 2003.)*

For reprints of the C-SSRS contact Kelly Posner, Ph.D., New York State Psychiatric Institute, 1051 Riverside Drive, New York, New York, 10032; inquiries and training requirements contact posnerk@nyspi.columbia.edu

SUICIDAL IDEATION		
Ask questions 1 and 2. If both are negative, proceed to "Suicidal Behavior" section. If the answer to question 2 is "yes", ask questions 3, 4 and 5. If the answer to question 1 and/or 2 is "yes", complete "Intensity of Ideation" section below.	**Lifetime: Time He/She Felt Most Suicidal**	**Past __ Months**
1. Wish to be Dead Subject endorses thoughts about a wish to be dead or not alive anymore, or wish to fall asleep and not wake up. ***Have you wished you were dead or wished you could go to sleep and not wake up?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
2. Non-Specific Active Suicidal Thoughts General non-specific thoughts of wanting to end one's life/commit suicide (e.g., *"I've thought about killing myself"*) without thoughts of ways to kill oneself/associated methods, intent, or plan during the assessment period. ***Have you actually had any thoughts of killing yourself?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
3. Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act Subject endorses thoughts of suicide and has thought of at least one method during the assessment period. This is different than a specific plan with time, place or method details worked out (e.g. thought of method to kill self but not a specific plan). Includes person who would say, *"I thought about taking an overdose but I never made a specific plan as to when, where or how I would actually do it…and I would never go through with it."* ***Have you been thinking about how you might do this?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
4. Active Suicidal Ideation with Some Intent to Act, without Specific Plan Active suicidal thoughts of killing oneself and subject reports having some intent to act on such thoughts, as opposed to *"I have the thoughts but I definitely will not do anything about them."* ***Have you had these thoughts and had some intention of acting on them?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
5. Active Suicidal Ideation with Specific Plan and Intent Thoughts of killing oneself with details of plan fully or partially worked out and subject has some intent to carry it out. ***Have you started to work out or worked out the details of how to kill yourself? Do you intend to carry out this plan?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
INTENSITY OF IDEATION		
The following features should be rated with respect to the most severe type of ideation (i.e., 1-5 from above, with 1 being the least severe and 5 being the most severe). Ask about time he/she was feeling the most suicidal. Lifetime - ***Most Severe Ideation:*** ***Type # (1-5)*** ***Description of Ideation*** Past X Months - ***Most Severe Ideation:*** ***Type # (1-5)*** ***Description of Ideation***	Most Severe	Most Severe
Frequency ***How many times have you had these thoughts?*** (1) Less than once a week (2) Once a week (3) 2-5 times in week (4) Daily or almost daily (5) Many times each day	____	____
Duration ***When you have the thoughts how long do they last?*** (1) Fleeting - few seconds or minutes (2) Less than 1 hour/some of the time (3) 1-4 hours/a lot of time (4) 4-8 hours/most of day (5) More than 8 hours/persistent or continuous	____	____
Controllability ***Could/can you stop thinking about killing yourself or wanting to die if you want to?*** (1) Easily able to control thoughts (2) Can control thoughts with little difficulty (3) Can control thoughts with some difficulty (4) Can control thoughts with a lot of difficulty (5) Unable to control thoughts (0) Does not attempt to control thoughts	____	____
Deterrents ***Are there things - anyone or anything (e.g., family, religion, pain of death) - that stopped you from wanting to die or acting on thoughts of committing suicide?*** (1) Deterrents definitely stopped you from attempting suicide (2) Deterrents probably stopped you (3) Uncertain that deterrents stopped you (4) Deterrents most likely did not stop you (5) Deterrents definitely did not stop you (0) Does not apply	____	____
Reasons for Ideation ***What sort of reasons did you have for thinking about wanting to die or killing yourself? Was it to end the pain or stop the way you were feeling (in other words you couldn't go on living with this pain or how you were feeling) or was it to get attention, revenge or a reaction from others? Or both?*** (1) Completely to get attention, revenge or a reaction from others (2) Mostly to get attention, revenge or a reaction from others (3) Equally to get attention, revenge or a reaction from others and to end/stop the pain (4) Mostly to end or stop the pain (you couldn't go on living with the pain or how you were feeling) (5) Completely to end or stop the pain (you couldn't go on living with the pain or how you were feeling) (0) Does not apply	____	____

C-SSRS—Baseline/Screening (Version 1/14/09)
Page 1 of 2

SUICIDAL BEHAVIOR *(Check all that apply, so long as these are separate events; must ask about all types)*	**Lifetime**	**Past __ Years**
Actual Attempt: A potentially self-injurious act committed with at least some wish to die, *as a result of act.* Behavior was in part thought of as method to kill oneself. Intent does not have to be 100%. If there is ***any*** intent/desire to die associated with the act, then it can be considered an actual suicide attempt. ***There does not have to be any injury or harm***, just the potential for injury or harm. If person pulls trigger while gun is in mouth but gun is broken so no injury results, this is considered an attempt. Inferring Intent: Even if an individual denies intent/wish to die, it may be inferred clinically from the behavior or circumstances. For example, a highly lethal act that is clearly not an accident so no other intent but suicide can be inferred (e.g., gunshot to head, jumping from window of a high floor/story). Also, if someone denies intent to die, but they thought that what they did could be lethal, intent may be inferred. ***Have you made a suicide attempt?*** ***Have you done anything to harm yourself?*** ***Have you done anything dangerous where you could have died?*** ***What did you do?*** ***Did you_____ as a way to end your life?*** ***Did you want to die (even a little) when you_____?*** ***Were you trying to end your life when you_____?*** ***Or Did you think it was possible you could have died from_____?*** ***Or did you do it purely for other reasons / without ANY intention of killing yourself (like to relieve stress, feel better, get sympathy, or get something else to happen)?*** (Self-Injurious Behavior without suicidal intent) If yes, describe:	**Yes** ☐ **No** ☐ Total # of Attempts _____	**Yes** ☐ **No** ☐ Total # of Attempts _____
Has subject engaged in Non-Suicidal Self-Injurious Behavior?	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
Interrupted Attempt: When the person is interrupted (by an outside circumstance) from starting the potentially self-injurious act *(if not for that, actual attempt would have occurred).* Overdose: Person has pills in hand but is stopped from ingesting. Once they ingest any pills, this becomes an attempt rather than an interrupted attempt. Shooting: Person has gun pointed toward self, gun is taken away by someone else, or is somehow prevented from pulling trigger. Once they pull the trigger, even if the gun fails to fire, it is an attempt. Jumping: Person is poised to jump, is grabbed and taken down from ledge. Hanging: Person has noose around neck but has not yet started to hang - is stopped from doing so. ***Has there been a time when you started to do something to end your life but someone or something stopped you before you actually did anything?*** If yes, describe:	**Yes** ☐ **No** ☐ Total # of interrupted _____	**Yes** ☐ **No** ☐ Total # of interrupted _____
Aborted Attempt: When person begins to take steps toward making a suicide attempt, but stops themselves before they actually have engaged in any self-destructive behavior. Examples are similar to interrupted attempts, except that the individual stops him/herself, instead of being stopped by something else. ***Has there been a time when you started to do something to try to end your life but you stopped yourself before you actually did anything?*** If yes, describe:	**Yes** ☐ **No** ☐ Total # of aborted _____	**Yes** ☐ **No** ☐ Total # of aborted _____
Preparatory Acts or Behavior: Acts or preparation towards imminently making a suicide attempt. This can include anything beyond a verbalization or thought, such as assembling a specific method (e.g., buying pills, purchasing a gun) or preparing for one's death by suicide (e.g., giving things away, writing a suicide note). ***Have you taken any steps towards making a suicide attempt or preparing to kill yourself (such as collecting pills, getting a gun, giving valuables away or writing a suicide note)?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
Suicidal Behavior: Suicidal behavior was present during the assessment period?	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐

Answer for Actual Attempts Only	Most Recent Attempt Date:	Most Lethal Attempt Date:	Initial/First Attempt Date:
Actual Lethality/Medical Damage: 0. No physical damage or very minor physical damage (e.g., surface scratches). 1. Minor physical damage (e.g., lethargic speech; first-degree burns; mild bleeding; sprains). 2. Moderate physical damage; medical attention needed (e.g., conscious but sleepy, somewhat responsive; second-degree burns; bleeding of major vessel). 3. Moderately severe physical damage; *medical* hospitalization and likely intensive care required (e.g., comatose with reflexes intact; third-degree burns less than 20% of body; extensive blood loss but can recover; major fractures). 4. Severe physical damage; *medical* hospitalization with intensive care required (e.g., comatose without reflexes; third-degree burns over 20% of body; extensive blood loss with unstable vital signs; major damage to a vital area). 5. Death	*Enter Code* _____	*Enter Code* _____	*Enter Code* _____
Potential Lethality: Only Answer if Actual Lethality=0 Likely lethality of actual attempt if no medical damage (the following examples, while having no actual medical damage, had potential for very serious lethality: put gun in mouth and pulled the trigger but gun fails to fire so no medical damage; laying on train tracks with oncoming train but pulled away before run over). 0 = Behavior not likely to result in injury 1 = Behavior likely to result in injury but not likely to cause death 2 = Behavior likely to result in death despite available medical care	*Enter Code* _____	*Enter Code* _____	*Enter Code* _____

APPENDIX 6. COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS) "SINCE LAST VISIT" VERSION

COLUMBIA-SUICIDE SEVERITY

RATING SCALE

(C-SSRS)

Since Last Visit

Version 1/14/09

Posner, K.; Brent, D.; Lucas, C.; Gould, M.; Stanley, B.; Brown, G.; Fisher, P.; Zelazny, J.; Burke, A.; Oquendo, M.; Mann, J.

Disclaimer:

This scale is intended to be used by individuals who have received training in its administration. The questions contained in the Columbia-Suicide Severity Rating Scale are suggested probes. Ultimately, the determination of the presence of suicidal ideation or behavior depends on the judgment of the individual administering the scale.

Definitions of behavioral suicidal events in this scale are based on those used in ***The Columbia Suicide History Form****, developed by John Mann, MD and Maria Oquendo, MD, Conte Center for the Neuroscience of Mental Disorders (CCNMD), New York State Psychiatric Institute, 1051 Riverside Drive, New York, NY, 10032. (Oquendo M. A., Halberstam B. & Mann J. J., Risk factors for suicidal behavior: utility and limitations of research instruments. In M.B. First [Ed.]* Standardized Evaluation in Clinical Practice, *pp. 103 -130, 2003.)*

For reprints of the C-SSRS contact Kelly Posner, Ph.D., New York State Psychiatric Institute, 1051 Riverside Drive, New York, New York, 10032; inquiries and training requirements contact posnerk@nyspi.columbia.edu

SUICIDAL IDEATION

Ask questions 1 and 2. If both are negative, proceed to "Suicidal Behavior" section. If the answer to question 2 is "yes", ask questions 3, 4 and 5. If the answer to question 1 and/or 2 is "yes", complete "Intensity of Ideation" section below.

Since Last Visit

1. Wish to be Dead — Yes ☐ No ☐
Subject endorses thoughts about a wish to be dead or not alive anymore, or wish to fall asleep and not wake up.
Have you wished you were dead or wished you could go to sleep and not wake up?
If yes, describe:

2. Non-Specific Active Suicidal Thoughts — Yes ☐ No ☐
General, non-specific thoughts of wanting to end one's life/commit suicide (e.g., *"I've thought about killing myself"*) without thoughts of ways to kill oneself/associated methods, intent, or plan during the assessment period.
Have you actually had any thoughts of killing yourself?
If yes, describe:

3. Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act — Yes ☐ No ☐
Subject endorses thoughts of suicide and has thought of at least one method during the assessment period. This is different than a specific plan with time, place or method details worked out (e.g., thought of method to kill self but not a specific plan). Includes person who would say, *"I thought about taking an overdose but I never made a specific plan as to when, where or how I would actually do it…and I would never go through with it."*
Have you been thinking about how you might do this?
If yes, describe:

4. Active Suicidal Ideation with Some Intent to Act, without Specific Plan — Yes ☐ No ☐
Active suicidal thoughts of killing oneself and subject reports having some intent to act on such thoughts, as opposed to *"I have the thoughts but I definitely will not do anything about them."*
Have you had these thoughts and had some intention of acting on them?
If yes, describe:

5. Active Suicidal Ideation with Specific Plan and Intent — Yes ☐ No ☐
Thoughts of killing oneself with details of plan fully or partially worked out and subject has some intent to carry it out.
Have you started to work out or worked out the details of how to kill yourself? Do you intend to carry out this plan?
If yes, describe:

INTENSITY OF IDEATION

The following features should be rated with respect to the most severe type of ideation (i.e., 1-5 from above, with 1 being the least severe and 5 being the most severe).

Most Severe Ideation: ____ _____

Type # (1-5) ***Description of Ideation***

Most Severe

Frequency — ____
How many times have you had these thoughts?
(1) Less than once a week (2) Once a week (3) 2-5 times in week (4) Daily or almost daily (5) Many times each day

Duration — ____
When you have the thoughts, how long do they last?
(1) Fleeting - few seconds or minutes
(2) Less than 1 hour/some of the time
(3) 1-4 hours/a lot of time
(4) 4-8 hours/most of day
(5) More than 8 hours/persistent or continuous

Controllability — ____
Could/can you stop thinking about killing yourself or wanting to die if you want to?
(1) Easily able to control thoughts
(2) Can control thoughts with little difficulty
(3) Can control thoughts with some difficulty
(4) Can control thoughts with a lot of difficulty
(5) Unable to control thoughts
(0) Does not attempt to control thoughts

Deterrents — ____
Are there things - anyone or anything (e.g., family, religion, pain of death) - that stopped you from wanting to die or acting on thoughts of committing suicide?
(1) Deterrents definitely stopped you from attempting suicide
(2) Deterrents probably stopped you
(3) Uncertain that deterrents stopped you
(4) Deterrents most likely did not stop you
(5) Deterrents definitely did not stop you
(0) Does not apply

Reasons for Ideation — ____
What sort of reasons did you have for thinking about wanting to die or killing yourself? Was it to end the pain or stop the way you were feeling (in other words you couldn't go on living with this pain or how you were feeling) or was it to get attention, revenge or a reaction from others? Or both?
(1) Completely to get attention, revenge or a reaction from others
(2) Mostly to get attention, revenge or a reaction from others
(3) Equally to get attention, revenge or a reaction from others and to end/stop the pain
(4) Mostly to end or stop the pain (you couldn't go on living with the pain or how you were feeling)
(5) Completely to end or stop the pain (you couldn't go on living with the pain or how you were feeling)
(0) Does not apply

SUICIDAL BEHAVIOR *(Check all that apply, so long as these are separate events; must ask about all types)*	**Since Last Visit**
Actual Attempt: A potentially self-injurious act committed with at least some wish to die, *as a result of act.* Behavior was in part thought of as method to kill oneself. Intent does not have to be 100%. If there is ***any*** intent/desire to die associated with the act, then it can be considered an actual suicide attempt. ***There does not have to be any injury or harm***, just the potential for injury or harm. If person pulls trigger while gun is in mouth but gun is broken so no injury results, this is considered an attempt. Inferring Intent: Even if an individual denies intent/wish to die, it may be inferred clinically from the behavior or circumstances. For example, a highly lethal act that is clearly not an accident so no other intent but suicide can be inferred (e.g., gunshot to head, jumping from window of a high floor/story). Also, if someone denies intent to die, but they thought that what they did could be lethal, intent may be inferred. ***Have you made a suicide attempt?*** ***Have you done anything to harm yourself?*** ***Have you done anything dangerous where you could have died?*** ***What did you do?*** ***Did you_____ as a way to end your life?*** ***Did you want to die (even a little) when you_____?*** ***Were you trying to end your life when you_____?*** ***Or did you think it was possible you could have died from_____?*** ***Or did you do it purely for other reasons / without ANY intention of killing yourself (like to relieve stress, feel better, get sympathy, or get something else to happen)?*** (Self-Injurious Behavior without suicidal intent) If yes, describe:	Yes ☐ No ☐ Total # of Attempts _____
Has subject engaged in Non-Suicidal Self-Injurious Behavior?	Yes ☐ No ☐
Interrupted Attempt: When the person is interrupted (by an outside circumstance) from starting the potentially self-injurious act *(if not for that, actual attempt would have occurred).* Overdose: Person has pills in hand but is stopped from ingesting. Once they ingest any pills, this becomes an attempt rather than an interrupted attempt. Shooting: Person has gun pointed toward self, gun is taken away by someone else, or is somehow prevented from pulling trigger. Once they pull the trigger, even if the gun fails to fire, it is an attempt. Jumping: Person is poised to jump, is grabbed and taken down from ledge. Hanging: Person has noose around neck but has not yet started to hang - is stopped from doing so. ***Has there been a time when you started to do something to end your life but someone or something stopped you before you actually did anything?*** If yes, describe:	Yes ☐ No ☐ Total # of interrupted _____
Aborted Attempt: When person begins to take steps toward making a suicide attempt, but stops themselves before they actually have engaged in any self-destructive behavior. Examples are similar to interrupted attempts, except that the individual stops him/herself, instead of being stopped by something else. ***Has there been a time when you started to do something to try to end your life but you stopped yourself before you actually did anything?*** If yes, describe:	Yes ☐ No ☐ Total # of aborted _____
Preparatory Acts or Behavior: Acts or preparation towards imminently making a suicide attempt. This can include anything beyond a verbalization or thought, such as assembling a specific method (e.g., buying pills, purchasing a gun) or preparing for one's death by suicide (e.g., giving things away, writing a suicide note). ***Have you taken any steps towards making a suicide attempt or preparing to kill yourself (such as collecting pills, getting a gun, giving valuables away or writing a suicide note)?*** If yes, describe:	Yes ☐ No ☐
Suicidal Behavior: Suicidal behavior was present during the assessment period?	Yes ☐ No ☐
Suicide:	Yes ☐ No ☐
Answer for Actual Attempts Only	Most Lethal Attempt Date:
Actual Lethality/Medical Damage: 0. No physical damage or very minor physical damage (e.g., surface scratches). 1. Minor physical damage (e.g., lethargic speech; first-degree burns; mild bleeding; sprains). 2. Moderate physical damage; medical attention needed (e.g., conscious but sleepy, somewhat responsive; second-degree burns; bleeding of major vessel). 3. Moderately severe physical damage; *medical* hospitalization and likely intensive care required (e.g., comatose with reflexes intact; third-degree burns less than 20% of body; extensive blood loss but can recover; major fractures). 4. Severe physical damage; *medical* hospitalization with intensive care required (e.g., comatose without reflexes; third-degree burns over 20% of body; extensive blood loss with unstable vital signs; major damage to a vital area). 5. Death	*Enter Code* _____
Potential Lethality: Only Answer if Actual Lethality=0 Likely lethality of actual attempt if no medical damage (the following examples, while having no actual medical damage, had potential for very serious lethality: put gun in mouth and pulled the trigger but gun fails to fire so no medical damage; laying on train tracks with oncoming train but pulled away before run over). 0 = Behavior not likely to result in injury 1 = Behavior likely to result in injury but not likely to cause death 2 = Behavior likely to result in death despite available medical care	*Enter Code* _____

APPENDIX 7. VALIDATED INVESTIGATOR GLOBAL ASSESSMENT SCALE FOR ATOPIC DERMATITIS

Validated Investigator Global Assessment scale for Atopic Dermatitis

vIGA-AD™

Instructions:

The IGA score is selected using the descriptors below that best describe the overall appearance of the lesions at a given time point. It is not necessary that all characteristics under Morphological Description be present.

Score	Morphological Description
0 – Clear	No inflammatory signs of atopic dermatitis (no erythema, no induration/papulation, no lichenification, no oozing/crusting). Post-inflammatory hyperpigmentation and/or hypopigmentation may be present.
1 – Almost clear	Barely perceptible erythema, barely perceptible induration/papulation, and/or minimal lichenification. No oozing or crusting.
2 – Mild	Slight but definite erythema (pink), slight but definite induration/papulation, and/or slight but definite lichenification. No oozing or crusting.
3 – Moderate	Clearly perceptible erythema (dull red), clearly perceptible induration/papulation, and/or clearly perceptible lichenification. Oozing and crusting may be present.
4 – Severe	Marked erythema (deep or bright red), marked induration/papulation, and/or marked lichenification. Disease is widespread in extent. Oozing or crusting may be present.

Notes:

1. In indeterminate cases, please use extent to differentiate between scores.

For example:

- *Patient with marked erythema (deep or bright red), marked papulation and/or marked lichenification that is limited in extent, will be considered "3 – Moderate".*

2. Excoriations should not be considered when assessing disease severity.

APPENDIX 8. DERMATOLOGY LIFE QUALITY INDEX

DLQI

Site No: Date: Score:
Name:
Address: Diagnosis:

The aim of this questionnaire is to measure how much your skin problem has affected your life OVER THE LAST WEEK. Please tick ☑ one box for each question.

1.	Over the last week, how **itchy**, **sore**, **painful** or **stinging** has your skin been?	Very much A lot A little Not at all	❒ ❒ ❒ ❒
2.	Over the last week, how **embarrassed** or **self conscious** have you been because of your skin?	Very much A lot A little Not at all	❒ ❒ ❒ ❒
3.	Over the last week, how much has your skin interfered with you going **shopping** or looking after your **home** or **garden**?	Very much A lot A little Not at all Not relevant	❒ ❒ ❒ ❒ ❒
4.	Over the last week, how much has your skin influenced the **clothes** you wear?	Very much A lot A little Not at all Not relevant	❒ ❒ ❒ ❒ ❒
5.	Over the last week, how much has your skin affected any **social** or **leisure** activities?	Very much A lot A little Not at all Not relevant	❒ ❒ ❒ ❒ ❒
6.	Over the last week, how much has your skin made it difficult for you to do any **sport**?	Very much A lot A little Not at all Not relevant	❒ ❒ ❒ ❒ ❒
7.	Over the last week, has your skin prevented you from **working** or **studying**?	Yes No Not relevant	❒ ❒ ❒
	If "No", over the last week how much has your skin been a problem at **work** or **studying**?	A lot A little Not at all	❒ ❒ ❒

8.	Over the last week, how much has your skin created problems with your **partner** or any of your **close friends** or **relatives**?	Very much ❒ A lot ❒ A little ❒ Not at all ❒ Not relevant ❒
9.	Over the last week, how much has your skin caused any **sexual difficulties**?	Very much ❒ A lot ❒ A little ❒ Not at all ❒ Not relevant ❒
10.	Over the last week, how much of a problem has the **treatment** for your skin been, for example by making your home messy, or by taking up time?	Very much ❒ A lot ❒ A little ❒ Not at all ❒ Not relevant ❒

Please check you have answered EVERY question. Thank you.

APPENDIX 9. CHILDREN'S DERMATOLOGY LIFE QUALITY INDEX

Site No.:
Name: Diagnosis: CDLQI
Age: SCORE:
Address: Date:

The aim of this questionnaire is to measure how much your skin problem has affected you OVER THE LAST WEEK. Please tick ✓ one box for each question.

1.	Over the last week, how **itchy**, **"scratchy"**, **sore** or **painful** has your skin been?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
2.	Over the last week, how **embarrassed** or **self conscious**, **upset** or **sad** have you been because of your skin?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
3.	Over the last week, how much has your skin affected your **friendships**?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
4.	Over the last week, how much have you changed or worn **different** or **special clothes/shoes** because of your skin?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
5.	Over the last week, how much has your skin trouble affected **going out**, **playing**, or **doing hobbies**?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
6.	Over the last week, how much have you avoided **swimming** or **other sports** because of your skin trouble?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
7.	Last week, was it **school time?** **OR** was it **holiday time?**	**If school time**: Over the last week, how much did your skin problem affect your **school work**? **If holiday time**: How much over the last week, has your skin problem interfered with your enjoyment of the **holiday**?	Prevented school Very much Quite a lot Only a little Not at all Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

8.	Over the last week, how much trouble have you had because of your skin with other people **calling you names**, **teasing**, **bullying**, **asking questions** or **avoiding you**?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
9.	Over the last week, how much has your **sleep** been affected by your skin problem?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
10.	Over the last week, how much of a problem has the **treatment** for your skin been?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐

APPENDIX 10. DERMATITIS FAMILY IMPACT QUESTIONNAIRE (DFI)

Child's Name: **Mother/Father/Carer** **Date:** **Score** ☐

The aim of this questionnaire is to measure how much your child's skin problem has affected you and your family OVER THE LAST WEEK. Please tick ☑ one box for each question.

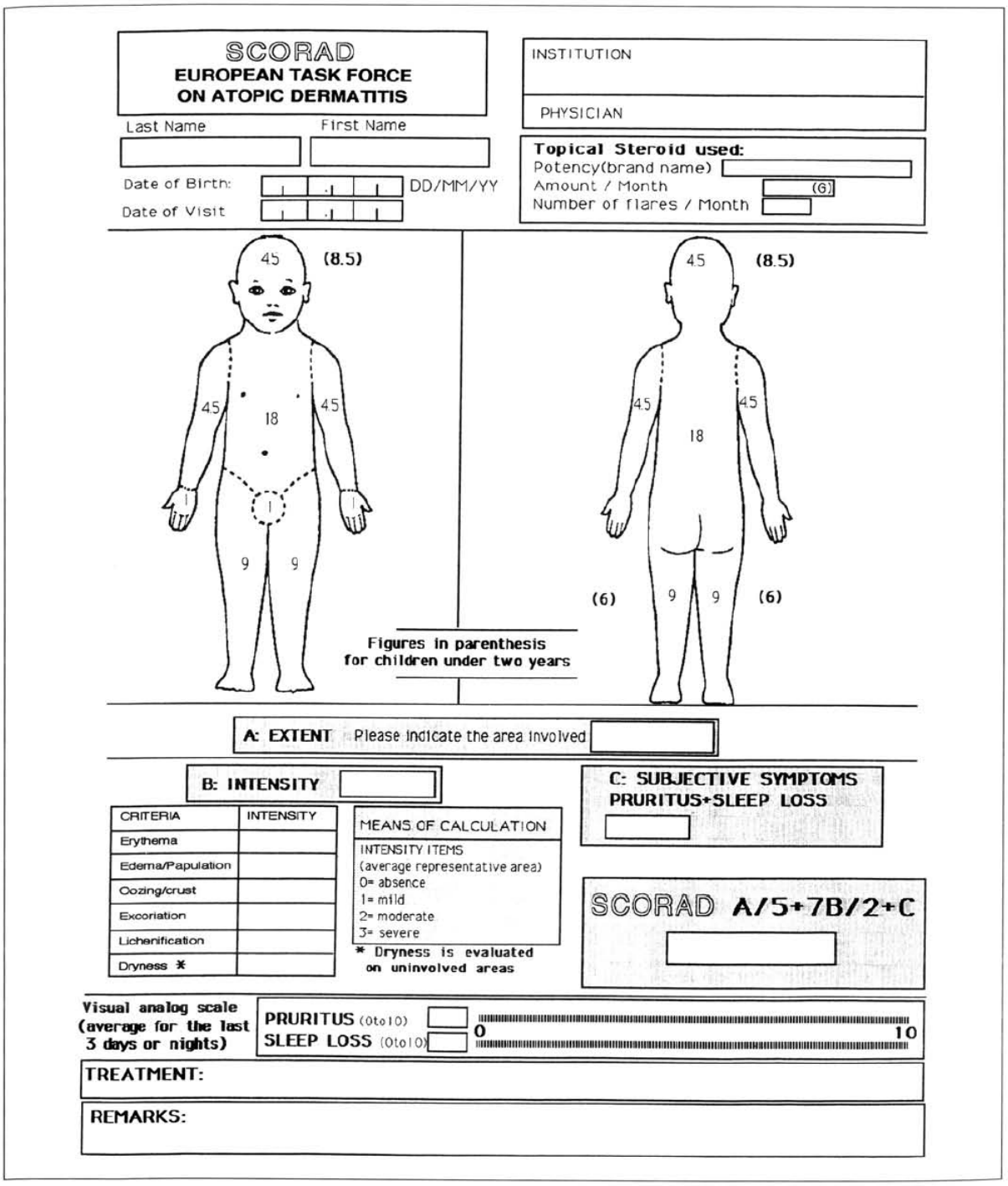
1.	Over the last week, how much effect has your child having eczema had on **housework**, e.g. washing, cleaning.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
2.	Over the last week, how much effect has your child having eczema had on **food preparation** and **feeding**.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
3.	Over the last week, how much effect has your child having eczema had on the **sleep of others in family**.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
4.	Over the last week, how much effect has your child having eczema had on **family leisure activities**, eg swimming.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
5.	Over the last week, how much effect has your child having eczema had on **time spent on shopping for the family.**	Very much A lot A little Not at all	❒ ❒ ❒ ❒
6.	Over the last week, how much effect has your child having eczema had on your **expenditure**, eg costs related to treatment, clothes, etc.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
7.	Over the last week, how much effect has your child having eczema had on causing **tiredness** or **exhaustion** in your child's parents/carers.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
8.	Over the last week, how much effect has your child having eczema had on causing **emotional distress** such as depression, frustration or guilt in your child's parents/carers.	Very much A lot A little Not at all	❒ ❒ ❒ ❒

9.	Over the last week, how much effect has your child having eczema had on **relationships** between the **main carer and partner** or between the **main carer and other children** in the family.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
10.	Over the last week, how much effect has **helping with your child's treatment** had on the main carer's life.	Very much A lot A little Not at all	❒ ❒ ❒ ❒

Please check you have answered EVERY question. Thank you

APPENDIX 11. SCORAD

SCORAD https://www.ncbi.nlm.nih.gov/pubmed/8435513



APPENDIX 12. PATIENT-ORIENTED ECZEMA MEASURE (POEM)

POEM
Patient-Oriented Eczema Measure

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

POEM for self-completion and/or proxy completion

Patient Details: _____

_____

_____

_____ Date: _____

Please circle one response for each of the seven questions below about your/your child's eczema. If your child is old enough to understand the questions then please fill in the questionnaire together. Please leave blank any questions you feel unable to answer.

1. Over the last week, on how many days has your/your child's skin been itchy because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

2. Over the last week, on how many nights has your/your child's sleep been disturbed because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

3. Over the last week, on how many days has your/your child's skin been bleeding because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

4. Over the last week, on how many days has your/your child's skin been weeping or oozing clear fluid because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

5. Over the last week, on how many days has your/your child's skin been cracked because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

6. Over the last week, on how many days has your/your child's skin been flaking off because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

7. Over the last week, on how many days has your/your child's skin felt dry or rough because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

Total POEM Score (Maximum 28): .....

POEM
Patient-Oriented Eczema Measure

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

POEM for self-completion and/or proxy completion

How is the scoring done?

Each of the seven questions carries equal weight and is scored from 0 to 4 as follows:

No days	= 0
1-2 days	= 1
3-4 days	= 2
5-6 days	= 3
Every day	= 4

Note:

- If one question is left unanswered this is scored 0 and the scores are summed and expressed as usual out of a maximum of 28
- If two or more questions are left unanswered the questionnaire is not scored
- If two or more response options are selected, the response option with the highest score should be recorded

What does a poem score mean?

To help patients and clinicians to understand their POEM scores, the following bandings have been established (see references below):

• 0 to 2	= Clear or almost clear
• 3 to 7	= Mild eczema
• 8 to 16	= Moderate eczema
• 17 to 24	= Severe eczema
• 25 to 28	= Very severe eczema

Do I need permission to use the scale?

Whilst the POEM scale is protected by copyright, it is freely available for use and can be downloaded from: www.nottingham.ac.uk/dermatology
We do however ask that you register your use of the POEM by e-mailing cebd@nottingham.ac.uk with details of how you would like to use the scale, and which countries the scale will be used in.

References

Charman CR, Venn AJ, Williams HC. The Patient-Oriented Eczema Measure: Development and Initial Validation of a New Tool for Measuring Atopic Eczema Severity From the Patients' Perspective. Arch Dermatol. 2004;140:1513-1519

Charman CR, Venn AJ, Ravenscroft JC, Williams HC. Translating Patient-Oriented Eczema Measure (POEM) scores into clinical practice by suggesting severity strata derived using anchor-based methods. Br J Dermatol. Dec 2013; 169(6): 1326–1332.

APPENDIX 13. NIAID DMID TOXICITY TABLE

NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DISEASES, DIVISION OF MICROBIOLOGY AND INFECTIOUS DISEASES (DMID) Toxicity Table for Use in Trials Enrolling Healthy Adults (2014) Modified

ABBREVIATIONS USED IN FOLLOWING TABLES:

Abbreviation/Term	Definition/Explanation
ALT	alanine aminotransferase
aPTT	activated partial thromboplastin time
AST	aspartate aminotransferase
AV block	atrioventricular block
bpm	beats per minute
BUN	blood urea nitrogen
CK	creatine kinase
CPK	creatine phosphokinase
FEV_1	forced expiratory volume in 1 second
g	Gram
HI	High
HPF	high power field
IU	international unit
IV	Intravenous
K/CUMM	$x10^3/mm^3$
LLN	lower limit of normal

Abbreviation/Term	Definition/Explanation
LO	Low
mEq	Milliequivalent
mmHg	millimeter of mercury
Ms	Millisecond
N	Normal
PT	prothrombin time
PTT	partial thromboplastin time
QTc	QT-interval corrected for heart rate
QTcB	Bazett's corrected QT interval
QTcF	Fridericia's corrected QT interval
RBC	red blood cell
Rx	Therapy
S	Second
U	Unit
ULN	upper limit of normal

ESTIMATING SEVERITY GRADE

For abnormalities NOT found elsewhere in the Toxicity Tables use the scale below to estimate grade of severity:

GRADE 1	**Mild**:	Transient or mild discomfort (<48 hours); no medical intervention/therapy required
GRADE 2	**Moderate**:	Mild to moderate limitation in activity - some assistance may be needed; no or minimal medical intervention/therapy required
GRADE 3	**Severe**:	Marked limitation in activity, some assistance usually required; medical intervention/therapy required hospitalizations possible.

LABORATORY RANGES

Where discrepancies in the ULN and LLN of laboratory ranges occur between those included in this document and those of the laboratory that performs the assays, the values provided by the laboratory will be used for assignment of severity grade.

CLINICAL ADVERSE EVENTS

Cardiovascular	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Arrhythmia	-	Asymptomatic, transient signs, no Rx required	Recurrent/persistent; symptomatic Rx required
Hemorrhage, blood loss	Estimated blood loss ≤100 mL	Estimated blood loss >100 mL, no transfusion required	Transfusion required
QTcF (Fridericia's correction)[a] or QTcB (Bazett's correction)	Asymptomatic, QTc interval 450-479 ms, *OR* Increase in interval<30 ms above baseline	Asymptomatic, QTc interval 480-499 ms, *OR* Increase in interval 30-59 ms above baseline	Asymptomatic, QTc interval ≥500 ms, *OR* Increase in interval ≥60 ms above baseline
PR interval (prolonged)	PR interval 0.20-0.25 s	PR interval >0.25 s	Type II 2nd degree AV block *OR* Ventricular pause >3.0 s
Respiratory	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Cough	Transient-no treatment	Persistent cough	Interferes with daily activities
Bronchospasm, acute	Transient wheeze; no treatment	Requires treatment; normalizes with bronchodilator and FEV_1 <80% predicted before bronchodilator	Minimal normalization with bronchodilator andFEV_1 <80% predicted after bronchodilator
Dyspnea	Does not interfere with usual and social activities	Interferes with usual and social activities, no treatment	Prevents daily and usual social activity or requires treatment

[a] Inclusion dependent upon protocol requirements

Respiratory	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Nasal discharge (rhinitis infective per **CTCAE 4.0**)	-	Localized; local intervention indicated (eg, topical antibiotic, antifungal, or antiviral)	-
Pharyngitis **(CTCAE 4.0)**	-	Localized; local intervention indicated (eg, topical antibiotic, antifungal, or antiviral)	IV antibiotic, antifungal, or antiviral intervention indicated; radiologic or operative intervention indicated
Pneumonitis (rales or rhonchi) **(CTCAE 4.0)**	Asymptomatic; clinical or diagnostic observations only; intervention not indicated	Symptomatic; medical intervention indicated; limiting instrumental ADL	Severe symptoms; limiting self-care ADL; oxygen indicated
Lung infection (**CTCAE 4.0**)	-	Moderate symptoms; oral intervention indicated (eg, antibiotic, antifungal, antiviral)	IV antibiotic, antifungal, or antiviral intervention indicated; radiologic, endoscopic, or operative intervention indicated
Gastrointestinal	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Nausea	No interference with activity	Some interference with activity	Prevents daily activities
Vomiting	No interference with activity or 1-2 episodes/24 hours	Some interference with activity or >2 episodes/24 hours	Prevents daily activity or requires IV hydration
Diarrhea	2-3 loose or watery stools or <400 g/24 hours	4-5 loose or watery stools or 400-800 g/24 hours	6 or more loose or watery stools or >800 g/24 hours or requires IV hydration

Urinary Tract	Mild (Grade 1)	Moderate (Grade 2)	Severe (Grade 3)
Urinary tract infection **(CTCAE 4.0)**	-	Localized; local intervention indicated (eg, oral or topical antibiotic, antifungal, antiviral)	IV antibiotic, antifungal, or antiviral intervention indicated; radiologic or operative intervention indicated
Reactogenicity	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Local reactions			
Pain	Does not interfere with activity	Repeated use of non-narcotic pain reliever >24 hours or interferes with activity	Any use of narcotic pain reliever or prevents daily activity
Tenderness	Discomfort only to touch	Discomfort with movement	Significant discomfort at rest
Erythema/redness [b]	2.5-5 cm	5.1-10 cm	>10 cm
Induration/swelling [c]	2.5-5 cm and does not interfere with activity	5.1-10 cm or interferes with activity	>10 cm or prevents daily activity
Systemic reactions			
Allergic reaction	Pruritus without rash	Localized urticaria	Generalized urticaria; angioedema or anaphylaxis
Headache	No interference with activity	Repeated use of non-narcotic pain reliever >24 hours or some interference with activity	Significant; any use of narcotic pain reliever or prevents daily activity
Fatigue	No interference with activity	Some interference with activity	Significant; prevents daily activity

[b] In addition to grading the measured local reaction at the greatest single diameter, the measurement should be recorded as a continuous variable.
[c] Induration/swelling should be evaluated and graded using the functional scale as well as the actual measurement.

Reactogenicity	Mild (Grade 1)	Moderate (Grade 2)	Severe (Grade 3)
Myalgia	No interference with activity	Some interference with activity	Significant; prevents daily activity
All other conditions	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Illness or clinical adverse event (as defined according to applicable regulations)	No interference with activity	Some interference with activity not requiring medical intervention	Prevents daily activity and requires medical intervention

LABORATORY AND VITAL SIGNS TOXICITY GRADING (Some laboratory values have been modified to be consistent with the normal ranges of the ACM laboratory used in the present study)

Blood, Serum, or Plasma Chemistries [a]	LO/HI/N [b]	Mild (Grade 1) [c]	Moderate (Grade 2)	Severe (Grade 3)
Sodium (mEq/L or mmol/L)	LO	131-<LLN	130	<130
	HI	>ULN-148	149-150	>150
Potassium (mEq/L or mmol/L)	LO	<LLN-3.2	<3.2-3.1	<3.1
	HI	>ULN-5.6	>5.6-5.7	>5.7
Glucose (mg/dL)	LO mmol/L	<LLN-3.0	<3.0-2.2	<2.2
	HI mmol/L	>ULN-8.9	>8.9-13.9	>13.9
Blood urea nitrogen	HI mmol/L	>8.9-17.8	>17.8-35.5	>35.5
Creatinine	N	115-151 (µmol/L)	152-177 (µmol/L)	>177 (µmol/L)
Calcium **(CTCAE 4.0)**	LO mmol/L	<LLN-2.0	<2.0-1.75	<1.75
	HI mmol/L	>ULN-2.9	>2.9-3.1	>3.1
Magnesium **(CTCAE 4.0)**	LO mmol/L	<LLN-0.5	<0.5-0.4	<0.4
Phosphorous **(CTCAE 4.0)**	LO mmol/L	<LLN-0.8	<0.8-0.6	<0.6
Creatine kinase (CPK or CK) **(CTCAE 4.0)**	HI	>ULN-2.5xULN	>2.5xULN-5xULN	>5xULN
Albumin	LO g/L	<30-28	<28-25	<25
Total protein	LO g/L	<LLN-52	<52-50	<50
Alkaline phosphatase (U/L) **(CTCAE 4.0)**	HI	>ULN-2.5xULN	>2.5xULN-5xULN	>5xULN
AST (U/L) **(CTCAE 4.0)**	HI	>ULN-3xULN	>3xULN-5xULN	>5xULN
ALT (U/L) **(CTCAE 4.0)**	HI	>ULN-3xULN	>3xULN-5xULN	>5xULN

Blood, Serum, or Plasma Chemistries [a]	LO/HI/N [b]	Mild (Grade 1) [c]	Moderate (Grade 2)	Severe (Grade 3)
Bilirubin, serum total (mmol/L) **(CTCAE 4.0)**	HI mmol/L	>ULN-1.5xULN	>1.5xULN-3xULN	>3xULN
Bilirubin, serum total (mg/dL) when ALT ≥105 (Hy's law)	HI	1.3-1.5	1.6-2.0	>2.0
Bilirubin, serum direct (mmol/L) (CTCAE 4.0)	HI mmol/L	>ULN-1.5xULN	>1.5xULN-3xULN	>3xULN
Amylase (U/L) **(CTCAE 4.0)**	HI	>ULN-1.5xULN	>1.5xULN-2xULN	>2xULN
Lipase (U/L) **(CTCAE 4.0)**	HI	>ULN-1.5xULN	>1.5xULN-2xULN	>2xULN
Uric acid (mg/dL/mmol/L) **(CTCAE 4.0)**	HI	>ULN – 10 mg/dL (0.59 mmol/L) without physiologic consequences	-	>ULN – 10 mg/dL (0.59 mmol/L) with physiologic consequences

[a] Depending upon the laboratory used, references ranges, eligibility ranges and grading may be split out by sex and/or age.
[b] Low, High, Not Graded (N).
[c] If initial boundary of grade 1 has gap from reference range or eligibility range, calculations based on the New England Journal of Medicine (NEJM) reference ranges.

Hematology	LO/HI/N [a]	Mild (Grade 1) [b]	Moderate (Grade 2)	Severe (Grade 3)
Hemoglobin (women) (g/dL)	LO	10.8-11.3	9.2-10.7	<9.2
Hemoglobin (men) (g/dL)	LO	12.0-12.5	10.0-11.9	<10.0
White blood cell count (K/CUMM)	HI	11.00-15.00	15.00-20.00	>20.00
	LO	2.50-3.50	1.50-2.49	<1.50
Lymphocytes (K/CUMM)	LO	0.76-0.90	0.50-0.75	<0.5
Neutrophils (K/CUMM)	LO	1.50-1.95	1.00-1.49	<1.00
Eosinophils (K/CUMM)	HI	0.58-0.74	0.75-1.50	>1.50
Platelets (K/CUMM)	LO	120-130	100-120	<100
Coagulation				
Prothrombin time (PT, seconds)	HI	> ULN-14.4	14.5-15.7	>15.7
Partial thromboplastin time (PTT or aPTT, seconds)	HI	>ULN-42.1	42.2-50.0	>50.0
Fibrinogen (mg/dL) **(CTCAE 4.0)**	HI	>ULN-500	501-600	>600
	LO	<LLN-0.75xLLN	<0.75xLLN-0.5xLLN	<0.5xLLN
Urine				
Protein (dipstick)	HI	1+	2+	>2+
Glucose (dipstick)	HI	1+	2+	>2+
Blood (microscopic) - red blood cells per high power field (RBC/HPF)	HI	5-10 for males 9-10 for females	11-50	>50 and/or gross blood

[a] Low, High, Not Graded.
[b] If initial boundary of grade 1 has gap from reference range or eligibility range, calculations based on the New England Journal of Medicine (NEJM) reference ranges.

Vital Signs	LO/HI/N [a]	Mild (Grade 1) [b]	Moderate (Grade 2)	Severe (Grade 3)
Fever (°C) [c]	HI	38.0-38.4	38.5-38.9	>38.9
Fever (°F)	HI	100.4-101.1	101.2-102.0	>102.1
Tachycardia - beats per minute	HI	101-115	116-130	>130 or ventricular dysrhythmias
Bradycardia - beats per minute	LO	40-45	35-40	<35
Hypertension (systolic) - mm Hg [d]	HI	141-150	151-160	>160
Hypertension (diastolic) - mm Hg	HI	91-95	96-100	>100
Hypotension (systolic) - mm Hg	LO	85-89	80-84	<80
Tachypnea - breaths per minute	HI	23-25	26-30	>30

[a] Low, High, Not Graded.
[b] If initial boundary of grade 1 has gap from reference range or eligibility range, calculations based on the New England Journal of Medicine (NEJM) reference ranges.
[c] Oral temperature; no recent hot or cold beverages or smoking. A protocol should select either °C or °F for inclusion.
[d] Assuming subject is awake, resting, and supine; for adverse events, 3 measurements on the same arm with concordant results.

[redacted]

Statistical Analysis Plan

Study Title:	**A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis**
Protocol Number and Version:	**ARQ-151-311, Original dated 31 October 2020** **Amendment 1 dated 08 June 2021** **Amendment 2 dated 18 July 2021** **Amendment 3 dated 23 May 2022** **Amendment 4 dated 28 September 2022**
Product:	**ARQ-151 Cream 0.15%**
Sponsor:	**Arcutis Biotherapeutics, Inc.** **3027 Townsgate Road, Suite 300** **Westlake Village, CA 91361**
Date:	**25 October 2022**
Version:	**Final 3.0**
Prepared by:	[redacted]

STATISTICAL ANALYSIS PLAN REVISION SUMMARY			
Version	**Version Date**	**Author**	**Summary of Changes**
Original V1.0	05-JUL-2022	[redacted]	Initial version
Final V2.0	11-Oct-2022	[redacted]	Primary estimand and multiple imputation procedure were updated as per FDA's request discussed at the Type B Pre-NDA meeting on [redacted] for a similar program. A baseline definition was added for randomized subjects who never received any study treatment (if any). Analysis day, and number of days in study derivations were updated accordingly.
Final V3.0	25-Oct-2022	[redacted]	Clarify that for subjects who discontinued early from the study due to adverse event or lack of efficacy, a subject will be considered as non-responder (for MI and non-responder imputation analyses) or missing (for observed case analyses) for all analysis visits (refer to Section 5.4) for which the subject's last dose day falls within the analysis visit window or is prior to the start of the analysis visit windows.

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(S):

Name	Title	Signature	Date (DD-MMM-YYYY)
[redacted]	Principal Biostatistician [redacted]		

APPROVALS:

Name	Title	Signature	Date (DD-MMM-YYYY)
[redacted]	Vice President, Biometrics [redacted]		
[redacted]	Vice President, Biometrics Arcutis Biotherapeutics, Inc.		
[redacted]	Vice President, Clinical Development Arcutis Biotherapeutics, Inc.		

TABLE OF CONTENT

ABBREVIATIONS

AD	Atopic Dermatitis
AE	Adverse Event
ANCOVA	Analysis of Covariance
ATC	Anatomical Therapeutic Chemical
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
EASI	Eczema Area and Severity Index
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EDC	Electronic Data Capture
EM	Expectation-Maximization
ET	Early Termination
HR	Heart Rate
IP	Investigational Product
ITT	Intent to Treat
IWRS	Interactive Web Response System
LOCF	Last Observation Carried Forward
MCMC	Markov-Chain Monte-Carlo
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent to Treat
PHQ-8	Patient Health Questionnaire-8
PHQ-A	Modified PHQ-9 for Adolescents
PMM	Predictive Mean Matching
POEM	Patient-Oriented Eczema Measure
PK	Pharmacokinetic
PP	Per Protocol
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
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-151-311. The analyses described in the SAP are based upon the protocol Amendment 4 dated 28 September 2022. In case of changes (note that any such changes are described in section 3.5 below) between the protocol and the SAP, the SAP will be used to guide the statistical analysis. Any deviations from the SAP will be described and justified in the final Clinical Study Report (CSR), as appropriate.

On [redacted] FDA provided the following advice in preliminary feedback prior to a pre-NDA meeting held with Arcutis for a similar program on [redacted]

> *For the primary estimand, you [Arcutis] proposed a treatment policy strategy to handle all intercurrent events, including treatment discontinuation due to adverse event or lack of efficacy. For intercurrent events of treatment discontinuation due to adverse events or lack of efficacy, we [FDA] recommend a composite strategy policy where subjects will be defined as non-responders, as we consider this to be the appropriate approach for handling such events. Your proposal to handle intercurrent events of treatment discontinuation using the treatment policy strategy can be used as part of a supportive estimand.*

The purpose of this version of the SAP is to document the change to the primary estimand and related multiple imputation strategies as requested by the FDA for a similar program.

This SAP has been developed prior to database lock, unblinding, and associated analyses. All final analyses will be performed after approval of the SAP, the clinical trial data are entered into the database, any discrepancies in the data are resolved, determination of the inclusion/exclusion of each subject from each analysis population, the database is locked, and the unblinding request form is signed.

2 STUDY OBJECTIVES AND ENDPOINTS

OBJECTIVES	ENDPOINTS
Efficacy	
To assess the efficacy of Roflumilast (also known as ARQ-151) cream 0.15% vs vehicle administered once daily	Primary efficacy endpoint:
	• Validated Investigator Global Assessment scale for Atopic Dermatitis (vIGA-AD) Success, defined as a vIGA-AD score of 'clear' (0) or 'almost clear' (1)

OBJECTIVES	ENDPOINTS
(QD) x 4 weeks to individuals 6 years of age and older with atopic dermatitis (AD).	plus at least a 2-grade improvement from Baseline at Week 4
	Secondary efficacy endpoints:
	• In subjects with a vIGA-AD score of 'Moderate' at randomization, vIGA-AD Success at Week 4 • In subjects ≥12 years old with baseline Worst Itch - Numeric Rating Score (WI-NRS) ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 4 • In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 2 • In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 1 • Achievement of at least a 75% reduction in the Eczema Area and Severity Index (EASI-75) at Week 4 • vIGA-AD of 'clear' or 'almost clear' at Week 4 • vIGA-AD Success at Week 2 • vIGA-AD Success at Week 1 • vIGA-AD of 'clear' or 'almost clear' at Week 2 • vIGA-AD of 'clear' or 'almost clear' at Week 1
	Exploratory efficacy endpoints:
	• Other continuous efficacy endpoints include change and percent change in average weekly WI-NRS and daily WI-NRS, EASI, % body surface area (BSA) affected by AD, Dermatology Life Quality Index (CDLQI/DLQI), the Dermatitis Family Impact (DFI), Scoring Atopic Dermatitis (SCORAD), and Patient-oriented Eczema Measure (POEM) at Week 1, Week 2 and Week 4. • Other categorical efficacy endpoints include EASI-50, EASI-90, EASI-100, and vIGA-AD of 'clear' at Week 4
Safety	

OBJECTIVES	ENDPOINTS
To assess the safety of Roflumilast cream 0.15% vs vehicle administered once daily (QD) x 4 weeks to individuals 6 years of age and older with AD	• Incidence and severity of adverse events (AEs) • Changes and percent change in clinical laboratory results • Changes and percent change in vital signs • The subject incidence of >5% weight loss or gain on study • Local tolerability assessments • Patient Health Questionnaire depression scale (PHQ-8) and Modified PHQ-9 for Adolescents (PHQ-A) • Children's Depression Inventory 2nd Edition (CDI-2) • Columbia-Suicide Severity Rating Scale (C-SSRS)
Pharmacokinetic	
To assess the systemic exposure of roflumilast and its N-oxide metabolite following Roflumilast cream 0.15% QD application x 4 weeks	• Plasma concentrations of roflumilast and its N-oxide metabolite

3 STUDY DESIGN

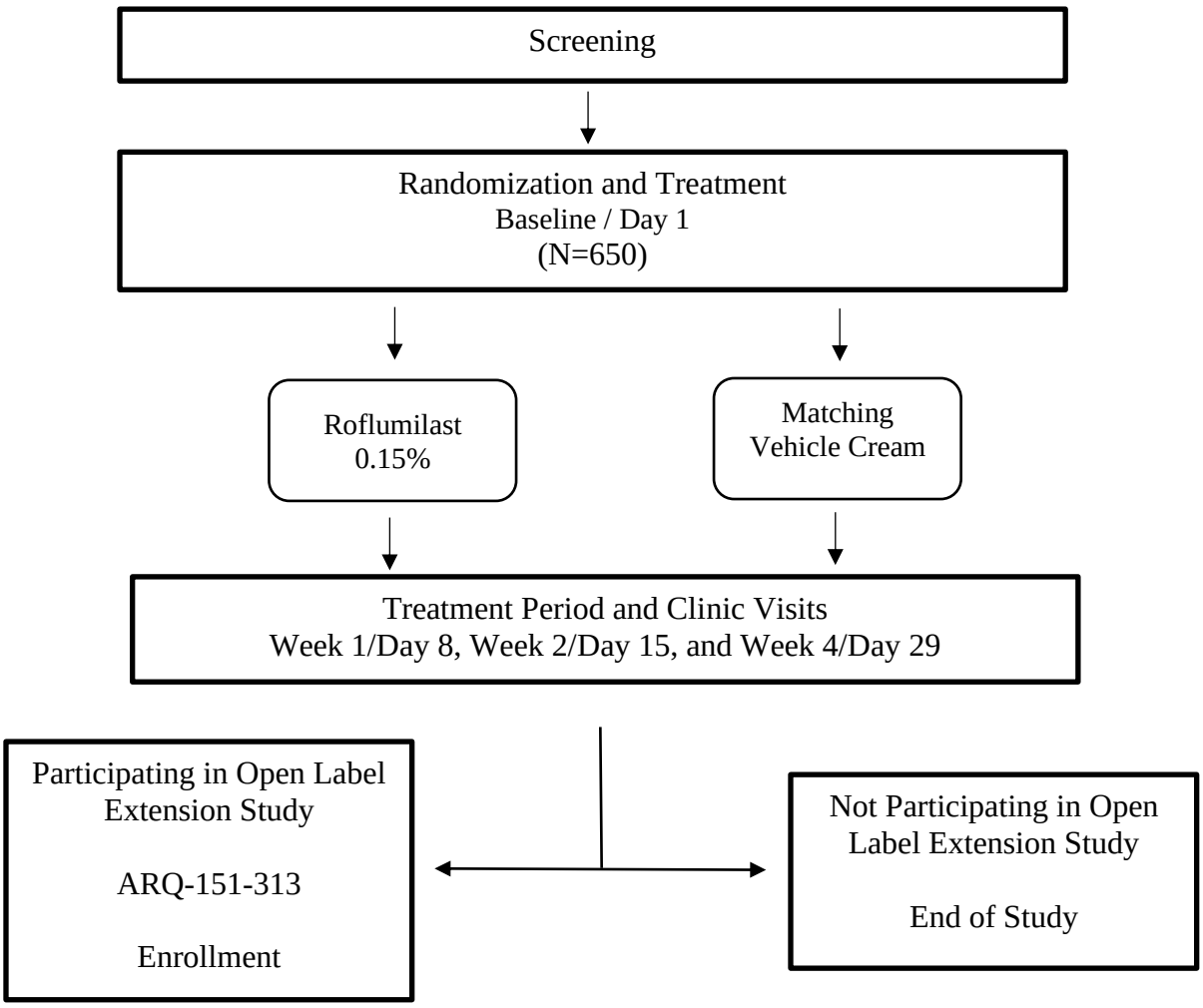
3.1 Overall Design

This is a Phase 3, parallel group, double blind, vehicle-controlled study in which Roflumilast cream 0.15% or vehicle is applied QD x 4 weeks to subjects with mild to moderate atopic dermatitis.

- Upon determination of eligibility, subjects will be randomized 2:1 to either Roflumilast cream 0.15% or matching vehicle cream. The randomization will be stratified by vIGA-AD score at baseline ('Mild' vs. 'Moderate') and by study site.
- At the Week 4 visit, subjects may be eligible to enroll in an open label extension study (ARQ-151-313) in which they will receive Roflumilast cream 0.15% QD.

The trial design is represented schematically in Figure 1.

Figure 1 **Study Schema**



A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis

Approximately 650 subjects with atopic dermatitis will be randomized 2:1 to receive either:

- Roflumilast cream 0.15% or Vehicle cream

Subjects will have ≥3% BSA involvement (excluding the scalp, palms, soles) with a vIGA-AD score of '2' (mild) or '3' (moderate) for study entry

Up to 50% of the subjects will be ≥ 18 years old

3.2 Schedule of Events

Table 1 Schedule of Visits and Assessments

Study Procedure	Screen	Baseline Day 1	Wk 1 Day 8	Wk 2 Day 15	Wk 4 Day 29 / ET
Visit	**1**	**2**	**3**	**4**	**5**
Visit Window	**-30 days**	**N/A**	**+/- 3 days**	**+/- 3 days**	**+/- 3 days**
Informed consent/assent	X				
Demographics	X				
Medical and surgical history	X				
Physical examination[a]	X	X			X
I/E criteria	X	X			
Hematology, Serum Chemistries, and Urine Analysis[b]	X[b]	X[b]			X[b]
Vital signs, height, weight[c]	X	X	X	X	X
vIGA-AD, EASI, BSA, SCORAD[d]	X	X	X	X	X
WI-NRS pruritus[e]	X	X	X	X	X
POEM[f]	X	X	X	X	X
Local Tolerability Assessment[g]		X	X	X	X
CDI-2, PHQ-8, PHQ-A, C-SSRS[h]	X	X	X	X	X
DLQI, CDLQI, DFI[i]	X	X	X	X	X
Medical Photography[j]		X	X		X
Serum pregnancy test (FOCBP only)	X				
Urine pregnancy test[k]		X	X	X	X
PK draws[l]					X
Drug/vehicle application in clinic[m]		X	X	X	
Dispense/Re-dispense study medication kit[n]		X	X[o]	X[o]	X[o]
Dispense/review diary	X	X	X	X	X
Weigh study medication kit[p]		X	X	X	X
Compliance determination[q]			X	X	X

Study Procedure	Screen	Baseline Day 1	Wk 1 Day 8	Wk 2 Day 15	Wk 4 Day 29 / ET
Adverse event assessment[r]	X	X	X	X	X
Concomitant medications	X	X	X	X	X
Study Exit [s]					X

a Limited physical examination: skin (including assessment of Fitzpatrick skin type at Screening only), lungs, and heart only

[b] For all subjects entering this study under Amendment 2, to be collected at Screening, but subsequent samples will be collected only for subjects ≥12 years old (Baseline/Day 1 and Week 4/Day 29/ET). For subjects 12 to 18 years of age, if Baseline/Day 1 is within 3 weeks of Screening, the Screening results may be used.

[c] Height will be collected at Screening only. 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 other 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. **The vIGA-AD assessment should be completed prior to other physician assessments**. SCORAD total score will range between 0 and 103.

[e] Subjects will self-assess their pruritus at home on a daily basis starting 7 days prior to the Baseline/Day 1 visit, and then every day thereafter. WI-NRS score will be determined by the subject assessing worst itch over the past 24 hours. The scale is from 0 (no itch) to 10 (worst itch) and this value will be recorded by the subject each day. Subjects will be trained at the Screening visit in the accurate completion of the WI-NRS. In addition, parents/caregivers of children and adolescent subjects will be trained at the Screening visit by study staff on how to assist the subject, if needed, in completing the WI-NRS.

[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] Local tolerability assessments should be recorded prior to study drug application for the Investigator assessment of skin irritation (Berger and Bowman skin irritation score). **Note for investigator tolerability assessments: reactions at the site of product application, which may occur post-Baseline, should be differentiated from the preexisting inflammation associated with the subject's atopic dermatitis.** The subject will assess burning/stinging (0-3 score) 10-15 minutes post drug application. **Note subject burning stinging assessment: at Day 29, subjects will provide a recall assessment of burning/stinging experienced post drug application on the previous day (Day 28).**

[h] Adolescents and adults will complete the C-SSRS (12 years of age and older). Adults will complete the PHQ-8. Adolescents (ages 12 to 17, inclusive) will complete the PHQ-A (PHQ-9 modified). Parents/caregivers will complete CDI-2 (parent report) for children 6-11 years of age, inclusive.

[i] The DLQI will be completed by subjects ≥17 years of age. The CDLQI will be completed for subjects 6 to 16 years old, inclusive. The Dermatitis Family Impact Questionnaire (DFI) will be completed by parents/caregivers for all subjects 6 to ≤17 years of age.

[j] Photography of AD lesion(s) selected by the Investigator will be performed at all investigational sites. All efforts will be made to de-identify the subjects. Canfield equipment will be used to capture photographs. Subjects who are unwilling to participate in the medical photography will be allowed to opt out of this procedure, as documented on the Informed Consent Form.

[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 entering this study under Amendment 2, a single PK trough draw will be collected at Day 29 only for subjects ≥12 years old. Ensure study medication was not applied in the area where PK will be drawn.

[m] Subjects to apply assigned IP during clinic visits, except for the Day 29/ET visit.

[n] It is expected that kits will be dispensed based on %BSA affected. See IP Handling Manual for details.

[o] On Day 8 and 15, dispensing of IP is optional. Site should review IP kit to ensure sufficient IP is available until the next visit and only dispense additional IP if needed. On Day 29, if the subject is unable to perform the Day 29 clinic visit due to COVID-19 restrictions (isolation, quarantine, etc.) then additional IP may need to be dispensed so IP can continue to be applied at home until the subject is able to return to the clinic to complete the Day 29 assessments (see IP Handling Manual for the process to dispense additional IP at or after Day 29).

[p] Every tube should be weighed and recorded when dispensed and returned. See IP Handling Manual for details.

[q] Compliance determination is described in the IP Handling Manual

[r] All AEs should be collected starting after the first application of the investigational product through the end of the study. All SAEs should be collected starting after the signing of the informed consent through 30 days after the last day of the application of the investigational product or the end of the study (whichever is later). Any AEs (whether serious or non-serious) and clinically abnormal laboratory test values(s) will be evaluated by the PI and treated and/or followed up for up to 30 days after end of treatment or until symptoms or value(s) return to normal, or acceptable level, as judged by the PI (if the subject is continuing into the ARQ-151-313 OLE study, then AEs from this study (ARQ-151-311) will only be followed until exit from this study).

[s] Subjects who enroll into the open label extension study (ARQ-151-313) must complete the ARQ-151-311 visit requirements at Week 4.

3.3 Treatment

Roflumilast cream 0.15% or vehicle cream will be administered QD for 28 days (+/- 3 days).

- Roflumilast cream 0.15%
- Vehicle cream

3.4 Randomization, Replacement, and Unblinding Procedures

Randomization will take place at the Baseline visit prior to first dosing. Subjects who meet all eligibility criteria will be randomized at a 2:1 ratio (drug:vehicle) to receive Roflumilast cream 0.15% QD or matching vehicle QD. The randomization will be stratified by vIGA-AD score at baseline ('Mild' vs. 'Moderate') and by study site according to a computer-generated randomization list. Kits containing tubes of study medication will be assigned to each subject using an internet-based response system (IWRS). A subject may receive more than one kit for the treatment period. The kits and tubes are blinded and each kit is numbered with a unique kit number.

The study is double-blinded, therefore neither the subjects nor the Investigator, sponsor and clinical personnel will be aware of which treatment an individual subject receives.

3.5 Changes to the Analysis from the Protocol

Rationale for change	Description of the change
A categorical summary 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.
To separate COVID patient from Per Protocol analysis set.	Original text: Per protocol (PP) population will include all subjects in the ITT population, who are at least 80% compliant with study medication application, and show no major deviations from the study protocol that would affect the interpretation of efficacy. In addition, subjects who miss the Week 4 vIGA-AD assessment specifically due to novel coronavirus disease-19 (COVID-19) disruptions will be excluded from per protocol population. Changed to: Per protocol population will include all subjects who are randomized, at least 80% compliant with study medication application, have a vIGA-AD assessment within the Week 4 visit window, and show no "major deviations" from the study protocol that would affect the interpretation of efficacy.
Added mITT population to address COVID-19 impact.	The mITT population includes all randomized subjects with the exception of subjects who missed the week 4 vIGA-AD assessment specifically due to COVID-19 disruption. This population will be used for sensitivity analysis for the primary endpoint.

4 POPULATIONS FOR ANALYSIS

4.1 Intent-to-Treat (ITT) Population

Intent-to-Treat population will include all subjects who are randomized, and all subjects will be analyzed according to the treatment group and stratum to which they were randomized.

4.2 Modified Intent-to-Treat (mITT) Population

The mITT population includes all randomized subjects with the exception of subjects who missed the week 4 vIGA-AD assessment specifically due to COVID-19 disruption. This population will be used for sensitivity analysis for the primary endpoint. All subjects will be analyzed according to the treatment group and stratum to which they were randomized.

4.3 Per Protocol (PP) Population

Per protocol population will include all subjects who are randomized, at least 80% compliant with study medication application, have a vIGA-AD assessment within the Week 4 visit window and show no "major deviations" from the study protocol that would affect the interpretation of efficacy. A complete list of major deviations from the study protocol will be created prior to unblinding and include a list of all subjects who will be excluded due to those major deviations. See section 7.2 for more details.

All subjects will be analyzed according to the actual treatment group they received and the stratum they belong to. Actual and randomized treatment will only differ if the subject received the wrong treatment throughout their participation in the study.

4.4 vIGA-AD Moderate ITT Population

vIGA-AD Moderate ITT population will be a subset of the ITT population with vIGA-AD score (randomized score) 'Moderate' at randomization.

All subjects will be analyzed according to the treatment group and the stratum to which they were randomized.

4.5 vIGA-AD Moderate PP Population

vIGA-AD Moderate PP population will be a subset of the PP population with vIGA-AD score (actual score) 'Moderate' at randomization.

All subjects will be analyzed according to the actual treatment group they received and the stratum they belong to. Actual and randomized treatment will only differ if the subject received the wrong treatment throughout their participation in the study.

4.6 WI-NRS Population

WI-NRS population will be a subset of the ITT population among subjects ≥12 years old with average weekly baseline WI-NRS score ≥ 4.

The subjects in WI-NRS population are those:

1. Completed at least 4 of 7 evaluable daily WI-NRS questionnaires during the last 7 days of the Screening period;
2. Have a mean baseline WI-NRS score ≥ 4.0, defined as the average of all non-missing scores reported during the last 7 days of the Screening period if at least 4 of 7 evaluable daily WINRS questionnaires available. If 4 or more evaluable daily questionnaires are missing, then the data will be treated as missing.

All subjects will be analyzed according to the treatment group and stratum to which they were randomized.

4.7 Safety Population

Safety population will include all subjects who are enrolled and received at least one confirmed dose of study medication.

Subjects will be analyzed based on the treatment group received and the stratum they belong to. Actual and randomized treatment will only differ if the subject received the wrong treatment throughout their participation in the study.

4.8 Pharmacokinetic (PK) Population

Pharmacokinetic population will include all subjects receiving the active drug with quantifiable plasma concentrations of roflumilast.

5 GENERAL CONSIDERATIONS

Formats and layouts of tables, listings, and figures (TLF) will be provided in a separate document (output general layout is described in Appendices

Appendix **1**).

5.1 Sample Size

There are approximately 650 subjects planned for this study. In order to have the desired power for the secondary endpoint of vIGA-AD Success in subjects with a vIGA-AD score of 'Moderate' at randomization, approximately 490 of the subjects to be accrued will have vIGA-AD score of 'Moderate' at randomization. Randomization will be stratified by vIGA-AD score ('Mild' vs. 'Moderate') and by study site.

The sample size of 650 subjects provides approximately 95% power to detect an overall 15% difference (Odds Ratio = 2.1) between treatment groups on vIGA-AD success at Week 4 at α=0.05 using a 2-sided stratified Cochran-Mantel-Haenszel (CMH) test. The results from a recent Phase 2 study (ARQ-151-212) of Roflumilast cream 0.15% compared to vehicle treatment were used to estimate the treatment difference. Specifically, in the Phase 2 trial, approximately 37% of subjects demonstrated vIGA-AD Success at Week 4 in the ARQ-151 0.15% group compared to 22% in the vehicle group. The sample size of 490 also provides approximately 90% power to detect an overall 17% difference (Odds Ratio = 2.1) between treatment groups (28% of vIGA-AD Success at Week 4 in vehicle treatment) on vIGA-AD Success at Week 4 among subjects with vIGA-AD score 'moderate' at randomization. The same testing method, the stratified CMH test, will be used as for the primary endpoint.

To control for familywise type I error at level of 0.05, the secondary endpoint of vIGA-AD Success at Week 4 in subjects with vIGA-AD of 'Moderate' at randomization will only be tested if the primary endpoint demonstrates statistical significance. In addition, the remaining secondary endpoints will be inferentially tested only if the primary and secondary endpoint (vIGA-AD Success at Week 4 for subjects with vIGA-AD score of 'Moderate' at randomization) comparisons are statistically significant using the hierarchical testing procedure by partitioning of alpha (see Section 6.5 for more details).

5.2 Baseline

Unless otherwise specified, baseline value will be defined as the last non-missing assessment prior to or concurrently with the first study treatment dosing* (including unscheduled/retest assessments). If the last non-missing assessment is performed on the same date as the first study treatment administration* and time is not available, it is assumed that the assessment took place prior to IP application*, per study site training, and the assessment will be considered as baseline, except for adverse events (AEs) and medications starting on the first study treatment dose administration date which will be considered postbaseline.

Average weekly baseline WI-NRS is defined as the average of all non-missing scores reported during the last 7 days prior to treatment* if at least 4 of 7 evaluable daily WI-NRS questionnaires

available. Daily baseline WI-NRS is defined as the last non-missing assessment prior to or concurrently with the first study treatment dosing*. Day 1 WI-NRS score will be used to calculate the value for week 1 only when it is collected after the application of the first study drug*. If Day 1 WI-NRS score is collected prior to or concurrently with the application of the first study drug*, then the Day 1 WI-NRS score will be included in baseline calculation.

For investigator/subject tolerability assessments, baseline is derived as the measurement taken on the day of first application of study drug*.

* For randomized subjects who discontinued study before the first application of study treatment, the date of randomization will be considered instead of the date of the first application of study treatment.

5.3 Reference Start Date and Analysis Day

Analysis day will be calculated from the first study treatment administration date* and will be used to derive start/end day of assessments or events.

Analysis day = (Date of event – Date of first dose administration*) + 1 if date of event is on or after the date of first dose administration of study treatment*;

= (Date of event – Date of first dose administration*) if date of event is before the date of first dose administration of study treatment*.

In the situation where the assessment/event date is partial or missing, analysis day will be missing.

* For randomized subjects who discontinued study before the first application of study treatment, the date of randomization will be considered instead of the date of the first application of study treatment.

5.4 Windowing Conventions

Visits will be analysed as scheduled. Unscheduled, early termination visits, and/or retest measurements will only be included if a scheduled measurement is not available and the early termination or unscheduled/retest measurement falls within the analysis visit windows as described in Table 2, Table 3, and Table 4 when appropriate. Unscheduled/retest measurements will be listed.

If there is more than one assessment for a given timepoint and analysis visit when a scheduled measurement is not available, the assessment closest to the target day will be considered.

Table 2 Analysis Visit Windows for Efficacy Endpoints, Vital Signs, CDI-2, PHQ-8, PHQ-A, C-SSRS and Local Tolerability Assessment

Analysis Visit	Target Day	Lower Limit	Upper Limit
Week 1	8	2	11
Week 2	15	12	22
Week 4	29	23	43

Table 3 Analysis Visit Windows for Clinical Laboratory, Physical Examination and PK assessment

Analysis Visit	Target Day	Lower Limit	Upper Limit
Week 4	29	23	43

Table 4 Windows for the derivation of Average Weekly WI-NRS

Days for calculation of weekly average	Week (Derived)
(-7, -1)*	Baseline
(1, 7)*	Week 1
(8, 14)	Week 2
(15, 21)	Week 3
(22, 28)	Week 4

* Day 1 WI-NRS score will be used to calculate the value for week 1 only when it is collected after the application of the first study drug (randomization for randomized subjects who were never treated with study drug). If Day 1 WI-NRS score is collected prior to or concurrently with the application of the first study drug (randomization for randomized subjects who were never treated with study drug), then the Day 1 WI-NRS score will be included in baseline calculation.

Note: With the caveat described in footnote '*', if more than one WI-NRS score is available on the same day, the worst score of the day will be considered in the analyses.

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 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 MI and non-responder imputation analyses) or missing (for observed case analyses) for all pre-specified analysis visits (refer to Section 5.4) for which the subject's last dose day falls within analysis visit window or is prior to the start of the analysis visit window.

- Average weekly WI-NRS = Average weekly WI-NRS pruritus score will be calculated as the sum of the daily WI-NRS scores reported during a specific week (in a 7-day period; refer to Section 5.4) of the study divided by the number of days with non-missing scores for that week. A minimum of 4 days of observations are needed to calculate an average weekly WI-NRS pruritus score. Otherwise, the corresponding average weekly WI-NRS pruritus score will be considered missing.

- With the following exception, a WI-NRS 4-point reduction = achievement of a 4- point reduction in average weekly WI-NRS pruritus score compared to average weekly WI-NRS baseline, calculated only for the subjects ≥12 years old with average weekly WI-NRS score of ≥ 4 at 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 WI-NRS 4-point reduction (for MI and non-responder imputation analyses) or missing (for observed case analyses) for all pre-specified analysis visits (refer to Section 5.4) for which the subject's last dose day falls within the analysis visit window or is prior to the start of the analysis visit window.

- EASI total score = $0.1 (E_h + I_h + E_{Xh} + L_h) A_h + 0.2 (E_u + I_u + E_{Xu} + L_u) A_u + 0.3 (E_t + I_t + E_{Xt} + L_t) A_t + 0.4 (E_l + I_l + E_{Xl} + L_l) A_l$
  for subjects '≥ 8 years old
  and
  $0.2 (E_h + I_h + E_{Xh} + L_h) A_h + 0.2 (E_u + I_u + E_{Xu} + L_u) A_u + 0.3 (E_t + I_t + E_{Xt} + L_t) A_t + 0.3 (E_l + I_l + E_{Xl} + L_l) A_l$
  for subjects < 8 years old

  where E, I, Ex, 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. Scalp, palms, and soles may be treated with investigational product in this study but will be excluded from the EASI assessment. If a subject turns 8 years old during the study, the formula used at Screening will continue to be used through the study duration of the subject's participation in the study.

- With the following exception, EASI-50, EASI-75, EASI-90, and EASI-100 = Achievement of at least a 50%, 75%, 90%, or 100% reduction from baseline in EASI total score, respectively. 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 reached EASI-50, EASI-75, EASI-90 and EASI-100 (for MI and non-responder imputation analyses) or missing (for observed case analyses) for all pre-specified analysis visits (refer to Section 5.4) for which the subject's last dose day falls within the analysis visit window or is prior to the start of the analysis visit window.

- DLQI Score = sum of the 10 questions (individual questions scored as Very much=3, A lot=2, A little=1, Not at all=0, Not relevant=0, Question 7: Yes=3, 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.

- CDLQI Score = sum of the 10 questions (individual questions scored as Very much=3, Quite a lot=2, Only a little=1, Not at all=0; Question 7: if the last week was school time, the question was scored as Very much=3, Quite a lot=2, Only a little=1, Not at all=0, with Prevented school recoded to 3, and if the last week was 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.

- DFI 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.

- PHQ-8 = sum of the 8 questions (individual questions scored as Not at all=0, Several days=1, More than half the days=2, and Nearly every day=3), 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 (=7).

- Modified PHQ-A = sum of the 8 questions (individual questions scored as Not at all=0, Several days=1, More than half the days=2, and Nearly every day=3), 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 (=7).

- CDI-2 total score is a sum of the 17 questions based on the scoring grid (individual questions scored as much or most of the time=0, often=1, some of the time=2, Not at all=3 for questions Q2, Q7, Q13, Q14, and Q16; individual questions scored as much or most of the time=3, often=2, some of the time=1, Not at all=0 for all other questions), ranging from

0 to 51. All CDI-2 raw scores can be prorated using the following formula, rounding to the nearest whole number:

- Prorated score = (Obtained raw score for scale) * (Total # of items on scale/subscale) / Total # items on scale/subscale with responses
- CDI-2 emotional problem scale is a sum of 9 questions (Q1, Q3-6, Q8, Q10-12). 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*9)/number of answered items (=8).
- CDI-2 functional problem scale is a sum of 8 questions (Q2, Q7, Q9, Q13-17). 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 (=7).
- CDI-2 total score is a sum of 17 questions. If more than 2 items are missing the score should not be calculated. If 1 item is missing the score is calculated as (sum of answered items*17)/number of answered items (=16). If 2 items are missing the score is calculated as (sum of answered items*17)/number of answered items (=15).

- POEM = 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 is scored 0 and the scores are summarized and expressed as usual out of a maximum of 28. If 2 or more questions are left unanswered the questionnaire is not scored.
- SCORAD = [Overall BSA affected by AD / 5] + [Intensity score*7/2] + subjective symptoms score (pruritus + sleep loss); SCORAD score will be set to missing if information for any of the three measures is missing.

5.6 Descriptive Statistics

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

Categorical variables will be presented as frequencies and percentages.

Summary tables will be presented by visit, when applicable.

Change from baseline will be calculated as:

Assessment value at postbaseline visit X – baseline value.

Percent change from baseline will be calculated as:

(Assessment value at postbaseline visit X – baseline value) × 100% / (baseline value)

Percent change from baseline will be missing in situation where baseline value equals to 0.

5.7 Statistical Tests

Unless otherwise specified, all statistical tests will be two-sided and will be performed with a significant level of 0.05. Confidence intervals (CIs) will be two-sided with 95% coverage.

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.

All data from retest, unscheduled measurements and ET visit assessments will be listed.

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

Covariates for this study include pooled study site and baseline vIGA-AD (vIGA-AD=2 - Mild vs. vIGA-AD=3 - Moderate at randomization). Subgroup analyses will be generated for the baseline covariates.

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

All subjects, regardless of completion status, will have available data assigned to the pre-specified analysis visit using the analysis windows defined in Section 5.4, including the last available data

of subjects who prematurely withdraws from the study. To comply with the definition of the primary estimand (Section 12.1.1), for subjects who discontinue due to lack of efficacy or adverse event, 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. Similarly, WI-NRS weekly averages will be removed from the source data used for multiple imputation if subject's last dose day falls within the analysis visit window or is prior to the start of the interval used to compute the weekly average as defined in Section 5.4. This procedure will ensure that the data collected on or after intercurrent events are not used in the imputation process.

For the primary efficacy endpoint of vIGA-AD success at Week 4 and the secondary endpoint of vIGA-AD success at Week 4 among subjects with a 'Moderate' randomized vIGA-AD score, the primary analysis will impute missing values using a Predictive Mean Matching (PMM) sequential-regression multiple imputation model for the ITT population. This is a three-step process.

1. The first step is to understand the pattern of missingness. In order to perform the multiple imputation, a monotone missing pattern has to be achieved. For example, if there exist values for baseline and Week 4 visits, but missing values for the Week 1 or 2 visits, the Markov-Chain Monte-Carlo (MCMC) method will be used to impute the small amount of missing data that may be missing at the intermediate visits that is required to make the missing data pattern monotone before applying the multiple imputation algorithm. This method uses a non-informative Jeffreys prior to derive the posterior mode from the expectation-maximization (EM) algorithm as the starting values for the MCMC method. The MCMC method will use the seed 6457149. The vIGA-AD score will be treated as a continuous variable for this step and the model will include the vIGA-AD scores at baseline, Week 1, Week 2, and Week 4. 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:

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

   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 treatment group, pooled study site, and vIGA-AD score outcomes at previous visits, using a seed of 482371. For other scales/questionnaires, the actual baseline vIGA score will also be included as an independent variable. 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 of the 25 to 250 completed dataset, the necessary derived variable will be computed as defined in Section 5.5 and analyzed using a CMH analysis, adjusted for the pooled study site and randomized vIGA-AD score for the primary efficacy endpoint and adjusted for the pooled site for the secondary efficacy endpoint. Results will be combined into one multiple imputation inference as follows:
    a. Common proportion of success, common Mantel-Haenszel (MH) proportion difference (and associated 95% CI), and common MH odds ratio (and associated common 95% CI) will be combined using PROC MIANALYZE based on Rubin's rule. Common MH odds ratios and associated common 95% CI will first be normalized using a log-transformation before being combined using PROC MIANALYZE. The resulting combined common MH odds ratio and associated combined common 95% CI will be back-transformed to the arithmetic scale before being presented in a table.
    b. For the combined common proportion of success, the associated combined common 95% CI will be calculated as per Lott and Reiter multiple imputation Wilson interval method[1].
    c. Two p-values will be produced for each analysis:
        i. The primary p-value will be obtained from a multiple imputation CMH test, where CMH general association statistics and their standard errors obtained

from the analysis of each multiply-imputed dataset will be transformed as per Wilson-Hilferty[2] before being combined using PROC MIANALYZE based on Rubin's rule. Because the Wilson-Hilferty transformation is a monotone transformation, the p-value for the CMH test is the one-sided p-value from the t distribution. This p-value will be the primary p-value used to evaluate the result according to the multiple testing strategy described in Section 6.5.

i. Should the common MH odds ratio from the analysis of at least one of the multiply-imputed datasets be not estimable, combined common MH odds ratio and associated combined common 95% CI and p-value will not be presented. Under such circumstance, conclusions will be based on the p-value obtained from a multiple imputation test of the proportion difference, where the common proportion difference using MH weights and associated common standard errors based on the Sato variance estimator obtained from the analysis of each multiply-imputed dataset will be combined using PROC MIANALYZE based on Rubin's rule.

Similar multiple imputation method will be used for the average weekly WI-NRS and EASI total score.

For the average weekly WI-NRS (refer to Section 5.5), it's the missing average weekly WI-NRS data (i.e., those that cannot be computed because only 3 or less non-missing assessments are available during a given week) at Weeks 1, 2, 3, and 4 that will be multiply imputed, not the missing daily WI-NRS data. Since the missing average weekly WI-NRS values have a precision of 1 decimal place, the MCMC imputation for the non-monotone missing data will be restricted to values between 0 and 10, rounded to the 1st decimal (i.e., 0.1). Imputation of missing data for WI-NRS will be based on the ITT population, i.e., all randomized subjects will be included in the imputation process.

For the EASI total score (refer to Section 5.5), it's the missing EASI total score data at Weeks 1, 2, and 4 that will be imputed, not the missing EASI question score data. Since the missing EASI total scores have a precision of 1 decimal place, the MCMC imputation for the non-monotone missing data will be restricted to values between 0 and 72, rounded to the 1st decimal (i.e., 0.1). Imputation of missing data for EASI total score will be based on the ITT population.

6.2.2 Non-responder Imputation Analysis

If assessment of vIGA-AD after baseline is missing, the subject will be considered as non-responder (for example, no vIGA-AD success in the analysis of vIGA-AD success). For the vIGA-

AD success endpoints, subjects with missing baseline vIGA-AD score will be considered as non-responder for these endpoints only. That is, for the secondary efficacy endpoints of vIGA-AD score of clear (0) or almost clear (1) and vIGA-AD score of clear (0), the post-baseline assessment at a specific visit will be established based on the vIGA-AD score at that visit only, regardless of the availability of the baseline vIGA-AD score.

Similar imputation method will be used for 4-point reduction on the average weekly WI-NRS and EASI-75. For EASI-75, subjects with missing baseline EASI total score will be considered as non-responder. For 4-point reduction on the average weekly WI-NRS, subjects with missing baseline average weekly WI-NRS will not be imputed since the analysis of WI-NRS endpoints will be performed based on the WI-NRS population from which subjects with a missing baseline average weekly WI-NRS are excluded.

6.2.3 Tipping Point Analysis

As a sensitivity analysis to the multiple imputation analysis as described in Section 6.2.1 for the vIGA-AD success primary endpoint and secondary endpoint of vIGA-AD success among subjects with Moderate disease at baseline, a tipping point analysis will be performed in order to determine the inflection point at which the inference under the missing not at random (MNAR) assumption changes substantially.

The sensitivity analysis for the primary endpoint will be performed by using a specified sequence of shift parameters. The range of shift parameters to be included in this analysis are 0 to 2 by 0.2 for active and -2 to 0 by 0.2 for Vehicle. The values at which the results of the primary analysis are shifted from significant (i.e., $\alpha \leq 0.05$) to non-significant (i.e., $\alpha > 0.05$) will be determined. Steps 1 and 3 of the analysis will be the same as for the multiple imputation analysis as described in Section 6.2.1. However, Step 2 of the analysis is where the shift parameters will be applied.

Imputed values for subjects who discontinue due to lack of efficacy or adverse event will be handled as described in Step 2 of Section 6.2.1 to ensure that these subjects are analyzed as non-responders at all visits on or after discontinuation of treatment.

6.3 Interim Analysis

No interim analysis is planned for this study.

6.4 Multicenter Studies

The study will be conducted at approximately 60 study sites in the US, Canada, and Poland. During the conduct of the study, additional countries and/or sites may be added if necessary.

Sites with less than 16 subjects in the ITT population will be pooled within a country:

a. Sites with less than 16 subjects will be ordered from lowest to highest in terms of number of ITT subjects. In case of ties, the ordering for tied sites will be determined according to the site identification number (from smallest to largest).
b. Sites will be combined beginning at the smallest until the resulting pooled site contains at least 16 ITT subjects with at least 1 subject in each treatment group. The sites pooled in this way will be considered as a single site in the statistical analyses.
c. The process described above will resume for the remaining sites not meeting the criterion of at least 16 ITT subjects with at least 1 subject in each treatment group. If the final set of pooled sites does not meet the criterion of at least 16 ITT subjects with at least 1 subject in each treatment group, the final set will be pooled with the preceding pooled site.
d. If there is only one site with less than 16 ITT subjects with at least 1 subject in each treatment group, then this site will be combined with the site with the second lowest number of subjects. As above, in the case of ties, the ordering for tied sites will be determined according to the site identification number (from smallest to largest).

As a sensitivity analysis of different pooling strategy, sites that have randomized less than 50% of the number of randomized subjects at the site with the largest number of randomized subjects, those sites will be pooled within each country. The sensitivity analysis of pooling strategy will be applied for the primary endpoint only.

6.5 Multiple Comparisons/Multiplicity

To control for familywise type I error at level of 0.05, the secondary endpoint of vIGA-AD Success at Week 4 for subjects with vIGA-AD score of 'Moderate' at randomization will only be tested if the primary endpoint demonstrates statistical significance. In addition, the remaining secondary endpoints will be inferentially tested only if the primary and the secondary endpoint (vIGA-AD Success at Week 4 for subjects with vIGA-AD score of 'Moderate' at randomization) comparisons are statistically significant using hierarchical testing procedure by partitioning of alpha.

Upon successful demonstration of statistical significance for the primary and above secondary endpoint, the remaining endpoints will be grouped into secondary endpoint family 1, comprised of the 4-point reduction on the WI-NRS endpoint, at Week 4, Week 2 and Week 1, and secondary endpoint family 2, comprised of the endpoints of EASI-75 at Week 4, vIGA-AD of 'clear' or 'almost clear' at Week 4, vIGA-AD of success at Week 2 and Week 1, vIGA-AD of 'clear' or 'almost clear' at Week 2 and Week 1. An alpha level of 0.03 will be used to test the endpoints in the secondary endpoint family 1 sequentially. An alpha level of 0.02 will be used to test the endpoints in secondary endpoint family 2 sequentially.

In addition to the partitioning of the overall 0.05 alpha into two families, the Fallback Method will be applied. The fallback method is a modification of the fixed-sequence method, providing opportunity to test an endpoint later in the sequence even if an endpoint tested early in the sequence has failed to show statistical significance. The order of the endpoints remains important. The appeal of the fallback method is that if an endpoint later in the sequence has a robust treatment effect while the preceding endpoint is unsuccessful, there is a modest amount of alpha retained as a fallback to allow interpretation of that endpoint without inflating the Type I error rate. Applying the fallback method begins by dividing the total alpha (not necessarily equally) among the endpoints and maintains a fixed sequence for the testing. In this study, the Fallback Method will be applied to the fixed sequence of testing Family 1, and then Family 2.

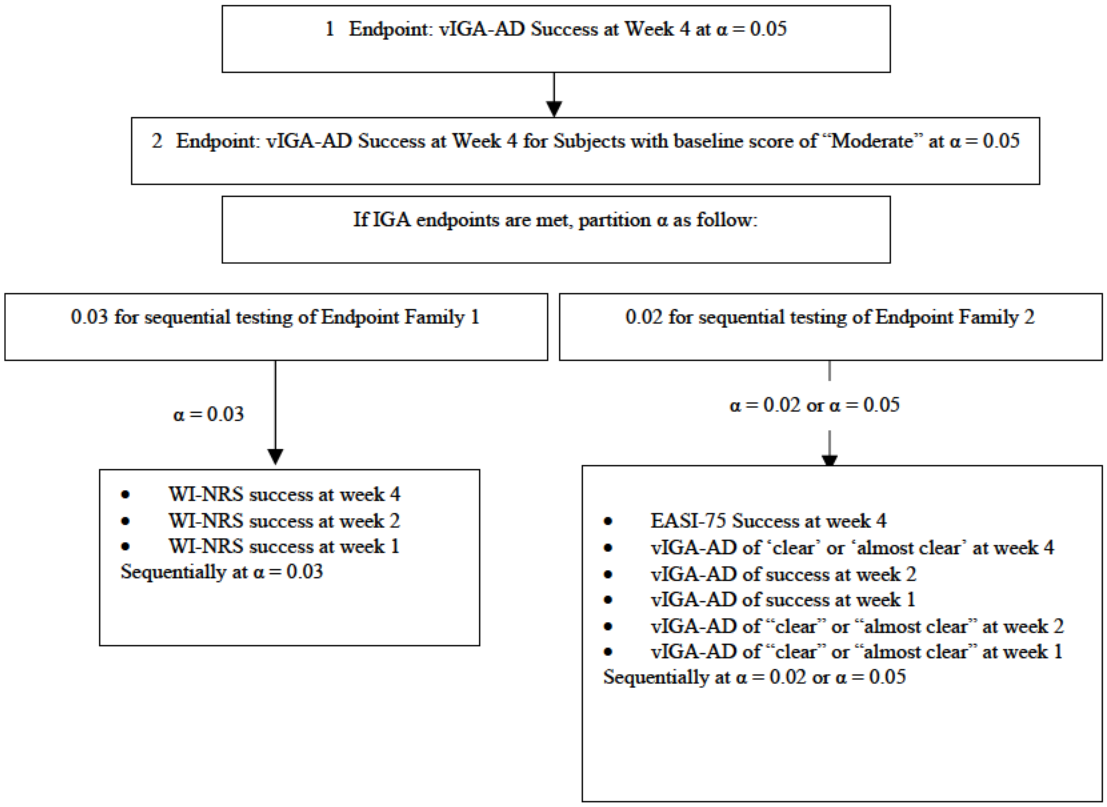
As the testing sequence progresses, a successful test preserves its assigned alpha as "saved" ("unused" or "accumulated") alpha that is passed along to the next test in the sequence, as is the case for the sequential method. This accumulated alpha is added to the prospectively assigned alpha (if any) of that next endpoint and the summed alpha is used for testing that endpoint. Thus, as sequential tests are successful, the alpha accumulates for the endpoints later in the sequence; these endpoints are then tested with progressively larger alphas.

In this study, the Fallback Method will be applied following this sequence:

Family 1: Testing will proceed at the 0.03 level sequentially within Family 1. Should all 3 endpoints in Family 1 be statistically significant at the 0.03 level, then the full 0.03 alpha will be carried to Family 2. Family 2 would then be tested at the full (α=0.02+0.03=0.05).

Should, anywhere during the sequential testing of Family 1, there is a p-value >0.03, the testing within Family 1 will stop, and no additional alpha can be carried over to Family 2.

Figure 2 Multiple Testing Scheme



Achievement of vIGA-AD success is a score of "clear" or "almost clear" plus a 2-grade improvement from baseline.

WI-WRS Success is a 4-point reduction in WI-NRS among subjects ≥12 years old with WI-NRS ≥ 4 at baseline.

EASI-75: achievement of at least a 75% reduction in the Eczema Area and Severity Index

6.6 Examination of Subgroups

Subset analysis for the following subgroups will be performed for the primary and secondary efficacy endpoints:

- Age group (6 – 11 years vs. 12 - 17 years vs. ≥18 years),
- Gender (male vs. female),
- Race (White vs. Black or African American vs. Asian vs. other),
- Ethnicity (Hispanic vs. Non-Hispanic),
- Randomized vIGA-AD score (mild (2) vs. moderate (3)),

- Actual baseline vIGA-AD score (mild (2) vs. moderate (3)),
- Baseline % BSA (<10% vs. ≥ 10%),
- %BSA categories based on tertiles,
- Baseline EASI total score (≤7 vs >7),
- Baseline EASI total score based on tertiles
- Fitzpatrick skin type at Screening (Type I, II and III vs. Type IV, V, and VI),
- Prior inadequate response, intolerance, or contraindication to Topical Corticosteroids (yes vs. no),
- Facial Involvement (yes vs. no)

For subgroup based on tertiles, tertiles will be derived using pooled data from both treatment group based on the ITT population.

Details on these analyses are described in Section 12.4.

7 STUDY SUBJECTS

7.1 Disposition of Subjects

All subjects who provide informed consent will be accounted for in this study. The number of subjects who were screened and who failed screening (screen failures) will be presented. The reasons for screen failure will be presented for all screened subjects who failed screening.

The number of subjects randomized will be presented by treatment group. The number and percentage of the subjects included in each analysis population will be provided by treatment group. The number and percentage of the subjects who completed the study, who discontinued the study, the reasons for study discontinuation, and early termination due to COVID-19 disruption will be presented by treatment group. The percentages will be calculated using the number of the randomized subjects as denominator.

Number of days in the study will be summarized with descriptive statistics by treatment group and overall. For each subject, the number of days in the study will be calculated as following:

Number of days in study = Date of completion/discontinuation – 1st dose date* +1

* For randomized subjects who discontinued study before the first application of study treatment, the date of randomization will be considered instead of the date of the first application of study treatment.

A listing of subject's disposition and randomization will be provided. Information on first screening for subjects who were rescreened, including the rescreened subject identifier and the reason for first screening failure, will be presented under the first screening subject identifier. The reason for screening failure will be listed as well.

A table of randomized strata vs. actual strata will be provided if there is any mis-randomization discrepancy.

7.2 Protocol Deviations

A data review will be conducted before database lock by the Medical Monitor and the Sponsor to classify protocol deviations as minor or major.

The number and percentage of subjects with at least one important protocol deviation (including important protocol deviations associated with COVID-19) will be summarized by deviation category and treatment group using the safety analysis set.

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

8 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Demographics and baseline characteristics will be summarized with descriptive statistics using the ITT and safety population. The list of demographics and baseline characteristics to be summarized will include:

- Age (years)
- Age Group: 6-11, 12-17, 6-17 and ≥18, 18-64 and ≥65
- Childbearing potential
- Sex at birth
- Ethnicity
- Race*
- Baseline Height (cm)

- Baseline Weight (kg)
- Baseline Body Mass Index (BMI) (kg/m^2): Adult, Child/Adolescent
- Fitzpatrick Skin Type
- Prior failure of Topical Corticosteroids, Topical Calcineurin Inhibitors, Eucrisa
- Atopic Dermatitis involvement on the face, on the eyelids
- Baseline vIGA-AD
- Average weekly baseline WI-NRS
- Daily baseline WI-NRS
- Baseline BSA (%)
- Baseline BSA (%) Group - <10% and ≥10%, tertile groups
- Baseline EASI total score
- Baseline EASI score group - ≤7 and >7, tertile groups
- Baseline SCORAD
- Baseline DLQI/CDLQI
- 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.

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

Child and Adolescent (6 to 17 years): After BMI is calculated using the same formula above for children and teens, it is expressed as a percentile obtained from either a graph or a percentile calculator. Percentiles will be calculated using data files and instructions provided by the CDC (https://www.cdc.gov/growthcharts/percentile_data_files.htm). Baseline height will be used to derive BMI for each visit since height is not collected at all visits.

A listing of all demographics, analysis population flag, reason not included in the efficacy analysis 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.

Surgical and medical history will be summarized by system organ class (SOC) and preferred term (PT) using the safety analysis set. 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 analysis set.

10 PRIOR AND CONCOMITANT MEDICATIONS

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

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

Incidence of prior and concomitant medications will be tabulated by ATC level 3 and PT using the safety analysis set. A subject with the same medication taken multiple times will be counted only once for the corresponding PT. 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. Prior and concomitant medications will be sorted alphabetically by ATC level and within each ATC level, the PT will be presented by descending order.

11 STUDY TREATMENT EXPOSURE AND TREATMENT COMPLIANCE

A summary of exposure related to Roflumilast cream and the vehicle will be presented using the safety population by treatment group. It will include descriptive statistics on the number of days on IP, as well as the number of investigational product applications based on diary, for each treatment group. The number of days on IP, will be calculated as follows:

[(last treatment date- first treatment date) + 1].

For each subject, investigational product application compliance (%) will be calculated as follows:

$$\frac{\text{Number of investigational product applications}}{\text{Number of expected investigational product applications}} \times 100$$

Number of investigational product applications will be calculated as number of expected investigational product applications minus number of doses missed. Number of doses missed and the date that the dose was missed were collected in eCRF.

Number of expected investigational product applications will be calculated as calculated as last treatment/interruption date – first treatment date + 1. If latest treatment date ≥ latest interruption date, then the latest treatment date will be used; otherwise, latest interruption date will be used in deriving the expected number of IP applications.

Descriptive statistics for the compliance as well as the number of missed applications, subjects with < 80%, [80% - 100%], and >100% compliance will be presented by treatment. Furthermore, the incidence of subjects who missed more than 3 consecutive doses and compliant subjects will be presented by treatment.

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

Total weight of study medication applied (determined by weighing the study medication before and after use) will be summarized by treatment using descriptive statistics. Weight of study medication used will be documented in source documents and in eCRF. Total weight of IP used is determined by subtracting minimum of (returned, remained) tube weight from the maximum of

(dispensed, prepared) tube weight for each tube that was dispensed and summing the weights. Calculate Actual IP Used: maximum of (dispensed, prepared) tube weight minus minimum of (returned, remained) tube weight = ___ grams

If a tube is not returned at the end of the study or a tube weight is missing, actual IP used for the tube and total weight of IP applied during study will be missing.

Number of days on IP and compliance, including compliance collected at each clinic visit, will be displayed in a listing of study treatment administration by subject, for each treatment. A listing of drug accountability including the kit number, tube number, dispensed and returned weight will also be provided.

12 EFFICACY ANALYSIS

12.1 Primary Efficacy Endpoint Analysis

12.1.1 Primary Efficacy Endpoint and Estimand

The vIGA-AD 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). 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 primary efficacy endpoint is vIGA-AD success, defined as an vIGA-AD score of ‘clear’ (0) or ‘almost clear’ (1) plus at least a 2-grade improvement from Baseline at Week 4.

The primary estimand is described by the following attributes:

Population: Patients with Atopic Dermatitis

Endpoint: vIGA-AD success at Week 4

Intercurrent events: In the course of the 4-week randomized treatment period, subjects may be exposed to possible known or unknown intercurrent events that could possibly impact the estimates of the estimand, such as treatment discontinuation due to a specific adverse effect or perhaps a lack of effect. A composite strategy will be implemented that handles subjects who discontinue due to lack of efficacy or adverse event as missing not at random differently than all other subjects. That is, subjects who discontinue due to lack of efficacy or adverse event will be treated as non-responders for all pre-specified analysis visits (refer to Section 5.4) for which the subject’s last dose day falls within the analysis window or is prior to the start of the analysis window (refer to

Section 5.5) while the "Treatment Policy Strategy" will be adopted for handling intercurrent events in this study other than discontinuation due to lack of efficacy or adverse event.

Population-level summary: ratio of the odds of achieving vIGA-AD success after 4 weeks of using roflumilast cream 0.15%, relative to the odds of success after 4 weeks using a matching vehicle cream in the ITT population.

The supportive population-level summary: the proportion difference between Roflumilast cream 0.15% and vehicle groups will be provided for the patients who achieve vIGA-AD success at week 4 in the ITT population.

12.1.2 Hypothesis Testing

Primary hypothesis testing on the odds ratio: The null hypothesis is that the vIGA-AD success does not differ between roflumilast cream 0.15% and matching vehicle cream. The alternative hypothesis is that the vIGA-AD success does differ between roflumilast cream 0.15% and matching vehicle cream.

Null Hypothesis (H_0): $P_R\ Q_V\ /\ P_V\ Q_R = 1.0$,
Alternative Hypothesis (H_A): $P_R\ Q_V\ /\ P_V\ Q_R \neq 1.0$, where

P_R = the proportion of vIGA-AD success in roflumilast cream 0.15%
P_V = the proportion of vIGA-AD success in matching vehicle cream
$Q_R = 1 - P_R$
$Q_V = 1 - P_V$.

12.1.3 Primary Endpoint Analysis

For the primary analysis, missing vIGA-AD scores will be imputed using multiple imputation as described in Section 6.2.1. These imputations will result in a minimum of 25 to a maximum of 250 complete analysis datasets, depending on the number of imputed monotone datasets that are required.

Percentages of subjects having a vIGA-AD Success (refer to Section 12.1.1) will be presented by visit and treatment group based on multiply imputed data in the ITT population along with a 95% Wilson CI. The common MH odds ratio and common MH proportion difference, adjusted for the randomization factors (i.e., randomized vIGA-AD score and pooled study site) will also be provided along with their common associated 95% CIs. Additionally, count and percentage of subjects having a vIGA-AD success, count and percentage of subjects in each category of the vIGA-AD scale, and descriptive statistics for the vIGA-AD scores, change in vIGA-AD score from

baseline and percent change in vIGA-AD score from baseline will be presented by visit and treatment group based on observed data in the ITT population.

The primary endpoint (vIGA-AD Success at Week 4) will then be analyzed using multiple imputation CMH test stratified by the randomization factors. To do so, a CMH analysis will be performed separately for each of the complete multiply imputed analysis data sets, and results will be combined into one multiple imputation inference using the methodology described in Section 6.2.1. Statistical significance will be concluded at the 5% significance level (2-sided). Should the odds ratio be not estimable for at least one multiply imputed dataset, the conclusion of the study will be based on the p-value obtained from a MH test, stratified by the randomization factors, for the common MH proportion difference at Week 4.

The following sensitivity analyses to the primary analysis of the primary endpoint will be performed:

- Multiply-imputed data (refer to Section 6.2.1) on mITT population (refer to Section 4.2) and
- Tipping Point analysis (refer to Section 6.2.3) on ITT population.
- Non-responder imputation (refer to Section 6.2.2) on ITT population.
- Observed data on ITT population (refer to Section 4.1).
- Observed data on PP population (refer to Section 4.3).

For the last three sensitivity analyses, count of subjects having vIGA-AD success will also be presented by visit and treatment group in addition to the percentage of subjects having vIGA-AD success.

To assess the impact of the pooling of the study sites on the primary analysis of the primary endpoint, the following analyses will be performed:

- The primary analysis of the primary efficacy endpoint will be repeated but with a different site pooling strategy (refer to Section 6.4).
- To assess the impact of site on the primary analysis endpoint, the proportion of subjects achieving vIGA-AD success and 95% CI within each site will be tabulated. No p-values will be provided. Forest plots of the proportions (and associated 95% CI) for each site and treatment will also be provided.

- An additional analysis to examine the impact of study site will examine the changes in p-values that occur after removal of subjects from a site. To do so, the primary analysis of the primary efficacy endpoint will be repeated but removing a different pooled study site for each iteration. Forest plots of the odds ratio (and associated 95% CI) for each pooled site removed will also be provided.

12.2 Secondary Endpoints Analysis

Secondary efficacy endpoints are based on the vIGA-AD scale, WI-NRS scale or EASI questionnaire. The list of secondary efficacy endpoints can be found in Section 2 and their derivation in Section 5.5.

- For more details about the vIGA-AD scale refers to Section 12.1.1.
- The WI-NRS scale is a single item scale assessing the subject-reported worst itch severity during the previous 24-hour period. The scale is from 0 to 10 ('no itch' to 'worst itch imaginable').
- For the EASI questionnaire, 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): 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%).

For each secondary efficacy endpoint, missing data will be imputed and data will be summarized as for the primary efficacy endpoint (refer to Section 12.1.3).

Upon successful demonstration of statistical significance for the primary efficacy endpoint, the secondary efficacy endpoint of vIGA-AD success among subjects with vIGA-AD score of 'Moderate' at randomization will be analyzed as described for the primary efficacy endpoint (refer to Section 12.1.3) but based on the vIGA-AD Moderate ITT population (refer to Section 4.4) and stratifying by pooled study site only. That is, as per the vIGA-AD Moderate ITT population definition all subjects included in this analysis will have a randomized vIGA-AD score of Moderate (3) and so, the CMH test cannot be stratified by randomized vIGA-AD score.

Upon successful demonstration of statistical significance for the primary and above secondary efficacy endpoints, the remaining secondary efficacy endpoints will be grouped into secondary

endpoint family 1, comprised of the 4-point reduction on the average weekly WI-NRS at Week 4, Week 2 and Week 1, and secondary endpoint family 2, comprised of the EASI-75 at Week 4, vIGA-AD of 'clear' or 'almost clear' at Week 4, vIGA-AD of success at week 2 and week 1, vIGA-AD of 'clear' or 'almost clear' at Week 2 and Week 1. The analysis of secondary endpoint family 1 will be performed as described for the primary efficacy endpoint (refer to Section 12.1.3) but based on the WI-NRS population (refer to Section 4.6) and an alpha level of 0.03 will be used to test these endpoints sequentially as described in Section 6.5. Similarly, the analysis of secondary endpoint family 2 will be performed as described for the primary efficacy endpoint (refer to Section 6.2.1) based on the ITT population (refer to Section 4.1) but an alpha level of 0.02 will be used to test these endpoints sequentially as described in Section 6.5.

Secondary Endpoint Family 1 (α=0.03, hierarchical testing)

- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the average weekly WI-NRS at Week 4
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the average weekly WI-NRS at Week 2
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the average weekly WI-NRS at Week 1

Secondary Endpoint Family 2 (α = 0.02, hierarchical testing)

- Achievement of at least a 75% reduction in the Eczema Area and Severity Index at Week 4 (EASI-75)
- vIGA-AD of 'clear' or 'almost clear' at Week 4
- vIGA-AD Success at Week 2
- vIGA-AD Success at Week 1
- vIGA-AD of 'clear' or 'almost clear' at Week 2
- vIGA-AD of 'clear' or 'almost clear' at Week 1

For the secondary efficacy endpoint of vIGA-AD Success at Week 4 based on the vIGA-AD moderate ITT population, the following sensitivity analyses to the primary analysis of this efficacy endpoint will be performed if and only if the hierarchical testing procedure (refer to Section 6.5) allows it:

- Tipping Point analysis (refer to Section 6.2.3) on vIGA-AD moderate ITT population.

- Non-responder imputation (refer to Section 6.2.2) on vIGA-AD moderate ITT population.
- Observed data on vIGA-AD moderate ITT population.
- Observed data on vIGA-AD moderate PP population (refer to Section 4.5).

For the last three sensitivity analyses, count of subjects having vIGA-AD success will also be presented at Week 4 by treatment group in addition to the percentage of subjects having vIGA-AD success.

For all other secondary efficacy endpoints, the following sensitivity analyses to the primary analysis of these secondary endpoints will be performed if and only if the hierarchical testing procedure (refer to Section 6.5) allows it:

- Non-responder imputation on WI-NRS population for the WI-NRS secondary efficacy endpoints and ITT population for all other secondary efficacy endpoints.
- Observed data on WI-NRS population for the WI-NRS secondary efficacy endpoints and ITT population for all other secondary efficacy endpoints.

The daily WI-NRS score, average weekly WI-NRS, and WI-NRS success flag will be listed.

The EASI total score and EASI-75 flag at each visit will be listed.

12.3 Exploratory Efficacy Endpoints

Exploratory efficacy endpoints are based on the vIGA-AD scale, WI-NRS scale, EASI questionnaire, %BSA affected by AD, CDLQI/DLQI questionnaires, DFI questionnaire, SCORAD tool or POEM tool. The list of exploratory efficacy endpoints can be found in Section 2 and their derivation in Section 5.5.

- For more details about the vIGA-AD scale, refer to Sections 12.1.1.
- For more details about the WI-NRS scale, refer to Section 12.2.
- For more details about the EASI questionnaire, refer to Section 12.2.
- The % 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, and soles).
- The CDLQI/DLQI is a self-administered validated questionnaire designed to measure the health-related quality of life of children/adult subjects suffering from a skin disease. It

consists of 10 questions concerning subjects' perception of the impact of skin disease on different aspects of their health-related quality of life over the last week. Questions 1 to 6 and 8 to 10 are rated from 0 (Not at all) to 3 (Very much). For children, if last week was a school time, question 7 is rated from 0 (Not at all) to (Prevented school) but if the last week was a holiday time, question 7 is rated from 0 (Not at all) to 3 (Very much). For adults, question 7 is rated from 0 (Not at all/Not relevant/No) to 3 (Yes).

- The DFI questionnaire measures how much having a child with AD affects the quality of life of other (adult) members of the family over the last week. It is designed to be completed by caregivers of subjects ≤ 17 years of age and consists of 10 questions rated from 0 (Not at all) to 3 (Very much).

- The SCORAD is a clinical tool to assess 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).

- The POEM is a tool used for monitoring atopic eczema severity. If focus on the illness as experience 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).

For the continuous exploratory efficacy endpoints of change and percent from baseline in average weekly WI-NRS, EASI total score, % BSA, DLQI/CDLQI score, DFI score, SCORAD score, and POEM score at Week 1, Week 2, and Week 4, descriptive statistics for the score, change from baseline and percent change from baseline will be presented by visit and treatment group based on observed data in the ITT population. Descriptive statistics will also be presented similarly for the average weekly WI-NRS at Week 3. For the continuous efficacy exploratory endpoint of daily WI-NRS, descriptive statistics will be presented similarly but on a daily basis instead of a weekly basis. Additionally, a plot of the mean (and standard error) daily WI-NRS scores over time for each treatment group will also be provided based on observed data. A similar plot will be provided for the percent change from baseline in daily WI-NRS.

For each categorical exploratory efficacy endpoint, count and percentages of subjects meeting the criteria for an exploratory efficacy endpoint will be presented by visit and treatment group based on observed data in the ITT population along with a 95% Wilson CI. Additionally, count and percentage of subjects meeting the criteria for vIGA-AD of Clear (0), EASI-50, EASI-90, and EASI-100 at Week 1 and Week 2 will also be presented by visit and treatment group based on observed data in the ITT population along with a 95% Wilson CI.

Continuous exploratory efficacy endpoints will be analyzed at Week 1, Week 2, and Week 4 and for daily WI-NRS assessments using an analysis of covariance (ANCOVA) with the factors of treatment, two stratification variables (pooled study site and randomized vIGA-AD score), and baseline of the variable under analysis as covariate. Statistical comparisons between the treatment groups will be obtained using contrasts. The Least Square (LS) mean and its standard error, difference in LS means between treatment group (i.e., active - vehicle), its standard error and associated 95% confidence interval, and p-value for difference from vehicle will be presented at each visit. These analyses will be performed based on observed data in the ITT population.

Categorical exploratory efficacy endpoints at Weeks 1, 2, and 4 will be analyzed using a CMH test adjusted for the two stratification variables (pooled study site and randomized vIGA-AD score). Common MH odds ratio, common MH proportion difference and their common associated 95% CI, adjusted for the randomization factors, will be provided. The p-value will be from a CMH test for the common MH odds ratio, unless the common MH odds ratio is not estimable. In such circumstances, the p-value will be from a MH test for the common MH proportion difference.

The analysis of these endpoints will be performed on the observed data with no imputation. The p-values will be nominal as no formal inferential testing will be done on exploratory efficacy endpoints.

12.4 Subgroup Analysis

With the following exception, analyses of the primary and secondary efficacy endpoints (refer to Sections 12.1 and 12.2, respectively) will be repeated by subgroups (refer to Section 6.6) based on the multiple imputation data using mITT, ITT or WI-NRS population, as applicable (refer to Section 12.2). The exception is that the subgroup analyses by randomized vIGA-AD score and by actual baseline vIGA-AD score will not be performed for the secondary efficacy endpoint of vIGA-AD Success based on the vIGA-AD Moderate ITT population.

The following alternatives could be implemented if the odds ratio and/or proportion difference are not estimable:

1. An unstratified model can be used if odds ratios are not estimable due to over-stratification, i.e., too few events for the number of strata.
2. If the between-imputation variance of odds ratios and/or proportion difference is 0, i.e., the estimates are equal across all imputations, the estimates from any imputation will be reported.

Should a specific subgroup have less than 10 subjects across both treatment groups, no statistical inference will be performed. Forest plots of subgroup analysis for primary efficacy endpoint and secondary efficacy endpoints will also be provided.

No subgroup analyses will be presented for the exploratory efficacy endpoint (refer to Section 2).

12.5 Summary of Primary and Secondary Efficacy Analysis

Table 5 provides a summary of the primary and sensitivity analyses that will be provided for primary and secondary efficacy endpoints.

Table 6 Summary of Primary and Secondary Efficacy Analyses

Efficacy Endpoint	**Primary Analysis**	**Sensitivity Analysis**
Primary (refer to Section 12.1)		
vIGA-AD Success (a score of '0' or '1' plus at least a 2-grade improvement) at week 4	ITT, multiple imputation (CMH)	#1 mITT, multiple imputation (CMH) #2 ITT, Tipping point (CMH)) #3 ITT, non-responder imputation (CMH) #4 ITT, observed data (CMH) #5 PP, observed data (CMH) #6 ITT, multiple imputation (CMH) and a different site pooling strategy (CMH) #7 ITT, multiple imputation (CMH) by pooled study site #8 ITT, multiple imputation (CMH) and removing one pooled study site at the time
Secondary (refer to Section 12.2)		

vIGA-AD Success (a score of '0' or '1' plus at least a 2-grade improvement) at week 4	vIGA-AD Moderate ITT, multiple imputation (CMH)	#1 vIGA-AD Moderate ITT, Tipping point (CMH) #2 vIGA-AD Moderate ITT, non-responder imputation (CMH) #3 vIGA-AD Moderate ITT, observed data (CMH) #4 vIGA-AD Moderate PP, observed data (CMH)
Family 1 (α=0.03): Average weekly WI-NRS Success (achievement of at least a 4-point reduction) at weeks 1,2, and 4	WI-NRS population, multiple imputation (CMH)	#1 WI-NRS, non-responder imputation (CMH) #2 WI-NRS, observed data (CMH)
Family 2 (α=0.02 or α=0.05): EASI-75 at week 4	ITT, multiple imputation (CMH)	#1 ITT, non-responder imputation (CMH) #2 ITT, observed data (CMH)
Family 2 (α=0.02 or α=0.05): vIGA-AD Success (a score of '0' or '1' plus at least a 2-grade improvement) at weeks 1 and 2	ITT, multiple imputation (CMH)	#1 ITT, non-responder imputation (CMH) #2 ITT, observed data (CMH)
Family 2 (α=0.02 or α=0.05): vIGA-AD score of '0' or '1' at weeks 1, 2 and 4	ITT, multiple imputation (CMH)	#1 ITT, non-responder imputation (CMH) #2 ITT, observed data (CMH)

13 SAFETY ANALYSIS

Safety analyses will be conducted using the safety population. Subjects will be analyzed based on the treatment received and the stratum they belong to.

No formal inferential statistics will be performed on safety assessments.

Also, treatment-emergent adverse events by SOC and PT table and summary of weight change will be presented by treatment groups for age 6 to 17 years and ≥ 18 years separately.

13.1 Adverse Events

Adverse events (AEs) will be coded according to the latest available version of Medical Dictionary for Regulatory Activities (MedDRA) (Version 25.0).

Treatment emergent adverse events (TEAEs) are defined as any AEs with onset on or after the first study drug application. 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. All reported TEAEs will be summarized by treatment group.

Overall summary will be presented, which will include the total number of events, and the number and percentage of subjects who experienced TEAE, TEAE by the strongest relationship, TEAE by the maximum severity, treatment-related TEAE by maximum severity, treatment-emergent serious AE (TESAE), 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.

The number and percentage of subjects who experience TEAE will be summarized by SOC and PT within SOC. In addition, similar table will be presented by treatment groups for age 6 to 17 years and ≥ 18 years separately. Unless otherwise specified, a subject experiencing the same TEAE multiple times will be counted only once for the corresponding PT. Similarly, if a subject experiences multiple TEAEs within the same SOC, the subject 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 the safety analysis set. 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 be summarized by SOC, PT and the maximum severity (mild/moderate/severe/life threatening/death related to AE). 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 TESAE, treatment-emergent Non-SAE, will be summarized by SOC and PT within SOC.

Frequency and percentage of subjects who experience TEAE on an application site will be summarized by SOC and PT.

A table and plot of most frequent TEAE (‘≥1%) by PT will be provided by treatment arms (overall TEAE, overall TESAE will be included in the same plot).

All the AEs will be listed. Any TEAE leading to death will also be included in the AE listing (if there is any). The TEAE related to application site will be flagged in the AE listing.

13.2 Clinical Laboratory

Descriptive statistics for the observed values in chemistry, hematology, and quantitative urinalysis, change from baseline and percent change from baseline values will be presented by treatment group at each scheduled visit. For qualitative urinalysis data, the number and percentage of the subjects for each level of result by treatment at each scheduled visit will be provided.

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

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

Laboratory data will be presented in SI units.

13.3 Vital Signs

Descriptive statistics will be presented for vital signs (systolic blood pressure, diastolic blood pressure, heart rate, body temperature, and weight). Observed values, change from baseline and percent change from baseline values will be presented by treatment group at each scheduled visit.

The number and percentage of subjects with gain or lose >5% from baseline in body weight over the course of the study will be summarized for overall and separately for children and adolescents (6 to 17 years) and adults (≥ 18 years). BMI will be summarized using descriptive statistics (n, mean, standard deviation, median, Q1, Q3, minimum, and maximum) on observed values, change from baseline and percent change from baseline values separately for children and adolescents (6 to 17 years) and adults (≥ 18 years). The BMI percentile rather than BMI kg/m^2 will be summarized for children and adolescents (6 to 17 years).

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

13.4 Local Tolerability Assessments

The investigator's assessment of the application site reaction will be summarized by visit using both categorical methods (number and percentage of subject with each score) as well as continuous methods (e.g., mean, standard deviation, etc.)

Local tolerability (burning/stinging sensation) assessed by the subject will be summarized using number and percentage similarly.

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

The Modified PHQ-A Assessment will be performed in adolescent subjects (12-17 years old, inclusive; question 9 has been removed since that is better evaluated by use of the C-SSRS tool).

The PHQ-8/Modified PHQ-A score is the sum of the responses for the questions, each question ranging from 0 (Not at all) to 3 – (Nearly every day). Five severity categories of depression are defined 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 score will be set to missing in case of at least one missing value. The number and percentage of subjects in each category will be summarized by treatment and visit. A summary of the shifts in depression category from baseline to each study visit will also be provided.

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 baseline will be calculated for CDI-2 total score and the 2 scales, i.e., emotional problems and functional problems, and will be summarized descriptively by treatment group and visit.

The CDI-2 total score will be categorized as follows:

1. Normal: Male Score < 22, Female Score < 21
2. Elevated: $22 \leq$ Male Score < 32, $21 \leq$ Female Score < 32
3. Very Elevated: Male or Female Score ≥ 32.

The proportion of subjects that meeting the criteria will be summarized by treatment group and visit.

A summary of the shifts in CDI-2 category from baseline to each study visit will also be provided.

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

The Columbia-Suicide Severity Rating Scale (C-SSRS) is a questionnaire used (for adolescents and adults 12 years old and older) for suicide assessment developed by multiple institutions, including Columbia University. The C-SSRS prospectively assesses Suicidal Ideation and Suicidal Behavior. At the Screening study visit, "Baseline/Screening" version of the C-SSRS will be used. This version assesses Suicidal Ideation and Suicidal Behavior during the subject's lifetime and during the past 6 months. For the Screening visit, "lifetime" experience of the subject with Suicidal Ideation and Suicidal Behavior will be summarized. On all subsequent visits, the "Since Last Visit" version will be used (Baseline/Day 1, Week 1/Day 8, Week 2/Day 15 or Week 4/Day 29).

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.

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 and treatment group at each study visit.

14 PHARMACOKINETICS ANALYSIS

Concentration data will be summarized by visit by age group (<18 vs. >=18) and overall, for active treatment group using descriptive statistics, reporting n, mean, standard deviation, median, Q1, Q3, minimum, and maximum, and geometric statistics including geometric mean and coefficient of variation. For computation of mean plasma concentrations, data that are below the limit of quantification (BLQ) will be set to 0.0001. The PK population will be used for these analyses.

PK data will be presented in the listing.

15 REFERENCES

1. Anne Lott & Jerome P. Reiter (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

2. Ratitch, B., Lipkovich, I., & O'Kelly, M. (2013). *Combining Analysis Results from Multiply Imputed Categorical Data*. PharmaSUG. https://www.pharmasug.org/proceedings/2013/SP/PharmaSUG-2013-SP03.pdf

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.

P-values equal to or above 0.0001 will be reported to 4 decimal places; p-values less than 0.0001 will be reported as "<0.0001"; p-values greater than 0.9999 will be reported as ">0.9999".

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 group, 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:

Study Treatment Full Name	Study Treatment Output Name
Roflumilast cream 0.15%	Roflumilast Cream 0.15%
Vehicle cream	Vehicle

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 day 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 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 date is the same as the first dose date and event occurred prior to study drug application (as flagged in CRF), impute to 00:00.
- If the event date is the same as the first dose date and event did not occur prior to study drug application (as flagged in CRF), 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)

Impute to 23:59.

Protocol ARQ-151-312

A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis

Sponsor:	Arcutis Biotherapeutics, Inc. 3027 Townsgate Road, Suite 300 Westlake Village, CA 91361
Sponsor Representative:	[redacted]
Medical Monitor:	[redacted]
IND Number:	[redacted]
Protocol Version:	Amendment 4
Date:	28 September 2022

GCP Statement

This study is to be performed in full compliance with the protocol, International Conference on Harmonisation Good Clinical Practices (ICH GCP), and applicable regulatory requirements. All required study documentation will be archived as required by regulatory authorities.

Confidentiality Statement

SITE INVESTIGATOR SIGNATURE PAGE

ARQ-151-312

A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis

SPONSOR: Arcutis Biotherapeutics, Inc.
3027 Townsgate Road, Suite 300
Westlake Village, CA 91361

ISSUE DATE: 28 September 2022

I have read this protocol, including all appendices, and I agree that it contains all necessary details for me and my staff to conduct this study as described. I will conduct this study as outlined herein, in accordance with the current International Conference on Harmonisation Good Clinical Practices (ICH GCPs) and applicable local and regional regulations.

I will provide all study personnel under my supervision copies of the protocol and access to all information provided by Arcutis Biotherapeutics, Inc. I will discuss the material with them to ensure that they are fully informed about ARQ-151 and the study.

I agree that I or my designee will completely inform all subjects in this study concerning the pertinent details and purpose of the study prior to their agreement to participate in the study in accordance with cGCPs and regulatory authority requirements. I will be responsible for maintaining each subject's consent form in the study file and providing each subject with a signed copy of the consent form.

I agree to maintain the confidentiality of all information received or developed in connection with this protocol.

Investigational Site Name: [redacted]

Print Investigator Name: [redacted]

Investigator Signature: [redacted] Date: [redacted]

SUMMARY OF CHANGES

The following sections have been changed in Amendment 4 of the ARQ-151-312 protocol:

Version/Date	Description
31 October 2020	Original Protocol
Amendment 1 08 June 2021	• Added Summary of Changes section. • Revised the proportion of adult subjects in the study from up to 25% to up to 50%. This change was made to the synopsis, study schema, and Section 4.2 "Number of Sites and Subjects".
Amendment 2 18 July 2021	• Updated Arcutis Biotherapeutics, Inc. address. • Updated Medical Monitor information. • Updated Protocol Synopsis to reflect protocol changes. • Clarification added that if an unscheduled visit is required for reasons other than safety, the following assessments are not required: – vIGA-AD and EASI – BSA affected with AD – Local tolerability assessment (by Investigator) • WI-NRS: added statement that subjects will be given instructions on how to complete this questionnaire for subjects 6 years of age and older. • Added clarification that only subjects 18 years and older will be enrolled at sites located in the province of Québec in Canada. • Updated Figure 1 to note that subjects who have undergone a bilateral tubal ligation/occlusion are considered females of childbearing potential using a highly effective method of contraception. • Updated subject ages for hematology, serum chemistries, and urine analysis: all subjects will be evaluated at Screening but only subjects ≥12 years old will be evaluated at Baseline/ Day 1 and at Week 4/Day 29/ET and have a PK sample collected at Week 4/Day 29/ET. • Updated that all AEs occurring after the first application of IP through the end of the study should be collected, and all SAEs should be collected starting at Screening. • Added estimand language in the primary analysis section. • Updated "key secondary endpoints" to "secondary endpoints" and updated the language to be consistent throughout the document. • Updated the multiple testing procedure to hierarchical testing for secondary endpoint family 2. • Added per protocol population. • Removed the language for continuous endpoint analysis using ANCOVA model and will be added in SAP. • Removed time to event endpoints.

Version/Date	Description
	• Added new endpoints of vIGA-AD success at Week 1 and Week 2, vIGA-AD of 'clear' or 'almost clear' at Week 1, Week 2 and Week 4 and WI-NRS success at Week 1 and Week 2. • Updated wording of endpoints to make it consistent throughout the documents.
Amendment 3 23 May 2022	• Updated Medical Monitor information. • Revised the wording for the Per protocol population with the addition of the bolded text: – Per protocol (PP) population will include all subjects in the ITT population, who are at least 80% compliant with study medication application, **have a vIGA-AD assessment within the Week 4 visit window,** and show no major deviations from the study protocol that would affect the interpretation of efficacy. • Revised the wording for the WI-NRS population with the addition of the bolded text: – WI-NRS population will be a subset of the ITT population **who are ≥12 years old at Baseline and have a** Baseline WI-NRS score ≥4. • Revised the secondary endpoints to be tested using the ITT population, with the exception of the WI-NRS endpoints, which will use the WI-NRS population. • Minor grammatical/editorial changes throughout the document.
Amendment 4 28 September 2022	• Updated Sponsor Representative. • Updated the method of handling the intercurrent events of discontinuation due to adverse event or lack of efficacy to identify these subjects as non-responders at any visit that occurred or would have occurred after the date of last dose of study treatment. • Minor grammatical/editorial changes throughout the document.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Definition
α	Alpha Level (significance level)
AE	Adverse Event
AMP	Adenosine Monophosphate
AD	Atopic Dermatitis
AUC	Area Under the Curve
BSA	Body Surface Area
CDI	Children's Depression Inventory
CDLQI	Children's Dermatology Life Quality Index
C_{max}	Maximum Concentration
cm	Centimeter
COPD	Chronic Obstructive Pulmonary Disease
CRF	Case Report Form
C-SSRS	Columbia-Suicide Severity Rating Scale
CTCAE	Common Terminology Criteria for Adverse Events
DFI	Dermatitis Family Impact
DNA	Deoxyribonucleic Acid
DLQI	Dermatology Life Quality Index
EASI	Eczema Area and Severity Index
FDA	U.S. Food and Drug Administration
FOCBP	Female of Child Bearing Potential
FSH	Follicle-Stimulating Hormone
GCP	Good Clinical Practices
GLP	Good Laboratory Practices
hr	Hour
IB	Investigational Brochure
IC_{50}	Half Maximal Inhibitory Concentration
ICF	Informed Consent Form
ICH	International Conference on Harmonisation
ID	Identification
IGA	Investigator Global Assessment

Abbreviation	Definition
IL	Interleukin
IND	Investigational New Drug
IP	Investigational Product
IRB	Institutional Review Board
ITT	Intent to Treat
IWRS	Interactive Web Response System
Kg	Kilogram
LED	Light Emitting Device
µg	Microgram
MedDRA	Medical Dictionary for Regulatory Activities
Mg	Milligram
Min	Minute
mL	Milliliter
MTD	Maximum Tolerated Dose
NCI	National Cancer Institute
NIH	National Institutes of Health
NOAEL	No Observed Adverse Effect Level
Ng	Nanogram
P-450	Cytochrome P450
PDE-4	Phosphodiesterase 4
PDMP	Protocol Deviation Management Plan
PHQ-8	Patient Health Questionnaire-8
PI	Principal Investigator
PK	Pharmacokinetics
POEM	Patient-Oriented Eczema Measure
QD	Once Daily ("quaque die")
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SCORAD	Scoring Atopic Dermatitis
SD	Standard Deviation
SUSAR	Suspected Unexpected Serious Adverse Reaction

Abbreviation	Definition
TCPS	Tri-Council Policy Statement
TEAE	Treatment Emergent Adverse Event
T_{max}	Time to Reach Maximum Concentration
TPA	Target Plaque Area
TPSS	Target Plaque Severity Score
US	United States
UVR	Ultraviolet Radiation
V79	Chinese Hamster Cell Line
vIGA-AD	Validated Investigator Global Assessment - Atopic Dermatitis
WI-NRS	Worst Itch - Numeric Rating Score

1. PROTOCOL SUMMARY

1.1. Synopsis

Protocol Title:	A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis
Clinical Indication:	Atopic Dermatitis
Investigational Product:	• ARQ-151 will be supplied as an emollient cream at 0.15% strength • Matching vehicle cream will contain only excipients of ARQ-151
Study Design:	This is a Phase 3, parallel group, double blind, vehicle-controlled study in which ARQ-151 cream 0.15% or vehicle is applied QD for 4 weeks to subjects 6 years of age and older with mild to moderate atopic dermatitis. At entry, subjects will have ≥3% BSA involvement (excluding the scalp, palms, soles) and mild or moderate atopic dermatitis (AD) based on vIGA-AD assessment. Upon determination of eligibility, subjects will be randomized 2:1 to either ARQ-151 cream 0.15% cream or matching vehicle cream. The randomization will be stratified by vIGA-AD score at Baseline/Day 1 ('Mild' vs. 'Moderate') and by study site. Subjects/caregivers will apply ARQ-151 cream 0.15% or vehicle cream QD to all AD affected areas and any newly appearing AD lesions that arise during the study, except on the scalp. Subjects/caregivers should maintain treatment of these areas with study drug for the duration of the study regardless of whether treatable areas of AD clear prior to Week 4/Day 29. At the Week 4 visit, subjects may be eligible to enroll in a 12-month, open label extension study (ARQ-151-313) evaluating ARQ-151 cream 0.15% QD.
Study Objective:	To assess the safety and efficacy of ARQ-151 cream 0.15% vs vehicle administered QD x 4 weeks to individuals with atopic dermatitis.
Study Sites:	Approximately 50 sites in the US and Canada. During the conduct of the study, additional countries and/or sites may be added if necessary.
Study Population:	Subjects will be male and female children and adolescents (6-17 years), and adults (≥18 years). Subjects will have mild to moderate atopic dermatitis involvement with a vIGA-AD score of '2' (Mild) or '3' (Moderate) for study entry. Up to ~50% of the subjects will be ≥18 years old. Approximately 650 subjects are planned to be randomized in this study.
Duration of Participation for Subjects:	Screening (up to 30 days) + Treatment phase (4 weeks) for a total of about 8 weeks. Upon completion of the treatment phase of the study (Week 4/Day 29) subjects may have the opportunity, subject to regulatory approval and enrollment has not been completed, to participate in an open-label extension study (ARQ-151-313) of up to 12 months.

Inclusion Criteria:	1. For adult subjects: Participants legally competent to sign and give informed consent. For pediatric and adolescent subjects: Informed consent of parent(s) or legal guardian, and, if age appropriate, assent by the subjects, as required by local laws. 2. Males and females, ages 6 years and older at time of signing Informed Consent (Screening). Only subjects 18 years and older will be enrolled at sites located in the province of Québec in Canada. 3. Diagnosed with mild to moderate atopic dermatitis according to the criteria of Hanifin and Rajka (1980) prior to or at the screening visit. Subjects must have at least 3 of the 4 basic features per Hanifin and Rajka (1. Pruritus; 2. Typical morphology and distribution [flexural lichenification in adults and facial and extensor eruptions in infants and children]; 3. Chronic or chronically relapsing dermatitis; or 4. Personal or family history of atopy), in addition to 3 or more minor criteria. 4. History of AD for at least 3 months in subjects 6-17 years of age or 6 months in subjects ≥18 years of age, as determined by the Investigator using information from the subject's medical chart, from the subject's physician, or through subject/parent/caregiver interview. Stable disease for the past 4 weeks with no significant flares in atopic dermatitis before screening. 5. EASI Score ≥5 at Baseline. EASI is evaluated for the entire body except the scalp, palms, and soles. 6. vIGA-AD score of 'Mild' ('2') or 'Moderate' ('3') at Baseline. The vIGA-AD is evaluated for the entire body except the scalp, palms, and soles. 7. Has AD involvement of ≥3% BSA (excluding the scalp, palms, soles) at Baseline. 8. Females of childbearing potential (FOCBP) must have a negative serum pregnancy test at Screening and a negative urine pregnancy test at Baseline/Day 1. In addition, sexually active FOCBP must agree to use at least one form of a highly effective or barrier method of contraception throughout the trial. The use of abstinence as a contraceptive measure is acceptable as long as this is a consistent part of a lifestyle choice and an acceptable backup method has been identified if the subject becomes sexually active. 9. Females of non-childbearing potential should either be pre-menarchal, or post-menopausal with spontaneous amenorrhea for at least 12 months (post-menopausal status should be confirmed with FSH testing) or have undergone surgical sterilization (permanent sterilization methods include hysterectomy, bilateral oophorectomy, or bilateral salpingectomy). 10. In good health as judged by the Investigator, based on medical history, physical examination, vital signs, serum chemistry labs, hematology values, and urinalysis. 11. Subjects and parent(s)/legal guardian(s) are considered reliable and capable of adhering to the Protocol and visit schedule, according to the judgment of the Investigator.

Exclusion Criteria:	1. Subjects with any serious medical condition or clinically significant laboratory, vital signs, or physical examination abnormality that would prevent study participation or place the subject at significant risk, as judged by the Investigator 2. Liver function tests excursions that exceed: • AST or ALT >2X ULN • Total bilirubin: – >1.5 x ULN or – >ULN and ≤1.5 x ULN AND direct bilirubin is >35% of total bilirubin • ALP ≥2x ULN 3. Subjects who cannot discontinue medications and treatments prior to the Baseline visit and during the study according to Excluded Medications and Treatments (Table 2). 4. Has unstable AD or any consistent requirement for high potency topical steroids to manage AD signs or symptoms. 5. Subjects who have significant active systemic or localized infection (eg, molluscum contagiosum), including known actively infected AD, or have had any infection that required oral or intravenous administration of antibiotics, antifungal or antiviral agents within 14 days prior to Baseline/Day 1. 6. Subjects who are unwilling to refrain from prolonged sun exposure and from using a tanning bed or other artificial light emitting devices (LEDs) for 4 weeks prior to Baseline/Day 1 and during the study. 7. Subjects with skin conditions other than AD that would interfere with evaluations of the effect of the study medication on AD, as determined by the Investigator. Subjects with any condition on the treatment area which, in the opinion of the Investigator, could confound efficacy measurements, eg, molluscum contagiosum. 8. Subjects with known genetic dermatological conditions that overlap with AD, such as Netherton syndrome. 9. Known allergies to excipients in ARQ-151 cream [redacted] 10. Subjects who cannot discontinue the use of strong cytochrome P-450 CYP3A4 inhibitors eg, indinavir, nelfinavir, ritonavir, clarithromycin, itraconazole, ketoconazole, nefazodone, saquinavir, suboxone and telithromycin for 2 weeks prior to Baseline/Day 1 and during the study period. 11. Subjects who cannot discontinue the use of strong cytochrome P-450 CYP3A4 inducers eg, efavirenz, nevirapine, glucocorticoids, barbiturates (including phenobarbital), phenytoin, rifampin, and carbamazepine for 2 weeks prior to Baseline/ Day 1 and during the study period.

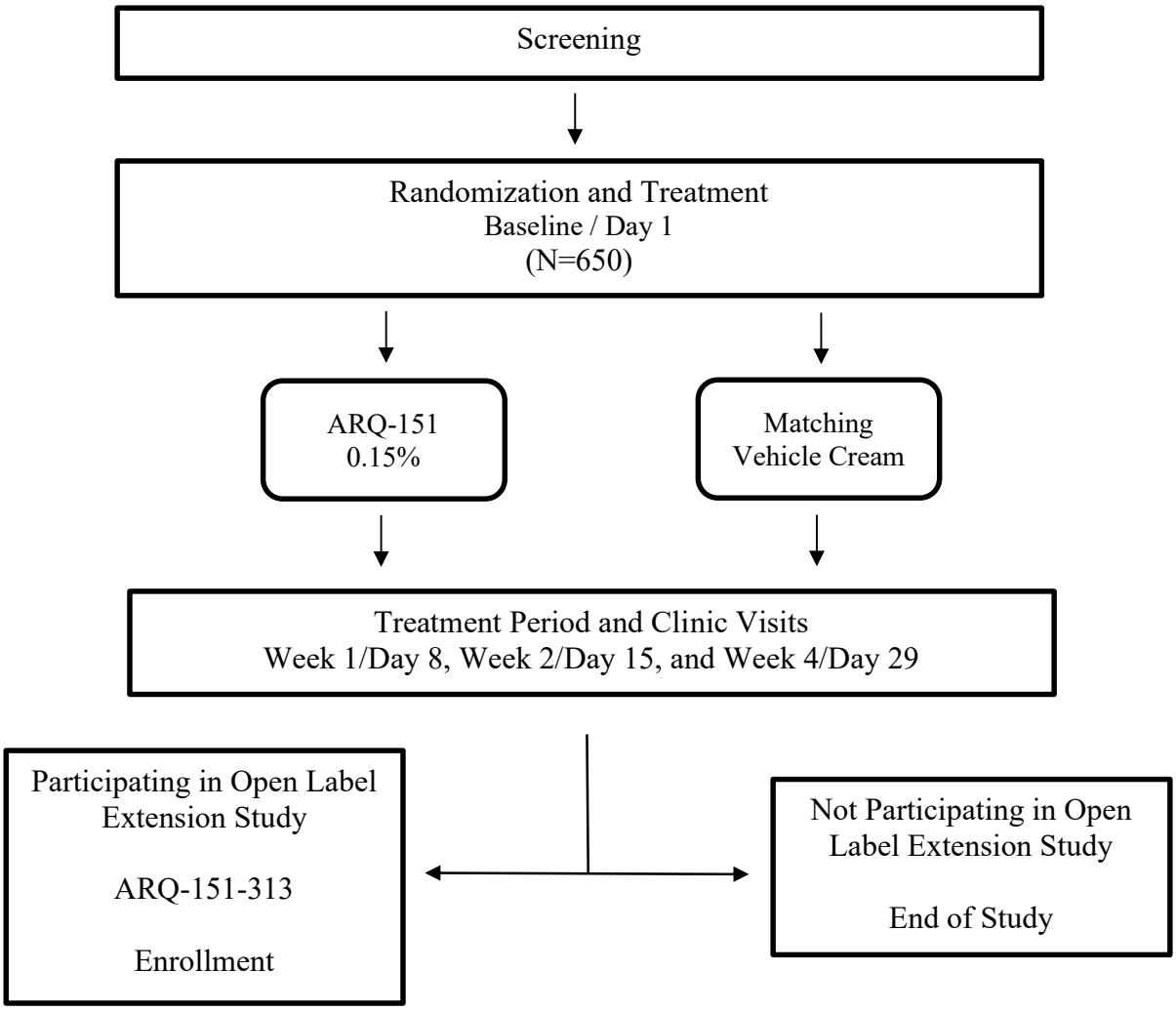
	12. Subjects who have received oral roflumilast (Daxas®, Daliresp®) within 4 weeks prior to Baseline/ Day 1. 13. Known or suspected: • Severe renal insufficiency – Severe renal insufficiency is defined as calculated creatinine clearance <30 mL/min. • Moderate to severe hepatic disorders (Child-Pugh B or C) 14. History of severe depression, suicidal ideation or behavior, Baseline/Screening C-SSRS (for adolescents and adults 12 years old and older) indicative of suicidal ideation or behavior, whether lifetime or recent/current. 15. Subjects with a PHQ-8 (adults) or modified PHQ-A (adolescents, 12-17 years old inclusive) score ≥10 at Screening or Baseline/Day 1 visits. 16. Subjects (6 to 11 years old, inclusive) with a CDI-2 (parent report) raw score ≥17 for females and ≥18 for males at Screening or Baseline/ Day 1 visits. 17. Females who are pregnant, wishing to become pregnant during the study, or are breast-feeding. 18. Previous treatment with ARQ-151. 19. Subjects currently undergoing allergy testing (eg, food allergy testing or skin prick testing), patch testing, food challenges, or allergy desensitization, or plan to do so during the study. 20. Subjects with any serious medical condition (eg, uncontrolled hypo- or hyper-thyroidism) or clinically significant laboratory abnormality that would prevent study participation or place the subject at significant risk, as determined by the Investigator. 21. Subjects with a history of a major surgery within 4 weeks prior to Baseline/Day 1 or subjects who have a major surgery planned during the study. 22. Subjects with a history of chronic alcohol or drug abuse within 6 months prior to Screening. 23. Current or a history of cancer within 5 years with the exception of fully treated skin basal cell carcinoma, cutaneous squamous cell carcinoma or carcinoma in situ of the cervix. 24. Parent(s)/legal guardian(s) who are unable to communicate, read, or understand the local language(s). Subjects who are unable to communicate, read or understand the local language, or who display another condition, which in the Investigator's opinion, makes them unsuitable for clinical study participation. 25. Subjects who are family members of the clinical study site, clinical study staff, or sponsor, or family members of enrolled subjects (subjects enrolled in other studies of ARQ-151) living in the same house.

Key Assessments:	**Safety Assessments:** • Safety will be monitored through local tolerability assessments, vital signs, physical examination, safety labs, Children's Depression Inventory 2 (CDI-2, parent report for children 6-11 years old, inclusive), modified PHQ-A (for adolescents 12-17 years old, inclusive), PHQ-8 (for adults), C-SSRS (for adolescents and adults 12 years old and older), and AEs. • All AEs that occur after the first application of IP through the end of the study should be collected. All SAEs should be collected starting at Screening. • The investigator or a properly trained and designated subinvestigator will perform local tolerability assessments at Baseline/Day 1, and Weeks 1, 2, and 4 (Days 8, 15, and 29). Subjects will have vital signs measured at each study visit. Height will be collected at Visit 1 (Screening) only. • A limited physical exam (skin, lungs, and heart only) will be performed at Screening, Baseline/Day 1 and Week 4/Day 29. Blood and urine samples for routine safety laboratory tests (hematology, serum chemistry, and urinalysis) for all subjects will be obtained at Screening, and for subjects ≥12 years of age at Baseline/Day 1 and Week 4/Day 29/ET along with a PK sample collection at Week 4/Day 29/ET. For all female subjects of childbearing potential, a urine pregnancy test will be administered at all clinic visits except for Screening where a serum pregnancy test will be performed. A negative pregnancy result is required for continued participation in the study, and results (of the urine pregnancy test) must be available prior to dispensing of study drug at study visits. **Efficacy Assessments:** • Efficacy assessments will include vIGA-AD (Appendix 7), EASI, WI-NRS, BSA, DLQI/CDLQI (Appendix 8 and Appendix 9), DFI (Appendix 10), SCORAD (Appendix 11), and POEM (Appendix 12). **Pharmacokinetic Assessment:** • A single PK assessment (trough) will be performed in all subjects with a blood sample collected at Week 4/Day 29.
Study Endpoints:	1. Primary Efficacy Endpoint: • IGA Success, defined as a vIGA-AD score of 'clear' or 'almost clear' PLUS a 2-grade improvement from Baseline at Week 4 2. Secondary Efficacy Endpoints: • In subjects with a vIGA-AD score of 'Moderate' at randomization, vIGA-AD Success at Week 4 • In subjects ≥12 years old with Baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 4 • In subjects ≥12 years old with Baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 2 • In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 1

	• Achievement of at least a 75% reduction in the Eczema Area and Severity Index (EASI-75) at Week 4 • vIGA-AD of 'clear' or 'almost clear' at Week 4 • vIGA-AD Success at Week 2 • vIGA-AD Success at Week 1 • vIGA-AD of 'clear' or 'almost clear' at Week 2 • vIGA-AD of 'clear' or 'almost clear' at Week 1 3. Pharmacokinetic endpoints include concentrations of roflumilast and its N-oxide metabolite.
Power and Sample Size:	Approximately 650 subjects are planned to be randomized in this study. To test the secondary endpoint of IGA success in subjects with a vIGA-AD score of 'Moderate' at randomization, approximately 490 of the subjects to be accrued will have vIGA-AD score of 'Moderate' at randomization. Randomization will be stratified by vIGA-AD score ('Mild' vs. 'Moderate') and by study site. This sample size provides approximately 95% power to detect an overall 15% difference between treatment groups on vIGA-AD success at Week 4 at α=0.05 using a 2-sided stratified Cochran-Mantel-Haenszel test. The results from a recent phase 2 study (ARQ-151-212) of ARQ-151 cream 0.15% compared to vehicle treatment were used to estimate the treatment difference. Specifically, in the phase 2 trial, approximately 37% of subjects demonstrated vIGA-AD success at Week 4 in the ARQ-151 0.15% group compared to 22% in the vehicle group. This sample size also provides approximately 90% power to detect an overall 17% difference between treatment groups on IGA success at Week 4 among subjects with vIGA-AD score 'Moderate' at randomization. The same testing method, the stratified Cochran-Mantel-Haenszel test, will be used as for the primary endpoint.
Statistical Analysis:	The analysis populations are defined as follows: • Intent-to-Treat (ITT) population will include all subjects who are randomized. • Per protocol (PP) population will include all subjects in the ITT population, who are at least 80% compliant with study medication application, have a vIGA-AD assessment within the Week 4 visit window, and show no major deviations from the study protocol that would affect the interpretation of efficacy. In addition, subjects who miss the Week 4 vIGA-AD assessment specifically due to novel coronavirus disease-19 (COVID-19) disruptions will be excluded from per protocol population. • vIGA-AD Moderate ITT population will be a subset of the ITT population with vIGA-AD score 'moderate' at randomization. • vIGA-AD Moderate PP population will be a subset of the PP population with vIGA-AD score 'Moderate' at randomization. • WI-NRS population will be a subset of the ITT population who are ≥12 years old at Baseline and have a Baseline WI-NRS score ≥4.

	• Safety population will include all subjects who are enrolled and received at least one confirmed dose of study medication. • PK population will include all subjects receiving the active drug with quantifiable plasma concentrations of roflumilast. To control for familywise type I error at level of 0.05, the secondary endpoint of IGA success at Week 4 for subjects with vIGA-AD score of 'Moderate' at randomization will only be tested if the primary endpoint demonstrates statistical significance. In addition, the remaining secondary endpoints will be inferentially tested only if the primary and above secondary endpoint (IGA success at Week 4 for subjects with vIGA-AD score of 'Moderate' at randomization) comparisons are statistically significant using hierarchical testing procedure by partitioning of the alpha and use of the Fallback Method. Descriptive statistics for continuous variables will include mean, median, standard deviation, Q1, Q3, min, max. Descriptive statistics for categorical variables will include frequencies and percentages. For missing data, the primary imputation method and sensitivity methods will be detailed in the SAP. The primary endpoint will be analyzed with a Cochran-Mantel-Haenszel test stratified by the randomization factors (disease severity determined by vIGA-AD and by study site). Categorical secondary efficacy analysis will be analyzed in the same manner as the primary endpoint. The incidence of adverse events will be summarized as well as laboratory parameters and vital signs.

1.2. Study Schema



A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis

Approximately 650 subjects with atopic dermatitis will be randomized 2:1 to receive either:

- ARQ-151 cream 0.15% or Vehicle cream

Subjects will have ≥3% BSA involvement (excluding the scalp, palms, soles) with a vIGA-AD score of '2' (mild) or '3' (moderate) for study entry

Up to ~50% of the subjects will be ≥18 years old

1.3. Schedule of Visits and Assessments

Study Procedure	Screen	Baseline Day 1	Wk 1 Day 8	Wk 2 Day 15	Wk 4 Day 29/ET
Visit	**1**	**2**	**3**	**4**	**5**
Visit Window	**-30 days**	**N/A**	**+/- 3 days**	**+/- 3 days**	**+/- 3 days**
Informed consent/assent	X				
Demographics	X				
Medical and surgical history	X				
Physical examination[a]	X	X			X
I/E criteria	X	X			
Hematology, Serum Chemistries, and Urine Analysis[b]	X[b]	X[b]			X[b]
Vital signs, height, weight[c]	X	X	X	X	X
vIGA-AD, EASI, BSA, SCORAD[d]	X	X	X	X	X
WI-NRS pruritus[e]	X	X	X	X	X
POEM[f]	X	X	X	X	X
Local Tolerability Assessment[g]		X	X	X	X
CDI-2, PHQ-8, PHQ-A, C-SSRS[h]	X	X	X	X	X
DLQI, CDLQI, DFI[i]	X	X	X	X	X
Medical Photography[j]		X	X		X
Serum pregnancy test (FOCBP only)	X				
Urine pregnancy test[k]		X	X	X	X
PK draws[l]					X
Drug/vehicle application in clinic[m]		X	X	X	
Dispense/Re-dispense study medication kit[n]		X	X[o]	X[o]	X[o]
Dispense/review diary	X	X	X	X	X
Weigh study medication kit[p]		X	X	X	X
Compliance determination[q]			X	X	X
Adverse event assessment[r]	X	X	X	X	X
Concomitant medications	X	X	X	X	X
Study Exit [s]					X

a Limited physical examination: skin (including assessment of Fitzpatrick skin type at Screening only), lungs, and heart only

b For all subjects entering this study under Amendment 2, to be collected at Screening, but subsequent samples will be collected only for subjects ≥12 years old (Baseline/Day 1 and Week 4/Day 29/ET). For subjects 12 to 18 years of age, if Baseline/Day 1 is within 3 weeks of Screening, the Screening results may be used.

Footnotes from table above:

c Height will be collected at Screening only. 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 other 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. **The vIGA-AD assessment should be completed prior to other physician assessments**. SCORAD total score will range between 0 and 103.

e Subjects will self-assess their pruritus at home on a daily basis starting 7 days prior to the Baseline/Day 1 visit, and then every day thereafter. WI-NRS score will be determined by the subject assessing worst itch over the past 24 hours. The scale is from 0 (no itch) to 10 (worst itch) and this value will be recorded by the subject each day. Subjects will be trained at the Screening visit in the accurate completion of the WI-NRS. In addition, parents/ caregivers of children and adolescent subjects will be trained at the Screening visit by study staff on how to assist the subject, if needed, in completing the WI-NRS.

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 Local tolerability assessments should be recorded prior to study drug application for the Investigator assessment of skin irritation (Berger and Bowman skin irritation score). **Note for investigator tolerability assessments: reactions at the site of product application, which may occur post-Baseline, should be differentiated from the preexisting inflammation associated with the subject's atopic dermatitis.** The subject will assess burning/stinging (0-3 score) 10-15 minutes post drug application. **Note subject burning stinging assessment: at Day 29, subjects will provide a recall assessment of burning/stinging experienced post drug application on the previous day (Day 28).**

h Adolescents and adults will complete the C-SSRS (12 years of age and older). Adults will complete the PHQ-8. Adolescents (ages 12 to 17, inclusive) will complete the PHQ-A (PHQ-9 modified). Parents/caregivers will complete CDI-2 (parent report) for children 6-11 years of age, inclusive.

i The DLQI will be completed by subjects ≥17 years of age. The CDLQI will be completed for subjects 6 to 16 years old, inclusive. The Dermatitis Family Impact Questionnaire (DFI) will be completed by parents/caregivers for all subjects 6 to ≤17 years of age.

j Photography of AD lesion(s) selected by the Investigator will be performed at all investigational sites. All efforts will be made to de-identify the subjects. Canfield equipment will be used to capture photographs. Subjects who are unwilling to participate in the medical photography will be allowed to opt out of this procedure, as documented on the Informed Consent Form.

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 entering this study under Amendment 2, a single PK trough draw will be collected at Day 29 only for subjects ≥12 years old (Baseline/Day 1 and Week 4/Day 29/ET). Ensure study medication was not applied in the area where PK will be drawn.

m Subjects to apply assigned IP during clinic visits, except for the Day 29/ET visit.

n It is expected that kits will be dispensed based on %BSA affected. See IP Handling Manual for details.

Footnotes from table above:

o On Day 8 and 15, dispensing of IP is optional. Site should review IP kit to ensure sufficient IP is available until the next visit and only dispense additional IP if needed. On Day 29, if the subject is unable to perform the Day 29 clinic visit due to COVID-19 restrictions (isolation, quarantine, etc.) then additional IP may need to be dispensed so IP can continue to be applied at home until the subject is able to return to the clinic to complete the Day 29 assessments (see IP Handling Manual for the process to dispense additional IP at or after Day 29).

p Every tube should be weighed and recorded when dispensed and returned. See IP Handling Manual for details.

q Compliance determination is described in the IP Handling Manual

r All AEs should be collected starting after the first application of the investigational product through the end of the study. All SAEs should be collected starting after the signing of the informed consent through 30 days after the last day of the application of the investigational product or the end of the study (whichever is later). Any AEs (whether serious or non-serious) and clinically abnormal laboratory test values(s) will be evaluated by the PI and treated and/or followed up for up to 30 days after end of treatment or until symptoms or value(s) return to normal, or acceptable level, as judged by the PI (if the subject is continuing into the ARQ-151-313 OLE study, then AEs from this study (ARQ-151-312) will only be followed until exit from this study).

s Subjects who enroll into the open label extension study (ARQ-151-313) must complete the ARQ-151-312 visit requirements at Week 4.

2. INTRODUCTION

2.1. Background

Roflumilast is a phosphodiesterase 4 (PDE-4) inhibitor approved globally to reduce the risk of exacerbations in patients with severe chronic obstructive pulmonary disease (COPD) associated with chronic bronchitis. Roflumilast and its active metabolite, roflumilast N-oxide, are high affinity selective inhibitors of PDE-4 (a major cyclic-3′,5′-adenosine monophosphate (cyclic AMP)-metabolizing enzyme), whose activity leads to accumulation of intracellular cyclic AMP. There are four different subtypes of PDE-4: PDE-4a, PDE-4b, PDE-4c, and PDE-4d, each with several isoforms (splicing variants). IC_{50} values of both roflumilast and roflumilast N-oxide for the different PDE-4 isoforms and subtypes are mostly sub-nanomolar and single digit nanomolar (Hatzelmann 2010). The PDE-4 family of enzymes are the most prevalent phosphodiesterases in immune cells and inhibition of PDE-4 subtypes has been associated with anti-inflammatory effects in many biological systems.

Roflumilast was initially developed as a 500 μg tablet for oral therapy in patients with COPD, and as such has been thoroughly evaluated in nonclinical studies. The safety profile is well-established. Oral roflumilast (500 μg tablet) was approved by Health Canada as DAXAS® in December 2010 and by the US FDA as DALIRESP® in February 2011 for the treatment of COPD. The study sponsor has conducted nonclinical studies in which roflumilast is applied dermally.

The dermal nonclinical program for ARQ-151 cream followed current International Conference on Harmonisation (ICH) guidelines and includes a 13-week dermal toxicity study in minipigs, a 13-week dermal toxicity study in mice, a 39-week dermal toxicity study in minipigs, a skin sensitization study in guinea pigs, a phototoxicity study, an eye irritation study and a 104-week carcinogenicity study in mice, the in-life portion of which is complete.

Refer to the current ARQ-151 Investigator's Brochure (IB) for the most current PDE-4 dermal and oral/systemic nonclinical and clinical information.

Atopic dermatitis is a chronic inflammatory skin disorder affecting children and adults, with the majority presenting with disease of mild to moderate severity. The use of topical corticosteroids and/or topical calcineurin inhibitors, in combination with emollients has been the mainstay for treating atopic dermatitis. In 2016, Eucrisa® (crisaborole), a PDE-4 inhibitor was approved for the topical treatment of mild to moderate atopic dermatitis in patients 2 years of age and older.

Topical calcineurin inhibitors block the activation of T-lymphocytes and diminish inflammation, but are accompanied by a boxed warning for the development of lymphomas and other lymphoproliferative diseases. Topical corticosteroids can cause skin atrophy and hypothalamic-pituitary axis suppression, and their use is often accompanied by poor adherence due to corticophobia (fear of using corticosteroids in patients or doctors). Eucrisa provides precedence for the effectiveness of topical PDE-4 inhibitors in atopic dermatitis, but it is a twice-daily ointment, its efficacy may be modest, and its use may be accompanied by burning, stinging, and local skin reactions. As a result, there is a need for the development of new topical products for the treatment of atopic dermatitis (Nygaard 2017).

The therapeutic use of PDE-4 inhibitors in AD is based on the recognized intracellular role of PDE-4 in keratinocytes (Dastidar 2007, Hanifin 1996). Circulating leukocytes in AD patients have PDE-4 activity, which has been associated with higher production of proinflammatory mediators and lower production of the anti-inflammatory mediator IL-10, in part due to hydrolyzation of cyclic adenosine monophosphate (cAMP) (Grewe 1982, Furue 2014, Baumer 2007). This consequently diminishes levels of cAMP, which leads to increased transcription of numerous cytokines, accelerating a number of intracellular functions involved in acute and chronic inflammation (Grewe 1982). Thus, targeting PDE-4 has been shown to directly attenuate inflammation due to inhibition of the breakdown of cAMP, consequently reducing the levels of tumour necrosis factor-α, IL-12, IL-23, and other signaling effectors (Murrell 2015, Nazarian 2009).

2.2. Conclusions on Toxicity Findings

The safety profile of oral roflumilast is well-established. An extensive systemic toxicity program that evaluated both roflumilast and its active N-oxide metabolite in multiple species via the oral route of administration was conducted to support registration of the 500 µg tablet for COPD.

The previously-conducted systemic toxicity program included studies to evaluate reproductive toxicity, genotoxicity and carcinogenicity, and the results of those studies are included in the labeling for oral roflumilast.

To support the development of ARQ-151 topical cream a GLP-compliant dermal toxicity program is being conducted. To date, no new risks have been identified through the dermal toxicity program. In 13-week dermal toxicity studies in minipigs and mice, and a 39-week dermal toxicity study in minipigs, no evidence of systemic toxicity was observed. Histopathological evaluation of skin in the minipig study included very slight to slight erythema at all treatment levels, and minor degrees of irritation such as hyperplasia. The NOAEL in both

studies was the 1% concentration of ARQ-151 (20 mg/kg), the highest dose administered and the maximum feasible concentration.

Across the dermal and systemic toxicology programs, the exposure to parent drug and N-oxide metabolite differs by route and species. While exposure to roflumilast and its active metabolite are likely to be higher following topical administration of ARQ-151 relative to oral administration, when the margins from the toxicity studies are considered as a whole, the NOAELs across routes and species provide assurance that the anticipated exposures with ARQ-151 cream will be safe.

2.3. Clinical Studies

2.3.1. Topical Roflumilast Cream

The formulation of topical roflumilast, ARQ-151 cream, has been evaluated in both plaque psoriasis (currently in Phase 3) and atopic dermatitis (through Phase 2).

2.3.2. Psoriasis Phase 2a (ARQ-151-101)

ARQ-151-101 (NCT03392168) was a Phase 2a study of two active doses of ARQ-151, 0.5% and 0.15% vs vehicle in the topical treatment of adult patients with chronic plaque psoriasis of up to 5% BSA involvement.

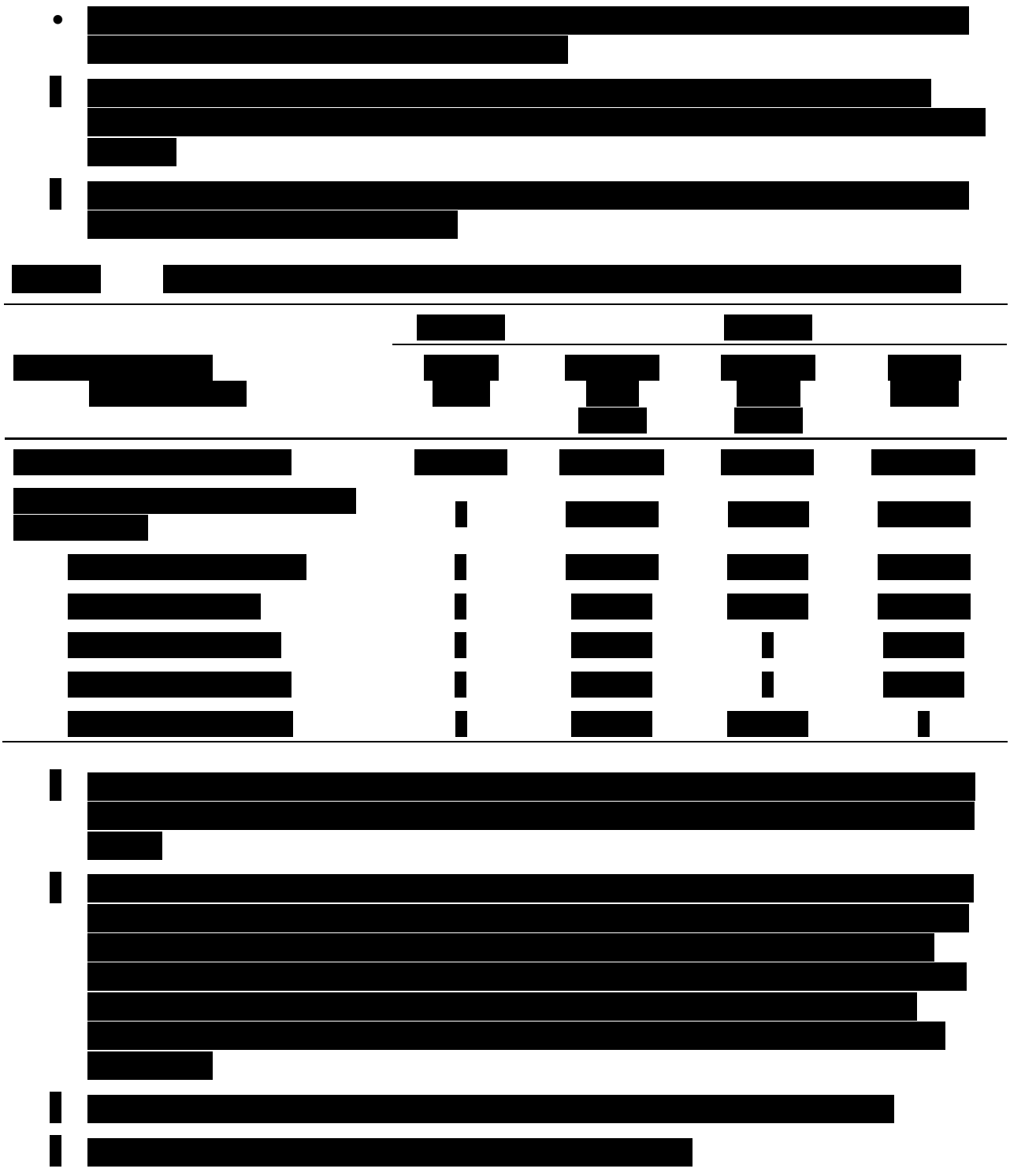
An initial cohort (Cohort 1) of 8 adult psoriasis subjects (no vehicle subjects) was treated as follows with the results indicated:

- Single dose application of ARQ-151 cream 0.5% to a 25 cm^2 area of psoriatic plaque on the trunk or extremities (not on the face, genital area, palms or soles)
- Local tolerability and systemic safety labs monitored
- PK assessments at baseline (pre-dose), 1, 2, 4, 6 and 24 hours
- Skin permeation of topically applied drug was ~0.4%
- Local tolerability and systemic safety labs were unremarkable

Six Cohort 1 subjects plus 83 additional psoriasis subjects were then enrolled into Cohort 2, an inter-individual, parallel group, randomized and blinded assessment of two concentrations of ARQ-151 drug product (0.15% and 0.5%) versus vehicle applied QD x 28 days, analyzing target psoriatic plaques for efficacy. Subjects were randomized 1:1:1 to receive 0.5% drug product, 0.15% drug product or vehicle to psoriatic plaques up to 5.0% of BSA. In each subject, up to 3 target plaques were identified for efficacy analysis.

PK assessments conducted on Day 1: 1, 2, 4 and 6 hours; Day 14: pre-dose (trough) and 1-hour post-dose; and Day 28: pre-dose (trough), 1, 2, 4, 6 and 24 hours.

Safety results follow:



Day 28 pharmacokinetic results of ARQ-151-101 are as follows:

- [redacted]
- [redacted]

2.3.3. Psoriasis Phase 2b (ARQ-151-201)

ARQ-151-201 (NCT03638258) was a parallel group, double blind, vehicle-controlled study in which ARQ-151 cream 0.15%, ARQ-151 cream 0.3%, or vehicle cream was applied QD for 12 weeks to over 300 adult subjects with 2% to 20% BSA of chronic plaque psoriasis and baseline IGA of Mild or greater. In this study, both ARQ-151 cream 0.3% and ARQ-151 cream 0.15% were safe and well tolerated, demonstrating similar safety and tolerability profiles compared to each other and compared to vehicle. The safety data are summarized below:

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

- [redacted]

Pharmacokinetic results of ARQ-151-201 are as follows:

- [redacted]

2.3.4. Atopic Dermatitis Phase 1 PK Study in Adults (ARQ-151-102)

ARQ-151-102 was an open label, Phase 1, pharmacokinetics and safety study of ARQ-151 Cream 0.15% and ARQ-151 Cream 0.05% administered QD in adult subjects with mild to moderate AD.

[redacted]

2.3.5. Phase 1 Study in Adolescents and Pediatrics (ARQ-151-105)

ARQ-151-105 (NCT04156191) is an ongoing open-label, Phase 1, pharmacokinetics, maximal usage PK, safety, and efficacy study of ARQ-151 cream 0.15% administered QD in adolescent and pediatric subjects with mild to moderate atopic dermatitis.

The study is being conducted in three parts, the first two of which are completed. The first part consisted of three cohorts in which subjects aged 2 to 17 years old had 1.5-35% BSA involvement (excluding the scalp, palms, soles) and mild or moderate atopic dermatitis based on vIGA-AD.

The second part of the study consisted of three cohorts in which subjects were evaluated under maximal use conditions (MUSE) and had BSA involvement (excluding the scalp, palms, soles) of ≥35% in subjects 2 to 11 years old (inclusive) or ≥25% in subjects 12 to <17 years old with mild or moderate atopic dermatitis based on vIGA-AD. At least 60% of the enrolled subjects had moderate atopic dermatitis.

The third part of the study will consist of one cohort (Cohort 7) in which subjects 2 to 5 years of age (inclusive) will be administered a lower concentration of ARQ-151 cream (0.05%) and evaluated under maximal use conditions (MUSE). Subjects will have BSA involvement (excluding the scalp, palms, soles) of ≥35% with mild or moderate atopic dermatitis based on vIGA-AD. At least 60% of the enrolled subjects had moderate atopic dermatitis.

Preliminary Study Results

[redacted]

2.3.6. Atopic Dermatitis Phase 2 Dose Ranging Study (ARQ-151-212)

ARQ-151-212 (NCT03916081) was a parallel group, double blind, vehicle-controlled, Phase 2 study that evaluated ARQ-151 cream 0.05% and 0.15% in the treatment of mild to moderate atopic dermatitis in 136 adolescent and adult subjects with 1.5 to 35% BSA of involvement.

Ninety-three female (68.4%) and 43 male (31.6%) subjects with mild to moderate AD participated in the study. Overall, the demographic and baseline disease characteristics were similar across all study groups. The mean age for all 136 study subjects was 41.6 years, including 8 adolescent subjects (between 12-17 years). The mean EASI score at Baseline for all study subjects was 9.04. The majority of subjects were in the moderate vIGA-AD category (77.9%). The mean BSA involvement was 9.5% for all study subjects.

[redacted]

2.3.7. Oral Roflumilast Tablet

Oral roflumilast (DALIRESP®) has been approved globally for the treatment of COPD and has been evaluated in nine Phase III/IV randomized double-blind clinical trials (Wedzicha 2016). Overall, the safety of oral roflumilast has been well established in its targeted population of mostly middle- and upper-aged individuals who currently smoke cigarettes or have smoked them extensively in the past. Adverse events (AEs) reported with roflumilast tablets have been consistent with those expected for oral PDE-4 inhibitors. In a pooled analysis of safety data from 6-month and 1-year clinical trials (N=8630), the most common AEs were diarrhea, weight loss and nausea. Other AEs reported more frequently with roflumilast treatment than with placebo were back pain, influenza, insomnia and decreased appetite (Michalski 2012, Wedzicha 2016).

In addition to the self-reported cases of weight loss in the 6-month and 1-year oral trials, clinically significant weight loss was also reported in two prospective studies that evaluated weight (Michalski 2012).

Psychiatric-related AEs were also greater in patients treated with roflumilast tablets (5.9%) compared to those treated with placebo (3.3%). The most common psychiatric-related AEs were insomnia, depression and anxiety. A small number of cases of completed suicide and suicide ideation have been reported in patients taking oral roflumilast in clinicals trials and also during post-marketing experience (Michalski 2012).

The only contraindication to oral roflumilast is use in patients with moderate to severe liver impairment (Child-Pugh B or C), where systemic levels of roflumilast may become elevated.

2.4. Rationale for Development

Atopic dermatitis is currently treated with topical calcineurin inhibitors and/or topical corticosteroids in combination with emollients. In 2016, Eucrisa® (crisaborole), a less potent PDE-4 inhibitor than roflumilast, was approved for the topical treatment of atopic dermatitis. Topical calcineurin inhibitors block the activation of T-lymphocytes and diminish inflammation, but are accompanied by a ‘black box’ warning for the development of lymphomas and other lymphoproliferative diseases. Topical corticosteroids can cause skin atrophy and hypothalamic-pituitary axis suppression, and their use is often accompanied by poor adherence due to corticophobia (fear of using corticosteroids in patients or doctors). Eucrisa provides precedence for the effectiveness of topical PDE-4 inhibitors in atopic dermatitis, but it is a twice-daily

ointment, its efficacy is modest, and its use is often accompanied by burning, stinging, and local skin reactions. In our Phase 2 AD study (ARQ-151-212), we observed excellent local toleration of ARQ-151 cream formulations. Since roflumilast is a more potent PDE-4 inhibitor than crisaborole (Hatzelmann 2010), we believe that ARQ-151 has potential to provide greater efficacy with better local toleration than Eucrisa.

This study will evaluate the safety and efficacy of ARQ-151 cream in children, adolescent, and adult subjects with mild to moderate atopic dermatitis.

2.4.1. Dose Selection

In ARQ-151-212, results for the primary efficacy endpoint, mean absolute change from baseline in EASI score at Week 4, were numerically higher in the ARQ-151 cream 0.05% and ARQ-151 cream 0.15% (p= 0.097) groups than in the vehicle group. Furthermore, the result of the sensitivity analysis of the primary endpoint at Week 4 was statistically significant (ARQ-151 cream 0.15%, p= 0.027). Statistical significance was reached for numerous other clinically important efficacy endpoints including % change from baseline in EASI score, EASI-75 responders, and patients achieving vIGA-AD score of clear or almost clear. Both doses of topical roflumilast (0.15% and 0.05%) had a similar and favorable safety and tolerability profile, with generally more favorable efficacy observed at the ARQ-151 cream 0.15% dose. Results of this study support the use of ARQ-151 cream 0.15% in studies of adult and adolescent subjects with mild to moderate AD. PK data from ARQ-151-105 support the use of the same concentration (0.15%) in subjects 6 to 11 years old and the use of a lower concentration in subjects 2 to 5 years old.

2.4.2 Risks and/or Benefits to Subjects

A favorable local and systemic benefit-risk profile has been observed in prior studies of ARQ-151. Subjects 6 years of age and older, included in this study, randomized to active treatment group may see an improvement in their atopic dermatitis with ARQ-151 0.15% cream, based on the activity of doses tested in atopic dermatitis (0.05%-0.15%) and psoriasis (0.15%-0.5%), and approval of a less potent topical PDE-4 inhibitor (crisaborole) for atopic dermatitis. Subjects may also see some benefit as the cream formulation of ARQ-151 may have a moisturizing effect.

Oral roflumilast has now been used for almost a decade in the treatment of COPD exacerbations and its safety record has been well-documented. The known adverse effects of oral treatment in the COPD population (nausea, vomiting, diarrhea, weight loss, psychiatric AEs (see Section 2.3.7) may be readily monitored as specified in this protocol. The profile that is emerging from studies of topical roflumilast appears different from the safety and tolerability profile of oral roflumilast. While oral PDE-4 inhibitors (DALIRESP, OTEZLA) have been associated with, in particular, a moderate incidence of GI AEs, these AEs, and perhaps others, appear to be reported far less frequently with topical PDE-4 inhibitors, including EUCRISA®, and ARQ-151 cream to date in clinical trials. For ARQ-151 cream, this may be related to the lack of 'peak to trough' C_{max} variation, lower C_{max} values than observed following oral administration, or bypassing of the gastrointestinal tract with topical administration.

This study has been designed with adequate safety monitoring practices (i.e., physical examinations, vital signs/weight, local skin toleration assessments, hematology, serum chemistry, urinalysis, PHQ-A/PHQ-8, CDI-2, C-SSRS and AE reporting).

3. STUDY OBJECTIVES AND ENDPOINTS

3.1. Study Objectives

3.1.1. Primary Objective

The primary objective of this study is to assess the safety and efficacy of ARQ-151 cream 0.15% vs vehicle administered QD x 4 weeks to individuals 6 years of age and older with atopic dermatitis.

3.2. Study Endpoints

3.2.1. Primary Efficacy Endpoint

The primary efficacy endpoint of this study is:

- IGA Success, defined as a vIGA-AD score of ‘clear’ or ‘almost clear’ PLUS a 2-grade improvement from Baseline at Week 4

3.2.2. Secondary Efficacy Endpoints

The secondary efficacy endpoints of this study are:

- In subjects with a vIGA-AD score of ‘Moderate’ at randomization, vIGA-AD Success at Week 4
- In subjects ≥12 years old with baseline WINRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 4
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 2
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 1
- Achievement of at least a 75% reduction in the Eczema Area and Severity Index (EASI-75) at Week 4
- vIGA-AD of ‘clear’ or ‘almost clear’ at Week 4
- vIGA-AD Success at Week 2
- vIGA-AD Success at Week 1
- vIGA-AD of ‘clear’ or ‘almost clear’ at Week 2
- vIGA-AD of ‘clear’ or ‘almost clear’ at Week 1

3.2.3. Pharmacokinetic Endpoint

- Pharmacokinetic endpoints include concentrations of roflumilast and its N-oxide metabolite.

4. INVESTIGATIONAL PLAN

4.1. Overall Study Design and Plan

This is a Phase 3, parallel group, double blind, vehicle-controlled study in which ARQ-151 cream 0.15% or vehicle is applied QD x 4 weeks to subjects with mild to moderate atopic dermatitis.

- At entry, subjects will have ≥3% BSA involvement (excluding the scalp, palms, soles) and mild or moderate atopic dermatitis (AD) based on vIGA-AD assessment.
- Upon determination of eligibility, subjects will be randomized 2:1 to either ARQ-151 cream 0.15% or matching vehicle cream. The randomization will be stratified by vIGA-AD score at baseline ('Mild' vs. 'Moderate') and by study site.
- Subjects/caregivers will apply ARQ-151 cream 0.15% or vehicle cream QD to all AD affected areas and any newly appearing AD lesions that arise during the study, except on the scalp. Subjects/caregivers should maintain treatment of these areas with study drug for the duration of the study regardless of whether treatable areas of AD clear prior to Week 4.
- At the Week 4 visit, subjects may be eligible to enroll in a 12-month, open label extension study (ARQ-151-313) in which they will receive ARQ-151 cream 0.15% QD.

4.2. Number of Sites and Subjects

A total of up to approximately 650 subjects will be randomized at approximately 50 study sites in the United States and Canada. Only subjects 18 years and older will be enrolled at sites located in the province of Québec in Canada. During the conduct of the study, additional countries and/or sites may be added if necessary. Subjects will be male and female children and adolescents (6-17 years), and adults (≥18 years). Subjects will have mild to moderate atopic dermatitis involvement with a vIGA-AD score of '2' (mild) or '3' (moderate) for study entry. Up to ~50% of the subjects will be ≥18 years old and up to 25% of subjects will have a vIGA-AD score of '2'.

4.3. Subject Participation

Subject participation involves a minimum of 5 clinic visits including Screening, Baseline, Week 1, Week 2, and Week 4. The interval between the Screening and Baseline visits could be up to 30 days, therefore the anticipated maximum duration of subject participation is approximately 8 weeks.

Upon completion of the treatment phase of the study (Week 4) subjects may have the opportunity, subject to regulatory approval and enrollment has not been completed, to participate in an open-label extension study (ARQ-151-313) of up to 12 months.

4.4. Numbering of Subjects

All subjects who sign an informed consent form will be assigned a unique 6-digit subject identification (ID) number by the IWRS system.

The subject identifier number is 6-digits (SXX-YYY) and will contain the study number-site number (where S = 2 and XX is this study site number such as 01, 02, etc.) and the subject number (YYY). It will be assigned in numerical order at the screening visit based on chronological order of screening dates.

Individuals who do not meet the criteria for participation in this trial (screen failure) may be rescreened once, if deemed acceptable by the investigator. Rescreened subjects should be assigned a different subject number than the initial screening. All procedures planned at the screening visit, including signature of a new consent form, will be performed.

In the case of a screening laboratory value abnormality, the test can be repeated once within the original screening time window, if the investigator believes there is a reasonable possibility that the subject would be eligible if re-tested. This would not be considered a screen failure and a new subject number would not be assigned.

The clinical site is responsible for maintaining a current log of subject ID number assigned to each subject. The subject ID number will be used to identify the subject throughout the study and is required to be entered on all clinical study documentation (eg, case report forms, labeling of clinical materials and sample containers, investigational product accountability logs, etc.).

4.5. Selection of Study Population

4.5.1. Inclusion Criteria

Subjects must fulfill all of the following inclusion criteria to be eligible for participation in the study:

1. For adult subjects: Participants legally competent to sign and give informed consent. For pediatric and adolescent subjects: Informed consent of parent(s) or legal guardian, and, if age appropriate, assent by the subjects, as required by local laws.
2. Males and females, ages 6 years and older at time of signing Informed Consent (Screening). Only subjects 18 years and older will be enrolled at sites located in the province of Québec in Canada.
3. Diagnosed with mild to moderate atopic dermatitis according to the criteria of Hanifin and Rajka (1980) prior to or at the screening visit. Subjects must have at least 3 of the 4 basic features per Hanifin and Rajka (1. Pruritus; 2. Typical morphology and distribution [flexural lichenification in adults and facial and extensor eruptions in infants and children]; 3. Chronic or chronically relapsing dermatitis; or 4. Personal or family history of atopy), in addition to 3 or more minor criteria.

4. History of AD for at least 3 months in subjects 6-17 years of age or 6 months in subjects ≥18 years of age, as determined by the Investigator using information from the subject's medical chart, from the subject's physician, or through subject/parent/caregiver interview. Stable disease for the past 4 weeks with no significant flares in atopic dermatitis before screening.
5. EASI Score ≥5 at Baseline. EASI is evaluated for the entire body except the scalp, palms, and soles.
6. vIGA-AD score of 'Mild' ('2') or 'Moderate' ('3') at Baseline/Day 1. The vIGA-AD is evaluated for the entire body except the scalp, palms, and soles.
7. Has AD involvement of ≥3% BSA (excluding the scalp, palms, soles) at Baseline.
8. Females of childbearing potential (FOCBP) must have a negative serum pregnancy test at Screening (Visit 1) and a negative urine pregnancy test at Baseline/Day 1. In addition, sexually active FOCBP must agree to use at least one form of a highly effective or barrier method of contraception throughout the trial. The use of abstinence as a contraceptive measure is acceptable as long as this is a consistent part of a lifestyle choice and an acceptable backup method has been identified if the subject becomes sexually active.
9. Females of non-childbearing potential should either be pre-menarchal, or post-menopausal with spontaneous amenorrhea for at least 12 months (post-menopausal status should be confirmed with FSH testing) or have undergone surgical sterilization (permanent sterilization methods include hysterectomy, bilateral oophorectomy, or bilateral salpingectomy).
10. In good health as judged by the Investigator, based on medical history, physical examination, vital signs, serum chemistry labs, hematology values, and urinalysis.
11. Subjects and parent(s)/legal guardian(s) are considered reliable and capable of adhering to the Protocol and visit schedule, according to the judgment of the Investigator.

4.5.2. Exclusion Criteria

Subjects who meet any of the following exclusion criteria will be excluded from participation in this study:

1. Subjects with any serious medical condition or clinically significant laboratory, vital signs, or physical examination abnormality that would prevent study participation or place the subject at significant risk, as judged by the Investigator
2. Liver function tests excursions that exceed:
   - AST or ALT >2X ULN
   - Total bilirubin:
     - >1.5 x ULN or
     - > ULN and ≤1.5 x ULN AND direct bilirubin is >35% of total bilirubin
   - ALP ≥2x ULN

3. Subjects who cannot discontinue medications and treatments prior to the Baseline visit and during the study according to Excluded Medications and Treatments (Table 2).
4. Has unstable AD or any consistent requirement for high potency topical steroids to manage AD signs or symptoms.
5. Subjects who have significant active systemic or localized infection (eg, molluscum contagiosum), including known actively infected AD, or have had any infection that required oral or intravenous administration of antibiotics, antifungal or antiviral agents within 14 days prior to Baseline/Day 1.
6. Subjects who are unwilling to refrain from prolonged sun exposure and from using a tanning bed or other artificial light emitting devices (LEDs) for 4 weeks prior to Baseline/Day 1 and during the study.
7. Subjects with skin conditions other than AD that would interfere with evaluations of the effect of the study medication on AD, as determined by the Investigator. Subjects with any condition on the treatment area which, in the opinion of the Investigator, could confound efficacy measurements, eg, molluscum contagiosum.
8. Subjects with known genetic dermatological conditions that overlap with AD, such as Netherton syndrome.
9. Known allergies to excipients in ARQ-151 cream [redacted]
10. Subjects who cannot discontinue the use of strong cytochrome P-450 CYP3A4 inhibitors eg, indinavir, nelfinavir, ritonavir, clarithromycin, itraconazole, ketoconazole, nefazodone, saquinavir, suboxone and telithromycin for 2 weeks prior to the Baseline/Day 1 and during the study period.
11. Subjects who cannot discontinue the use of strong cytochrome P-450 CYP3A4 inducers eg, efavirenz, nevirapine, glucocorticoids, barbiturates (including phenobarbital), phenytoin, rifampin, and carbamazepine for 2 weeks prior to the Baseline/Day 1 and during the study period.
12. Subjects who have received oral roflumilast (Daxas®, Daliresp®) within the past 4 weeks.
13. Known or suspected:
    - Severe renal insufficiency
      - Severe renal insufficiency is defined as calculated creatinine clearance <30 mL/min.
    - Moderate to severe hepatic disorders (Child-Pugh B or C)
14. History of severe depression, suicidal ideation or behavior, Baseline/Screening C-SSRS (for adolescents and adults 12 years old and older) indicative of suicidal ideation or behavior, whether lifetime or recent/current.

15. Subjects with a PHQ-8 (adults) or modified PHQ-A (adolescents, 12-17 years old inclusive) score ≥10 at Screening or Baseline visits
16. Subjects (6 to 11 years old, inclusive) with a CDI-2 (parent report) raw score ≥17 for females and ≥18 for males at Screening or Baseline/Day 1 visits
17. Females who are pregnant, wishing to become pregnant during the study, or are breast-feeding.
18. Previous treatment with ARQ-151.
19. Subjects currently undergoing allergy testing (eg, food allergy testing or skin prick testing), patch testing, food challenges, or allergy desensitization, or plan to do so during the study.
20. Subjects with any serious medical condition (eg, uncontrolled hypo- or hyper-thyroidism) or clinically significant laboratory abnormality that would prevent study participation or place the subject at significant risk, as determined by the Investigator.
21. Subjects with a history of a major surgery within 4 weeks prior to Baseline/Day 1 or subjects who have a major surgery planned during the study.
22. Subjects with a history of chronic alcohol or drug abuse within 6 months prior to Screening.
23. Current or a history of cancer within 5 years with the exception of fully treated skin basal cell carcinoma, cutaneous squamous cell carcinoma or carcinoma in situ of the cervix.
24. Parent(s)/legal guardian(s) who are unable to communicate, read, or understand the local language. Subjects who are unable to communicate, read or understand the local language(s), or who display another condition, which in the Investigator's opinion, makes them unsuitable for clinical study participation.
25. Subjects who are family members of the clinical study site, clinical study staff, or sponsor, or family members of enrolled subjects (subjects enrolled in other studies of ARQ-151) living in the same house.

4.6. Randomization

Randomization will take place at the Baseline visit after the Investigator confirms that the subject meets all eligibility criteria listed in Section 4.5.

Subjects will be randomly assigned to apply ARQ-151 cream 0.15% QD, or matching vehicle QD. Assignment of drug or vehicle will be made at a 2:1 ratio (drug:vehicle) and stratified by vIGA-AD score ('Mild' vs. 'Moderate'), and by study site according to a computer-generated randomization list. Kits containing tubes of study medication will be assigned to each subject using an internet-based response system (IWRS). A subject may receive more than one kit for the treatment period. The kits and tubes are blinded and each kit is numbered with a unique kit number.

4.7. Study Restrictions

4.7.1. Prohibitions and Concomitant Therapy

Prohibited medications and products are detailed in Table 2 (Excluded Medications and Treatments).

Generally, the addition of new medications, including nonprescription medications, during the course of the study is discouraged. However, the short-term use of a medication may be authorized by the Investigator. The Investigator must make the decision to authorize the use of any such a medication only after consideration of the clinical situation, the potential for masking symptoms of a more significant underlying event, and whether the use of the medication will compromise the outcome or validity of the clinical investigation. Other medications may be authorized by the Investigator for conditions other than AD. If medication is required, the name, strength, frequency, duration of use, and reason for use will be recorded in source documents and entered into the CRFs. Medications which have been used chronically by subjects, in particular statins and anti-hypertensives, are allowed for use during the study, except as prohibited in Table 2. No rescue medication for AD is allowed during this study up to Week 4.

Table 2: Excluded Medications and Treatments

Excluded Medications and Treatments	Washout Period Prior to Day 1
Approved biologics such as dupilumab	6 months
Investigational biologics	6 months
• Systemic treatments that could affect AD; eg, corticosteroids, retinoids, calcineurin inhibitors, hydroxycarbamide (hydroxyurea), methotrexate, cyclosporine, azathioprine, hydroxychloroquine, mycophenolate mofetil, or other immunosuppressive therapies, or systemic treatment with nonsedating antihistamines in a nonstable regimen. • Systemic treatments with nonsedating antihistamines (eg, cetirizine, desloratadine, loratadine) in a stable regimen is allowed.	4 weeks or 5 half-lives, whichever is longer
PUVA or NBUVB phototherapy	4 weeks
Topical products containing urea	1 week
Sedating antihistamines and other over the counter remedies containing sedating antihistamine, such as sleep aids (eg, ZzzQuil™ LIQUICAPS® SLEEP-AID), and cough/cold remedies (eg, Theraflu® night time, NyQuil™ Cold & Flu Night time)	1 week
Topical corticosteroids, calcineurin inhibitors, or Eucrisa®. Topical antibacterial medications or products, including soaps, dilute bleach baths, or sodium hypochlorite-based products anywhere on the body.	2 weeks
Strong cytochrome P-450 CYP3A4 inhibitors eg, indinavir, nelfinavir, ritonavir, clarithromycin, itraconazole, ketoconazole, nefazodone, saquinavir, suboxone and telithromycin	2 weeks

Table 2: Excluded Medications and Treatments (Continued)

Excluded Medications and Treatments	Washout Period Prior to Day 1
Strong cytochrome P-450 CYP3A4 inducers eg, efavirenz, nevirapine, glucocorticoids, barbiturates (including phenobarbital), phenytoin, rifampin, and carbamazepine	2 weeks
Systemic antibiotics	2 weeks
Tanning beds, other light emitting devices	4 weeks
Oral roflumilast (Daxas®, Daliresp®)	4 weeks
All other investigational drugs	4 weeks or 5 half-lives, whichever is longer

- Eye / ear drop and nasal corticosteroid preparations are allowed. Inhaled corticosteroid preparations are allowed if used for a stable condition and at a stable dose for >28 days before screening and are continued at the same dose throughout the study.
- All subjects should apply medication each evening, except on clinic visit days when the study product will be applied at the clinical site. If the subject takes an evening shower/bath, the ARQ-151 cream or vehicle can be applied as soon as the skin is nearly dry, but no later than 20 minutes before going to bed. Subjects are not to wash areas where ARQ-151 cream or vehicle has been applied until at least 4 hours after study drug application. Non-medicated emollients or moisturizers will be allowed once daily in a stable regimen as normally used by the subjects. For subjects that apply a non-medicated emollient or moisturizer after an evening shower/bath, the study drug must be applied first to the treatment areas. The non-medicated emollient or moisturizer can then be applied but only to other untreated areas of the subject's skin.
- Sunscreens will be allowed daily, as needed by the subjects when applied at least 2 hours after application of randomized study drug.
- Concomitant other medications for chronic conditions (eg, NSAIDs, statins, anti-hypertensives) are permitted unless specifically prohibited in the Protocol.
- Topical antibiotics, topical antihistamines, or any other topical agents are not allowed to be applied to treated areas.

4.8. Treatment

4.8.1. Drug Supplies, Packaging and Labeling

ARQ-151 cream or matching vehicle will be supplied in 45 gram tubes. The tubes will be packaged in kits, containing multiple tubes of investigational product. The number of kits dispensed to a subject will be based on the BSA involvement of atopic dermatitis. The kits and tubes will be labeled in a blinded manner. The kit(s) dispensed to a subject will be labeled with a unique number.

The Sponsor will supply sufficient quantities of the study drug and matching vehicle to each site to allow for completion of this study.

Records will be made of the receipt and dispensing of the study drugs supplied. At the conclusion of the study, any used/unused study drugs will be returned to the Sponsor or designee, or destroyed, as per Sponsor instructions.

Refer to the most current version of the IP Handling Plan for details on accountability, storage, and management of ARQ-151.

4.8.2. Blinding

This is a double-blind study, therefore neither the subjects nor the Investigator and clinical personnel will be aware of which treatment an individual subject receives.

4.8.3. Breaking Treatment Codes

In the event of a medical emergency where breaking the blind is required to provide medical care to the subject, the Investigator may obtain treatment assignment directly from the IWRS system for that subject. Refer to the current version of the ARQ-151-312 IWRS User Manual for details on unblinding. Treatment assignment should, however, remain blinded unless the assignment knowledge is necessary to determine subject emergency medical care. The rationale for unblinding must be clearly explained in source documentation and on the CRF, along with the date on which the treatment assignment was obtained. The investigator is requested to contact the Medical Monitor promptly in the event of any treatment unblinding.

Blinding of study treatment is critical to the integrity of this clinical trial and therefore, if a subject's treatment assignment is disclosed to the Investigator, the subject will have the study treatment discontinued. All subjects will be followed until study completion unless consent to do so is specifically withdrawn by the subject.

4.8.4. Treatment Administration

Initial treatment with the IP will occur on Day 1. ARQ-151 cream 0.15% is administered once daily as a topical product to cover the skin surface at an application rate of approximately 2 mg/cm^2.

At Baseline visit, the study staff will demonstrate to the subject/caregiver(s) how to apply ARQ-151 cream using the first tube from the kit that is assigned to the subject. Study site staff will be trained to ensure a unit dose (a pea size unit of ARQ-151 cream will cover approximately 1% BSA) is properly squeezed from the tube and applied to atopic dermatitis lesion(s) as a thin film and rubbed in using the index and middle finger, rubbing in thoroughly but gently, until the 'white' has disappeared. The subject/caregiver will then practice squeezing a similar amount onto their index and middle finger and apply a thin film to other areas to be treated. At Baseline/Day 1, the study staff will ensure that the subject/caregiver's application technique is correct and that a thin layer is applied as instructed (which represents an application rate of approximately 2 mg/cm^2).

Re-training will be conducted at subsequent visits as needed (i.e., if the returned tube(s) weighs substantially different than the expected weight).

Subjects/caregivers will be instructed to apply investigational product once daily to all treatment areas identified by the Investigator at Baseline using a Body Diagram (see Appendix 1).

Note:

- All subjects should apply medication each evening (except on clinic visit days when the investigational product will be applied at the clinic). If the subject takes an evening shower/bath, the ARQ-151 cream or vehicle can be applied as soon as the skin is nearly dry, but no later than 20 minutes before going to bed. Subjects are not to wash areas where ARQ-151 cream or vehicle has been applied until at least 4 hours after study drug application.
- Caregivers should wash their hands with soap and water after applying IP to a child.
- Parents/guardians/caregivers who are pregnant, or women of childbearing potential who are trying to become pregnant, or who are breastfeeding, or planning to breastfeed during the study should avoid accidental exposure by either avoiding applying investigational product or by wearing gloves during its application.
- Subjects should maintain treatment of areas with study drug for the duration of the study regardless of whether treatable areas of AD clear prior to Week 4 visit.
- New lesions that develop during the study should be treated (except scalp). An unscheduled visit is not required for starting treatment of new lesions.

Each investigational product tube will be weighed prior to dispensing at the Baseline visit and at each subsequent visit. Investigational product tubes must be returned by subjects at each study visit, both empty and full, and will be weighed. If the subject's actual use is substantially different than the expected use for the subject's BSA (see IP Handling Plan), the subject/ caregiver will be retrained on the study drug application technique.

4.8.5. Treatment Compliance

Investigational product tubes will be weighed at each clinic visit.

Subjects/caregivers will complete a daily diary recording the date and time of each dose applied, any missed doses, and a comment section should the subjects have a comment, eg, record potential AEs. Site personnel will review the diaries and use the information to question the subject regarding compliance and AEs and then record appropriate information in source documents and complete Case Report Forms (CRFs). If a subject misses a dose, they should be instructed to return to the protocol investigational product administration schedule (i.e. if subject forgets a dose they should wait until that evening and apply as usual).

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

Compliance will be assessed by review of the dosing diary. Weight of investigational product applied (via dispensed and returned tube weights) will be measured for reporting purposes.

If the diary shows less than 80% of expected daily applications (but not more than 3 consecutive missed doses), the subject is using too little study drug and retraining must be conducted and documented.

Compliance will be documented in source and in eCRF.

4.8.6. Removal of Subjects from Study Treatment

Subject treatment with study drug in this trial may be discontinued for any of the following reasons:

1. Occurrence of any medical condition or circumstance that, in the opinion of the Investigator does not allow the subject to adhere to the requirements of the Protocol.
2. Adverse Events as described in Section 5.1.12 and Section 5.9. The Investigator must follow the subject until the AE resolves or satisfactorily stabilizes.
3. Treatment must be discontinued immediately in the event of a female subject's pregnancy.
4. Subject's decision to discontinue treatment with study drug.
5. C-SSRS (Section 5.1.10) indicative of suicidal ideation.
6. PHQ-8 (Section 5.1.7) or modified PHQ-A (Section 5.1.8) score ≥15 if determined by Investigator in consultation with mental health professional.
7. CDI-2 (Section 5.1.9) raw total score of ≥32 if determined by Investigator in consultation with mental health professional.
8. Requirement for use of prohibited concomitant medication after consultation with the Sponsor and Medical Monitor.
9. Subject's repeated failure to comply with protocol requirements or study related procedures.
10. The subject interrupts trial study drug application for more than 50% of scheduled doses.
11. Termination of the study by the Sponsor, FDA, or other regulatory authorities.

4.8.7. Removal of Subjects from the Study

A subject may be removed from study participation for any of the following reasons:

- Subject death.
- Subject's decision to withdraw from the study.
- Subject is lost to follow-up. A subject will be considered lost to follow-up after three phone and three email attempts and documentation of a certified letter sent to the subject's address.
- Termination of the study by the Sponsor, FDA, or other regulatory authorities.

5. STUDY PROCEDURES

The Schedule of Visits and Assessments (Section 1.3) summarizes the clinical procedures to be performed at each visit. Individual clinical procedures are described in detail below.

5.1. Safety Assessments

This study assesses the safety and efficacy of ARQ-151 cream. Safety will be determined by evaluating physical examinations, local tolerability assessments, vital signs/weight, clinical laboratory parameters, either PHQ-8 (adults, ≥18 years old) or modified PHQ-A (adolescents, 12-17 years old) or Children's Depression Inventory 2 (CDI-2, parent report for children 6-11 years old, inclusive), C-SSRS (12 years and older) and AEs as outlined in the Schedule of Visits and Assessments (Section 1.3).

Additional evaluations/testing may be deemed necessary by the PI and/or the Sponsor for reasons related to subject safety.

5.1.1. Screening

Before a subject's participation in the clinical study, the Investigator is responsible for obtaining written informed consent from the subject or written assent from adolescent subjects and consent from their parent(s) or legal guardian(s) for children and/or adolescents after adequate explanation of the study design, anticipated benefits, and the potential risks. A subject is considered a participant of the trial once the ICF or written assent for adolescent subjects is completely signed.

Subjects must provide informed consent/assent as per their age group at screening. During the study, if a subject changes age group, the subject must provide informed consent/assent relative to his/her current age group. Subjects will continue with the assessments specific to their age group at the time of consent/assent at Screening.

The following procedures/assessments will be performed at the Screening Visit (within 4 weeks after signing the informed consent):

- Review of medical and surgical history
- Collection of demographic data including sex, age, race, ethnicity
- Vital signs including temperature, heart rate, and blood pressure
- Collection of body weight (kg), and height (cm)
- Atopic dermatitis assessments (eg, vIGA-AD, BSA, EASI, SCORAD)
- Limited physical examination of skin (including assessment of Fitzpatrick skin type at Screening only), lungs, and heart
  - Fitzpatrick skin phototype will be rated as follows:
    - I: Always burns easily; never tans (sensitive)
    - II: Always burns easily; tans minimally (sensitive)
    - III: Burns moderately; tans gradually (light brown) (normal)

IV: Burns minimally; always tans well (moderate brown) (normal)

V: Rarely burns; tans profusely (dark brown) (insensitive)

VI: Never burns, deeply pigmented (insensitive)

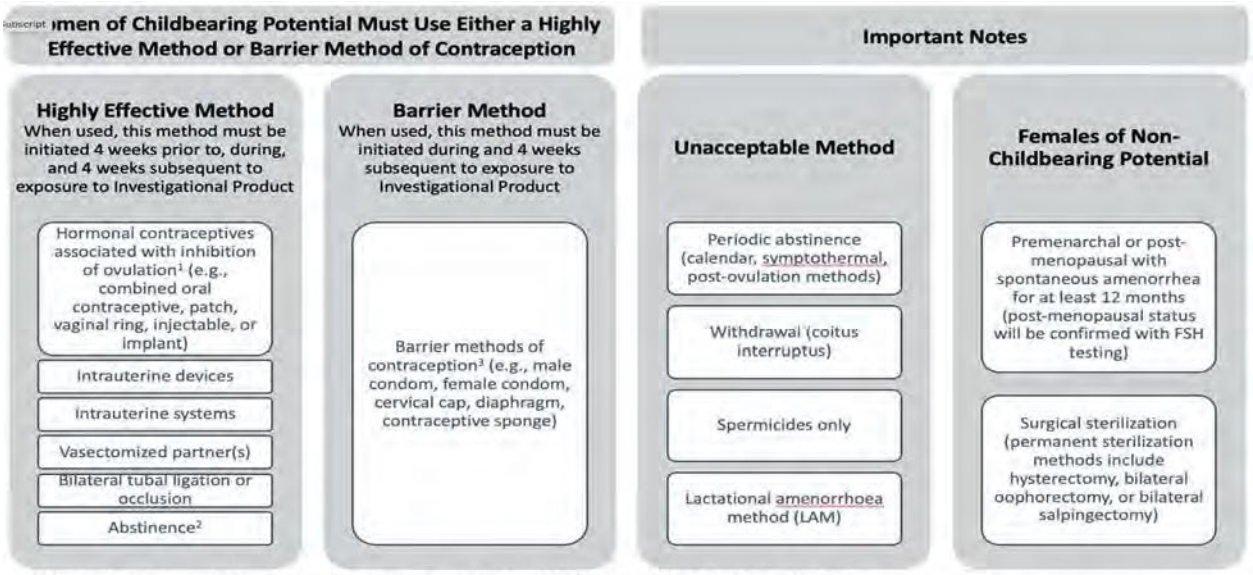
- Laboratory tests: hematology, chemistry, urinalysis, serum pregnancy test (for female subjects of child bearing potential)
- Completion of WI-NRS, CDI-2, DLQI, CDLQI, DFI, C-SSRS, POEM, and PHQ (-8 or -A)
- Collection of concomitant medications and adverse events

Subjects may be re-screened one time and the subject will be assigned a new Subject ID.

5.1.2. Contraception Requirements

Females of childbearing potential (FOCBP) must have a negative serum pregnancy test at Screening and a negative urine pregnancy test at Baseline (Day 1). In addition, sexually active FOCBP must agree to use at least one form of a highly effective or barrier method of contraception throughout the trial according to Contraception Requirements (Figure 1).

Figure 1: Contraception Requirements for Female Subjects



5.1.3. Baseline (Day 1)

Randomization will take place at the Baseline visit after the subject has been found to be fully eligible for participation. A subject will be considered enrolled into the study upon the first IP application.

If the Baseline visit occurs within 21 days of Screening for subjects 12 to 18 years old, the Screening laboratory test results may be used.

5.1.4. Physical Examination

Physical examinations will be performed according to the Schedule of Visits and Assessments (Section 1.3). The physical exam will be limited to skin, lungs and heart only.

5.1.5. Vital Signs, Height and Weight

Vital signs will be performed according to the Schedule of Visits and Assessments (Section 1.3). Blood pressure, heart rate, and temperature will be collected in seated position after 5 mins of rest. For weight measurement, subjects will be instructed to void prior to weight being taken and to remove any objects of significant weight (eg, jackets, outerwear, shoes, cell phones, wallet, key chains, etc.). Weight should be obtained using a calibrated weight scale and the same scale, whenever possible, should be used for a subject throughout the duration of the study. 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 (eg, 55.5 pounds or 25.1 kilograms). For subjects <18 years of age, measure the weight in triplicate and report the average weight in EDC. An unexplained, clinically significant weight loss should be reported to the Medical Monitor.

Height will be measured at Screening only.

5.1.6. Laboratory Tests

All tests listed in Table 3 below will be performed according to the Schedule of Visits and Assessments (Section 1.3) unless otherwise noted. Hematology, serum chemistries, and urine analysis will be collected for all subjects at Screening, but subsequent samples will be collected only for subjects ≥12 years old (Baseline/Day 1 and Week 4/Day 29/ET). No food restrictions are required for the collection of specimens. In addition, laboratory safety tests may be performed at various unscheduled time points, if deemed necessary by the Investigator.

Table 3: Laboratory Tests

Hematology	Serum Chemistry
Hemoglobin Hematocrit Total and differential leukocyte count Red blood cell count with indices and morphology Platelet count	Blood Urea Nitrogen Bilirubin (total and direct) Alkaline phosphatase Aspartate aminotransferase Alanine aminotransferase Albumin Sodium Potassium Chloride Glucose Creatinine
Urinalysis	**Additional Tests**
pH Specific gravity Protein* Glucose Ketones Bilirubin Blood* Nitrite* Urobilinogen Leukocyte esterase*	Urine pregnancy test** (for females of child bearing potential only) Serum pregnancy test (hCG)*** FSH test, (post menopausal) ***

* If urinalysis is positive for protein, blood, nitrite and/or leukocyte esterase, a microscopic examination (for red blood cells, white blood cells, bacteria, casts, and epithelial cells) will be performed.
** At Baseline and Weeks 1, 2, and 4 for FOCBP only
*** At screening only

5.1.7. Patient Health Questionnaire Depression Scale (PHQ-8)

The PHQ-8 Assessment (Appendix 2) will be performed in adult subjects according to the Schedule of Visits and Assessments (Section 1.3).

PHQ-8 score is the sum of the responses for the 8 questions.

Five severity categories of depression are defined 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)

A subject with a PHQ-8 score of 10-14 should be referred to a mental health professional for evaluation, regardless of the reason for their change in score.

A subject with a PHQ-8 score ≥15 should be immediately referred to a mental health professional, and Investigators in consultation with the mental health professional should consider interruption or discontinuation of the study drug.

5.1.8. Patient Health Questionnaire Depression Scale (Modified PHQ-A)

The Modified PHQ-A Assessment (Appendix 3) will be performed in adolescent subjects (12-17 years old, inclusive).

Modified PHQ-A score is the sum of the responses for five severity categories of depression defined 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)

A subject with a modified PHQ-A score of 10-14 should be referred to a mental health professional for evaluation, regardless of the reason for their change in score.

A subject with a modified PHQ-A score ≥15 should be immediately referred to a mental health professional, and Investigators in consultation with the mental health professional should consider interruption or discontinuation of the study drug.

5.1.9. Children's Depression Inventory 2 (CDI-2)

The CDI-2 Assessment will be performed according to the Schedule of Visits and Assessments (Section 1.3) for subjects 6 to 11 years old, inclusive.

The CDI-2 quantifies depressive symptomatology using reports from children/adolescents, teachers, and parents or caregivers. It is recommended for use in initial evaluation and is appropriate when there is a need for an assessment and robust description of a child's depressive symptoms.

This study will use the CDI Parent Report Form. An example of the Parent report form is presented in Appendix 4.

A subject with a CDI-2 raw score of ≥21 for females and ≥22 for males should be referred to a mental health professional for evaluation.

A subject with a CDI-2 raw total score of ≥32 should be referred to a mental health professional, and Investigators in consultation with the mental health professional should consider interruption or discontinuation of the study drug.

5.1.10. Columbia-Suicide Severity Rating Scale (C-SSRS)

C-SSRS Assessments will be performed according to the Schedule of Visits and Assessments (Section 1.3) for subjects 12-years old and older.

The administration schedule of the C-SSRS will be:

- The "Baseline/Screening" version (Appendix 5) will be used at Screening to provide a pre-treatment assessment.
- On all subsequent visits, the "Since Last Visit" version (Appendix 6) will be used (Baseline/Day 1, Week 1/Day 8, Week 2/Day 15 or Week 4/Day 29).
- A score greater than 0 at the Screening or Baseline visit in suicidal ideation may indicate the need for mental health intervention. The investigator should not enroll the subject in the study.
- Any score greater than 0 in the suicidal ideation score may indicate the need for mental health intervention. The Investigator should give consideration for the subject to discontinue from the study drug and prompt referral to an identified mental health professional and/or an appropriate emergency room. The Medical Monitor should be contacted.

The C-SSRS administer will be trained via the C-SSRS training video. A training certificate for the administer(s) will be on file in the trial master file at the site.

The Investigator must review the completed C-SSRS. A qualified mental health care provider must be available, immediately if needed, to refer the subject for further evaluation.

5.1.11. Local Tolerability Assessment

The Investigator Local Tolerability Assessment will be performed according to the Schedule of Visits and Assessments (Section 1.3).

Application site reactions will be graded at each timepoint. Irritation reactions are graded using the scale detailed in the following section (Berger 1982).

Reactions at the site of product application, which may occur post-Baseline, should be differentiated from the preexisting inflammation associated with the subject's atopic dermatitis.

The investigator assessments will be conducted by the investigator or a properly trained and designated subinvestigator prior to study drug application in the clinic.

Dermal Response

0. no evidence of irritation
1. minimal erythema, barely perceptible
2. definite erythema, readily visible; minimal edema or minimal papular response
3. erythema and papules
4. definite edema
5. erythema, edema and papules
6. vesicular eruption
7. strong reaction spreading beyond application site

Other Effects

A. = slight glazed appearance
B. = marked glazing
C. = glazing with peeling and cracking
D. = glazing with fissures
E. = film of dried serous exudates
F. = small petechial erosions and/or scabs
G. = no other effects

The Subject Local Tolerability Assessment will be performed according to the Schedule of Visits and Assessments (Section 1.3).

The subject will assess burning/stinging (0-3 score):

Grade	Sensation Following Drug Application
0 (none)	No sensation
1 (mild)	Slight warm, tingling sensation; not really bothersome
2 (moderate)	Definite warm, tingling sensation that is somewhat bothersome
3 (severe)	Hot, tingling/stinging sensation that has caused definite discomfort

This assessment will be administered by the site 10 to 15 minutes after study drug application in the clinic at Baseline and at every clinic visit.

- Note: for subject burning stinging assessment at Day 29, subjects will provide a recall assessment of burning/stinging experienced post drug application on the previous day (Day 28).

5.1.12. Adverse Events

Adverse events (AEs) will be collected and assessed throughout the study according to the Schedule of Visits and Assessments (Section 1.3). The Investigator is responsible for ensuring that all adverse events observed by the clinical staff or reported by the subject that occur after the first application of investigational product through the end of the study are recorded in the subject's medical record and the eCRF.

The Investigator is responsible for ensuring that all serious adverse events observed by the clinical staff or reported by the subject that occur after signing of the informed consent through 30 days after the last day of the application of the investigational product or the end of study (whichever is later) are recorded in the subject's medical record and are submitted per SAE reporting requirements (Section 5.7.5).

AEs (whether serious or non-serious) and clinically significant abnormal laboratory test value(s) will be evaluated by the PI and treated and/or followed up for up to 30 days after end of treatment or until the symptoms or value(s) return to normal, or acceptable levels, as judged by the PI. Refer Section 5.7 for further details on Adverse Events.

5.2. Efficacy Evaluations

For efficacy evaluation subjects will have ≥3% BSA of AD involvement (excluding the scalp, palms, soles). Palms and soles may be treated with investigational product in this study, but will not be counted towards vIGA-AD, EASI, or BSA assessments. EASI is evaluated for the entire body except the scalp, palms, and soles.

5.2.1. Validated Investigator Global Assessment Scale for Atopic Dermatitis

Validated Investigator Global Assessment scale for Atopic Dermatitis (vIGA-AD) assessments should be completed prior to other physician assessments.

vIGA-AD assessment will be performed according to the Schedule of Visits and Assessments (Section 1.3). The vIGA-AD 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). Each grade is defined by a distinct and clinically relevant morphologic description that minimizes inter-observer variability (see Appendix 7). vIGA-AD is evaluated for the entire body except the scalp, palms, and soles.

Note: All atopic dermatitis lesions on a subject will be treated including the face, trunk, genitals/skin folds, or limbs (excluding the scalp). The palms and soles will be treated but will not be counted towards any measurements of efficacy (EASI, vIGA-AD, BSA).

Every effort must be made for the same Evaluator to complete the IGA for the subject at every study visit.

IGA will be assessed at clinic visits prior to the subject applying Investigational Product at the site.

5.2.2. Eczema Area and Severity Index (EASI)

EASI scores (Hanifin 2001) will be performed according to the Schedule of Visits and Assessments (Section 1.3)

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; eg, 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:

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

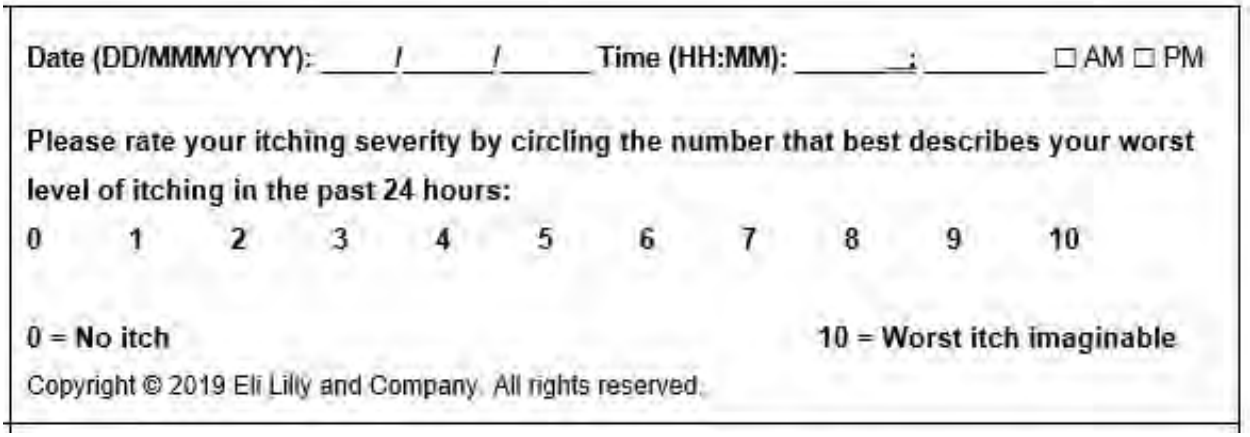
Where E, I, E*x*, L, and A denote erythema, induration, excoriation, lichenification and area, respectively, and h, u, t, and l denote head, upper extremities, trunk, and lower extremities, respectively.

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

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.

5.2.3. Worst Itch Numerical Rating Scale (WI-NRS)

The WI-NRS has been developed as a simple, single item to assess the patient-reported severity of this symptom at its highest intensity during the previous 24-hour period. (Newton 2019). The WI-NRS will be determined by the subject's recording of daily assessment of worst itch over the past 24 hours. The scale is from '0 to 10' ("no itch" to "worst itch imaginable" or "worst imaginable itch").


Date (DD/MMM/YYYY): ____/____/____ Time (HH:MM): ____:____ ☐ AM ☐ PM

Please rate your itching severity by circling the number that best describes your worst level of itching in the past 24 hours:

0 1 2 3 4 5 6 7 8 9 10

0 = No itch 10 = Worst itch imaginable

The WI-NRS will be **self-reported by the subjects.** Subjects will be reminded not to review responses from the previous assessment when completing the WI-NRS. Subjects will be trained at the Screening visit in the accurate completion of the WI-NRS. In addition, parents/caregivers of children and adolescent subjects will be trained at the Screening visit by study staff on how to assist the subject, if needed, in completing the WI-NRS. If subjects, despite the assistance of a parent/caregiver, are unable to complete the WI-NRS no entry should be made in the subject diary.

WI-NRS Assessments will be performed according to the Schedule of Visits and Assessments (Section 1.3) starting 7 days prior to the Baseline/Day 1 clinic visit (during the 7 days prior to Baseline/Day 1 the subject will record the WI-NRS value every day) and until Week 4/ET.

5.2.4. Dermatology Life Quality Index (DLQI) and Children's Dermatology Life Quality Index (CDLQI)

The DLQI (ages 17+ years) and CDLQI (ages 6-16 years, inclusive) will be performed according to the Schedule of Visits and Assessments (Section 1.3). The DLQI/CDLQI is a simple, self-administered and user-friendly validated questionnaire. The DLQI/CDLQI is designed to measure the health-related quality of life of adult 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. Subjects/caregivers will complete the CDLQI/DLQI. Refer to Appendix 8 for the DLQI and Appendix 9 for the CDLQI.

5.2.5. Dermatitis Family Impact Questionnaire (DFI)

This questionnaire measures how much having a child with atopic dermatitis affects the quality of life of other (adult) members of the family. To be completed by parents/guardians/caregivers of subjects ≤17 years of age (Appendix 10).

5.3. Other Evaluations

5.3.1. Body Surface Area (BSA)

BSA assessments will be performed according to the Schedule of Visits and Assessments (Section 1.3).

The BSA affected for atopic dermatitis will be determined by the subject's hand method, where the subject's hand (including fingers) surface area is assumed to equal 1% of body surface area (excluding the scalp, palms, soles).

5.3.2. SCORAD ("SCORing Atopic Dermatitis")

SCORAD is a clinical tool for assessing the severity (i.e. extent, intensity) of atopic dermatitis as objectively as possible. It gives approximate weights of 60% to intensity and 20% each to spread (extent) and subjective signs (insomnia, etc.). See Appendix 11.

5.3.3. Patient-Oriented Eczema Measure (POEM)

The Patient-Oriented Eczema Measure (POEM) is a tool used for monitoring atopic eczema severity. It focuses on the illness as experienced by the patient.

POEM is a 5-point scale measuring the frequency of each of seven AD symptoms (dryness, itching, flaking, cracking, sleep loss, bleeding, and weeping) over the past week scored as occurring "no days" (0), "1 to 2 days" (1), "3 to 4 days" (2), "5 to 6 days" (3) or "every day" (4). Total score ranges from 0–28, with higher score indicating greater symptom impact. See Appendix 12. The self/proxy report questionnaire will be used in this study (for children unable to read and/or understand the POEM questionnaire, the parent/guardian/caregiver will complete the questionnaire).

5.3.4. Pharmacokinetics Assessment

PK draws will be performed according to the Schedule of Visits and Assessments (Section 1.3) for all subjects at all sites under Amendment 2:

- A single PK assessment (trough) will be performed in subjects ≥12 years old with a blood sample collected at Week 4/Day 29/ET.
- No PK sample will be collected Week 4/Day 29 for subjects 6-11 years old.

Ensure study medication is not applied in the area where PK will be drawn.

5.3.5. Medical Photography

Photography of AD lesion(s) selected by the Investigator will be performed by all sites at all investigational visits, except Week 2/Day 15. All efforts will be made to de-identify the subjects. Canfield equipment will be used to capture photographs.

Photography should be focused on single lesions or specific body sections (eg, arm). Body or half body photos should only be taken if necessary. 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.

Refer to the current Photography Manual for instructions regarding photography.

5.4. Final Study Visit – End of Study

The approximate final study visit will occur at Week 4/Day 29. The procedures performed during this visit are as described in the Schedule of Visits and Assessments (Section 1.3). A 3-day scheduling window is allowed for this visit. Adverse events will be recorded as reported by the participant or and followed to resolution or stabilization (as necessary).

5.5. Early Termination Visit

If a subject is withdrawn or wishes to exit the study, a termination visit will be scheduled. This visit should include the procedures and assessments that would be performed at the Week 4/ Day 29 visit.

5.6. Unscheduled Visit

Unscheduled visits may be necessary to repeat testing following abnormal laboratory results, for follow-up of AEs, or for any other reason, as warranted in the judgement of the Investigator.

The following information will be collected for all subjects:

- Concomitant medications/procedures
- AEs

The following information also will be collected:

- vIGA-AD and EASI
- BSA affected with AD
- Local tolerability assessment (by Investigator)

However, if an unscheduled visit is required for reasons other than safety (eg, procedures such as labs or images that were either missed at the regular subject visit or need to be repeated), the vIGA-AD and EASI, BSA affected with AD, and Local Tolerability assessment (by Investigator) are not required.

The rules for how to tally vIGA-AD, BSA or other proportions of categorical responses will be described in the Statistical Analysis Plan.

5.7. Adverse Events

5.7.1. Adverse Event Definition

An AE is defined as any untoward medical occurrence associated with the use of a drug in humans, whether or not considered drug related. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not related to the medicinal (investigational) product.

A treatment emergent adverse event (TEAE) is defined as an AE that started post application of study medication at the Baseline visit or was present at treatment initiation but worsened during treatment, through study completion.

The definition of AEs includes worsening of a pre-existing medical condition. Worsening indicates that the pre-existing medical condition has increased in severity, frequency, and/or duration or has an association with a worse outcome. When recording such events, provide descriptions that the pre-existing condition has changed (eg, worsening hypertension for a subject with pre-existing hypertension). A pre-existing condition that has not worsened during the study or involves an intervention, such as elective cosmetic surgery or a medical procedure while on study, is not considered an AE.

Progression of atopic dermatitis including new or worsening of anticipated clinical signs or symptoms, which are collected as clinical efficacy variables and assessed as unequivocally associated with the disease progression and /or lack of efficacy, should NOT be reported as

adverse events unless the disease progression is greater than anticipated in the natural course of the disease.

In general, abnormal laboratory findings without clinical significance (based on the investigator's judgment) are not to be recorded as AEs. However, abnormal laboratory findings that result in new or worsening clinical sequelae, or that require therapy or adjustment in current therapy, are considered AEs. Where applicable, clinical sequelae (not the laboratory abnormality) are to be recorded as the AE.

5.7.2. Serious Adverse Event Definition

The definitions and reporting requirements of the Food and Drug Administration (FDA)/ICH Guidelines for Clinical Safety Data Management, Definitions and Standards for Expedited Reporting, Topic E2A will be adhered to. Statements regarding mandatory reporting of all serious unexpected adverse drug reactions (SUSARs) to Health Canada [as per C.05.014 (1) of the FDR] will be adhered. If any AEs are serious, as defined by ICH Guidelines for Clinical Safety, required procedures will be followed.

An SAE is defined as any AE that, in the view of either the PI or Sponsor, meets at least 1 of the following serious criteria:

- Fatal
- Life-threatening (places the subject at immediate risk of death)
- Requires inpatient hospitalization or prolongation of existing hospitalization
- Results in a persistent or significant incapacity/disability
- Congenital anomaly/birth defect.
- Other important medical events that, based upon appropriate medical judgment, may jeopardize the subject and may require medical or surgical intervention to prevent one of the outcomes listed in the above definition.

Hospitalization does not include the following:

- Rehabilitation facilities, hospice facilities or respite care (eg, caregiver relief)
- Nursing homes or skilled nursing facilities
- Emergency room visits
- Same day surgeries (as outpatient/same day/ambulatory procedures)
- <24 hour admissions for observation or evaluation

Hospitalization or prolongation of hospitalization in the absence of a precipitating, clinical AE is not in itself an SAE. Examples include:

- Admission for treatment of a preexisting condition that did not worsen
- Hospitalizations for cosmetic elective surgery, social, and/or convenience admissions
- Pre-planned treatments or surgical procedures should be noted in the baseline documentation for the individual subject.

- Diagnostic and therapeutic procedures, such as surgery, should not be reported as AEs; however, the medical condition for which the procedure was performed should be reported if it occurs during the reporting period and meets the definition of an AE. For example, an acute appendicitis that begins during the AE reporting period should be reported as an AE, and the resulting appendectomy should be recorded as treatment of the AE.

Unexpected is defined as an AE that is not listed in the IB or is not listed at the specificity or severity that has been observed; or is not consistent with the risk information described in the general investigational plan or study document.

If a SAE occurs to a subject on this study, contact the Medical Monitor within one business day of knowledge of event.

5.7.3. Suspected Unexpected Serious Adverse Reaction (SUSAR)

SUSAR is defined as a serious adverse reaction, the nature or severity of which is not consistent with the known study treatment information. A serious event or reaction is not defined as a SUSAR when: 'it is serious but expected' or it does not fit the definition of an SAE, whether expected or not.

5.7.4. Safety Review with Subject

At each subsequent clinic visit after the screening visit, subjects will be queried with an open-ended question such as: 'How have you been feeling since your last visit?' Additionally, the study staff will review subject diaries and, if it appears that a potential AE was recorded, study staff will query the subject and determine if an AE occurred.

AEs (whether serious or non-serious) and clinically significant abnormal laboratory test value(s) will be evaluated by the PI and treated and/or followed for up to one month after end of treatment until the symptoms or value(s) return to normal, or acceptable levels, as judged by the Investigator.

Where appropriate, medical test(s) and/or examination(s) will be performed to document resolution of event(s). Outcome may be classified as resolved, improved, unchanged, worse, fatal or unknown (lost to follow-up).

5.7.5. Adverse Event Reporting

The Investigator is responsible for recording all adverse events, observed by the clinic staff or reported by the subject that occur after the first application of investigational product through the end of the study. All SAEs should be reported starting after the signing of the informed consent through 30 days after the last day of the application of the investigational product or the end of the study (whichever is later).

Any AEs (whether serious or non-serious) and clinically abnormal laboratory test value(s) will be evaluated by the PI and treated and/or followed up for up to 30 days after end of treatment or until symptoms or value(s) return to normal, or acceptable level, as judged by the PI (if the subject is continuing into the ARQ-151-313 OLE study, then AEs from this study (ARQ-151-312) will only be followed until they exit from this study).

All adverse events that meet the criteria for "serious" (i.e., SAEs) will be reported to the Sponsor via fax or e-mail within 24 hours of becoming aware of the event, whether or not the serious events are deemed drug-related. Reporting should be done by sending the completed SAE form to the following e-mail address (faxing can also be done as a second option in case e-mailing is not possible).

Safety Contact Information: [redacted]
E-mail: [redacted]

All serious event reporting will adhere to ICH E6: Guideline for Good Clinical Practice and ICH E2A: Clinical Safety Data Management: Definitions and Standards for Expedited Reporting. The IRB will be notified of the Alert Reports as per FDA, ICH and the IRB's policies and procedures. The sponsor, or delegate, will be responsible for reporting SAEs to health authorities per local reporting requirements. The Investigator will be responsible for reporting events to their respective IRBs in accordance to the IRB requirements.

The Investigator will review each adverse event and assess its relationship to Investigational Product (unrelated, unlikely, possibly, probably, likely). Each sign or symptom reported will be graded on the NIH NCI CTCAE toxicity grading scale 5-point severity scale (Grade 1, 2, 3, 4 and 5). The date and time of onset, time relationship to drug dosing, duration, and outcome (resolved, improved, unchanged, worse, fatal, or unknown/lost to follow-up) of each event will be noted.

- The relationship of each AE to the Investigational Product will be assessed using the following definitions:

Unrelated	The AE must clearly be caused by the subject's clinical state, or the study procedure/conditions. Definitely not related to drug. Temporal sequence of an AE onset relative to administration of drug not reasonable. Another obvious cause of an AE.
Unlikely	Time sequence is unreasonable. There is another more likely cause for an AE.
Possibly	Corresponds to what is known about the drug. Time sequence is reasonable. Could have been due to another equally, likely cause.
Probably	Is a known effect of the drug. Time sequence from taking drug is reasonable. Ceases on stopping the drug. Cannot be reasonably explained by the known characteristics of the subject's clinical state.
Likely	Is a known effect of the drug (eg, listed in Physicians' Desk Reference, Compendium of Pharmaceuticals and Specialties, IB). Time sequence from taking drug is reasonable. Event stops upon stopping drug, event returns upon restarting drug.

The following CTCAE toxicity grading scale 5-point severity scale definitions for rating maximum severity will be used:

Grade 1	Mild; asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated.
Grade 2	Moderate; minimal, local or noninvasive intervention indicated; limiting age-appropriate instrumental activities of daily living.*
Grade 3	Severe or medically significant but not immediately life-threatening; Hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care activities of daily living.** Note**:** An experience may be severe but may not be serious, eg, severe headache).
Grade 4	Life-threatening consequences; urgent intervention indicated.
Grade 5	Death related to AE.

Note: A semi-colon indicates 'or' within the description of the grade.

* Instrumental activities of daily living refer to preparing meals, shopping for groceries or clothes, using the telephone, managing money, etc.

** Self-care activities of daily living refer to bathing, dressing and undressing, feeding self, using the toilet, taking medications, and not bedridden.

AEs will be coded using the most current MedDRA® version available at the start of the study.

5.8. Reporting Pregnancy

During study participation, all subjects should be instructed to contact the investigator immediately if they suspect they might be pregnant (eg, missed or late menstrual period). If pregnancy is confirmed, Investigational Product must be discontinued immediately, the subject should be referred to an obstetrician experienced in reproductive toxicity for evaluation and counseling, and the subject should be followed until conclusion of the pregnancy.

The investigator is responsible for reporting all available pregnancy information on the pregnancy report and submitting to the Medical Monitor within 24 hours of becoming aware of the event, although pregnancy itself is not considered an AE. Follow-up information detailing the outcome of the pregnancy and the health of the newborn should be reported as it becomes available. Any pregnancy complication must be reported as a SAE. In addition, any pregnancy resulting in a congenital abnormality or birth defect of the newborn, or neonatal death occurring within 30 days of the birth must be reported as a SAE, regardless of causality. Any infant death that occurs after the 30-day reporting period that the investigator suspects is related to Investigational Product must also be reported as an SAE.

Partner pregnancies of a male subject do not need to be reported.

5.9. Treatment Stopping Rules

If a subject has non-cutaneous adverse events of concern, clinically significant laboratory values or any condition that the investigator determines could possibly be related to the study drug, the Investigator should immediately contact the Medical Monitor to discuss if the subject should be discontinued from treatment with investigational product.

Treatment for any individual subject will be discontinued if the subject experiences:

- A serious adverse event (SAE) or a clinically significant non-serious AE which in the opinion of the Principal Investigator or Medical Monitor warrants discontinuation from the study for that subject's well-being.
- A treatment-emergent severe (Grade 3) laboratory abnormality (confirmed by repeat sample; see Appendix 13 for more information).

A subject with a PHQ-8 or modified PHQ-A score ≥15 should receive immediate referral to a mental health professional, and Investigators in consultation with the mental health professional should consider interruption or discontinuation of the study drug.

- Subjects with a PHQ-8 or modified PHQ-A score of 10-14 should be referred to a mental health professional for evaluation, regardless of the reason for their change in score

Subjects with a CDI-2 raw total score of ≥32 should be referred to a mental health professional, and Investigators in consultation with the mental health professional should consider interruption or discontinuation of the study drug

- Subjects with a CDI-2 raw score of ≥21 for females and ≥22 for males should be referred to a mental health professional for evaluation

A subject that is experiencing suicidal ideation and behavior should be referred immediately to a qualified mental health care provider and consideration given to discontinuation from study drug.

As noted above, study treatment must be discontinued immediately in the event of a female subject's pregnancy.

Treatment should be interrupted:

- If a subject develops an application site reaction with the clinical appearance of an 'irritation reaction', and with a severity of a Dermal Response Score of 5 (erythema, edema and papules) or greater on the scale of Berger and Bowman, treatment should be interrupted for up to one week and may then be resumed if the reaction has, in the opinion of the Investigator, adequately resolved.

Treatment should be discontinued:

- If the application site reaction reoccurs, treatment should be discontinued permanently, and the subject followed until the reaction resolves.

For cases of suspected allergic contact dermatitis, the medical monitor and sponsor should be notified and there should be discussion about performing patch testing to further evaluate. Patch testing is encouraged in such cases.

In the event of a medical emergency where unblinding is required to provide medical care to the subject, refer to the most current IWRS User Manual (Section 4.8.3). Contact the Medical Monitor and the Sponsor promptly.

6. DATA ANALYSIS

Data will be handled and processed according to the Contract Research Organization's Standard Operating Procedures, which are written based on the principles of GCP.

6.1. Statistical Methods

The methodology presented below is a summary of the more detailed analysis plan that will be presented in the Statistical Analysis Plan (SAP). The SAP will be finalized before the database is locked and unblinded. Any changes to the methods described in the final SAP will be described and justified in the clinical study report.

All statistical processing will be performed using SAS® (Version 9.4) unless otherwise stated.

6.2. Determination of Sample Size

There are approximately 650 subjects planned for this study. In order to test the secondary endpoint of IGA success in subjects with a vIGA-AD score of 'Moderate' at randomization, approximately 490 of the subjects to be accrued will have vIGA-AD score of 'Moderate' at randomization. Randomization will be stratified by vIGA-AD score ('Mild' vs. 'Moderate') and by study site.

This sample size provides approximately 95% power to detect an overall 15% difference between treatment groups on vIGA-AD success at Week 4 at α=0.05 using a 2-sided stratified Cochran-Mantel-Haenszel test. The results from a recent phase 2 study (ARQ-151-212) of ARQ-151 cream 0.15% compared to vehicle treatment were used to estimate the treatment difference. Specifically, in the phase 2 trial, approximately 37% of subjects demonstrated vIGA-AD success at Week 4 in the ARQ-151 0.15% group compared to 22% in the vehicle group. This sample size also provides approximately 90% power to detect an overall 17% difference between treatment groups on IGA success at Week 4 among subjects with vIGA-AD score 'moderate' at randomization. The same testing method, the stratified Cochran-Mantel-Haenszel test, will be used as for the primary endpoint.

To control for familywise type I error at level of 0.05, the secondary endpoint of vIGA-AD success at Week 4 in subjects with vIGA-AD of 'Moderate' at randomization will only be tested if the primary endpoint demonstrates statistical significance. In addition, the remaining secondary endpoints will be inferentially tested only if the primary and above secondary endpoint (vIGA-AD success at week 4 for subjects with vIGA-AD score of 'Moderate' at randomization) comparisons are statistically significant using the hierarchical testing procedure by partitioning of the alpha and the Fallback Method.

6.3. Subjects to Analyze

The analysis populations are defined as follows:

- Intent-to-Treat (ITT) population will include all subjects who are randomized.

- Per protocol (PP) population will include all subjects in the ITT population, who are at least 80% compliant with study medication application, have a vIGA-AD assessment within the Week 4 visit window, and show no major deviations from the study protocol that would affect the interpretation of efficacy. In addition, subjects who miss the Week 4 vIGA-AD assessment specifically due to novel coronavirus disease-19 (COVID-19) disruptions will be excluded from per protocol population.
- vIGA-AD Moderate ITT population will be a subset of the ITT population with vIGA-AD score 'moderate' at randomization.
- vIGA-AD Moderate PP population will be a subset of the PP population with vIGA-AD score 'Moderate' at randomization
- WI-NRS population will be a subset of the ITT population who are ≥12 years old at Baseline and have a Baseline WI-NRS score ≥ 4.
- Safety population will include all subjects who are enrolled and received at least one confirmed dose of study medication.
- Pharmacokinetic (PK) population will include all subjects receiving the active drug with quantifiable plasma concentrations of roflumilast.

6.4. Background and Demographic Characteristics

Descriptive statistics will be used to summarize demographic characteristics (age, sex, ethnicity, and race) and background characteristics for the randomized subjects.

6.5. Study Medication Compliance

The number of study drug applications by each subject based on diary data will be summarized using summary statistics (mean, standard deviation [SD], median, minimum, and maximum), and categorically.

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

The amount of investigational product used by each subject based on tube weight will be summarized by treatment using descriptive statistics, and categorically.

Investigational product application compliance will be calculated based on number of applications divided by the expected number (amount) of investigational product applications for each subject. Compliance will be summarized descriptively by treatment group.

6.6. Safety Analysis

Descriptive statistics will be calculated for quantitative safety data and frequency counts will be compiled for classification of qualitative safety data. The safety population will be used for these analyses.

6.7. Efficacy Analysis

The Primary Efficacy Endpoint will be tested in all randomized subjects and defined as:

- IGA Success, defined as a vIGA-AD score of 'clear' or 'almost clear' PLUS a 2-grade improvement from Baseline at Week 4

The primary estimand is the ratio of the odds of achieving vIGA-AD success after 4 weeks of using ARQ-151 (roflumilast cream 0.15%), relative to the odds of success after 4 weeks of using a matching vehicle cream. In the course of the 4-week randomized treatment period, subjects may be exposed to possible known or unknown inter-current events that could possibly impact the estimates of the estimand, such as treatment discontinuation due to a specific adverse effect or perhaps a lack of effect. A composite strategy will be implemented that handles subjects who discontinue due to lack of efficacy or adverse event as missing not at random, differently than all other subjects. The "Treatment Policy Strategy" has been adopted for handling all other known or unknown intercurrent events in this study other than discontinuation due to lack of efficacy or adverse event. Subjects who discontinue due to lack of efficacy or adverse event will be treated as non-responders at all analysis visits that occurred or would have occurred on or after the date of last dose of treatment application. Odds ratio of achieving vIGA-AD success for ARQ-151 (roflumilast cream 0.15%) relative to vehicle after 4 weeks will be evaluated accordingly. This estimand shall be estimated using the CMH approach. The primary efficacy endpoint is success in IGA of disease severity, defined as an vIGA-AD score of 'clear' or 'almost clear' PLUS a 2-grade improvement from Baseline at Week 4.

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

To control for the familywise type I error at level of 0.05, a hierarchical testing scheme will be used to test the following secondary endpoint:

- IGA success at Week 4 in subjects with a vIGA-AD score of 'Moderate' at randomization

Upon successful demonstration of statistical significance for the primary and above secondary endpoint, the remaining endpoints will be grouped into secondary endpoint family 1, comprised of the 4-point reduction on the WI-NRS endpoint, at Week 4, Week 2 and Week 1, and secondary endpoint family 2, comprised of the endpoints of EASI-75 at Week 4, vIGA-AD of 'clear' or 'almost clear' at Week 4, vIGA-AD of success at Week 2 and Week 1, vIGA-AD of 'clear' or 'almost clear' at Week 2 and Week 1. An alpha level of 0.03 will be used to test the endpoints in the secondary endpoint family 1 sequentially. An alpha level of 0.02 will be used to test the endpoints in secondary endpoint family 2 sequentially. See Figure 2 for details.

The endpoints listed below will be tested using the ITT population, with the exception of the WI-NRS endpoints, which will use the WI-NRS population.

In addition to the partitioning of the overall 0.05 alpha into two families, the Fallback Method will be applied. The fallback method is a modification of the fixed-sequence method, providing opportunity to test an endpoint later in the sequence even if an endpoint tested early in the

sequence has failed to show statistical significance. The order of the endpoints remains important. The appeal of the fallback method is that if an endpoint later in the sequence has a robust treatment effect while the preceding endpoint is unsuccessful, there is a modest amount of alpha retained as a fallback to allow interpretation of that endpoint without inflating the Type I error rate. Applying the fallback method begins by dividing the total alpha (not necessarily equally) among the endpoints and maintains a fixed sequence for the testing. In this study, the Fallback Method will be applied to the fixed sequence of testing Family 1, and then Family 2.

As the testing sequence progresses, a successful test preserves its assigned alpha as "saved" ("unused" or "accumulated") alpha that is passed along to the next test in the sequence, as is the case for the sequential method. This accumulated alpha is added to the prospectively assigned alpha (if any) of that next endpoint and the summed alpha is used for testing that endpoint. Thus, as sequential tests are successful, the alpha accumulates for the endpoints later in the sequence; these endpoints are then tested with progressively larger alphas.

In this study, the Fallback Method will be applied following this sequence:

Family 1: Testing will proceed at the 0.03 level sequentially within Family 1. Should all 3 endpoints in Family 1 be statistically significant at the 0.03 level, then the full 0.03 alpha will be carried to Family 2. Family 2 would then be tested at the full (α=0.02+0.03=0.05).

Should, anywhere during the sequential testing of Family 1, there be a p-value >0.03, the testing within Family 1 will stop, and no additional alpha can be carried over to Family 2.

Secondary Endpoint Family 1 (α=0.03, hierarchical testing within Family 1)

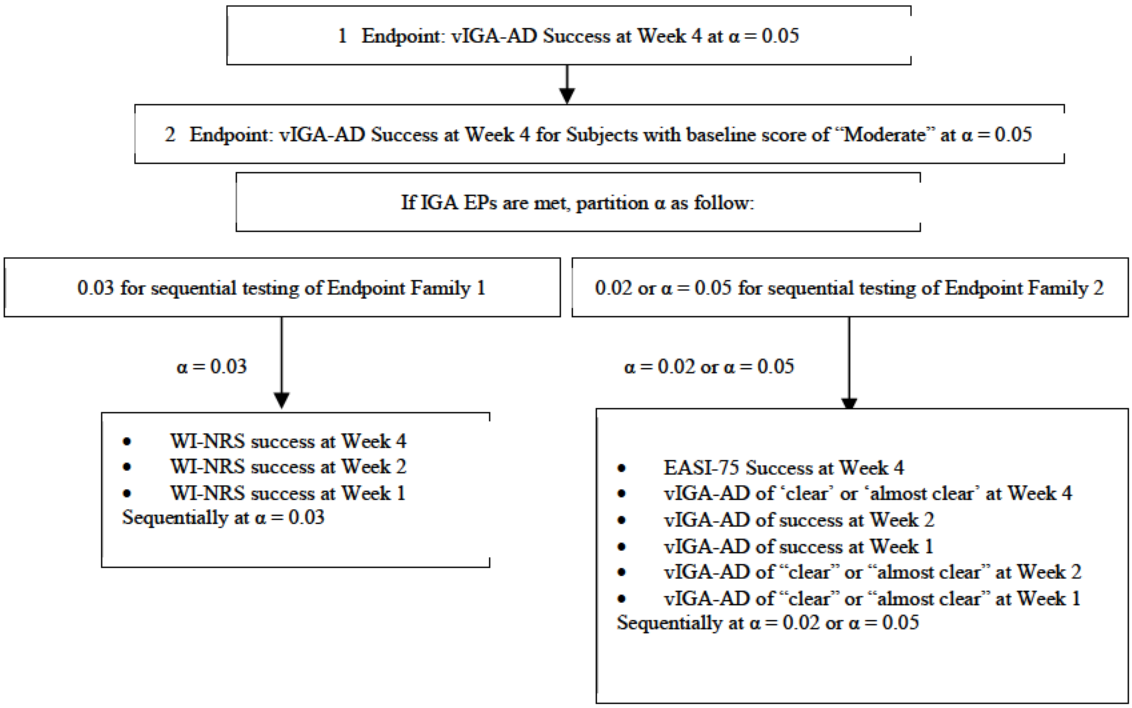
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the WI-NRS at Week 4
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the WI-NRS at Week 2
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the WI-NRS at Week 1

Secondary Endpoint Family 2 (α = 0.02 or 0.05 pending Family 1 results, hierarchical testing within Family 2)

- Achievement of at least a 75% reduction in the Eczema Area and Severity Index at Week 4 (EASI-75)
- vIGA-AD score of 'clear' or 'almost clear' at Week 4
- vIGA-AD of 'clear' or 'almost clear' at Week 4
- vIGA-AD Success at Week 2
- vIGA-AD Success at Week 1
- vIGA-AD of 'clear' or 'almost clear' at Week 2
- vIGA-AD of 'clear' or 'almost clear' at Week 1

A supplemental analysis of the primary endpoint, v-IGA Success at Week 4, and the first secondary endpoint, vIGA-AD Success at Week 4, for subjects with baseline score of "Moderate", will be performed using the PP population and the vIGA-AD Moderate PP population, respectively.

Figure 2: Multiple Testing Scheme



Achievement of vIGA-AD success is a score of "clear" or "almost clear" plus a 2-grade improvement from baseline.
WI-WRS Success is a 4-point reduction in WI-NRS among subjects ≥12 years old with WI-NRS ≥4 at baseline.
EASI-75: achievement of at least a 75% reduction in the Eczema Area and Severity Index

6.8. Adverse Events

All TEAEs occurring during the study will be recorded and classified on the basis of MedDRA terminology for the safety population. TEAEs are defined as those AEs with an onset on or after the time of first study drug application. All reported TEAEs will be summarized by treatment group.

Descriptive statistics will be calculated for quantitative safety data and frequency counts will be compiled for classification of qualitative safety data.

Adverse Events:

The subject incidence of treatment-emergent adverse events (TEAE) will be summarized overall, by severity, and by attribution.

Clinical Laboratory Results:

Shifts in clinical laboratory parameters from baseline to worst post-baseline grade will be provided.

Vital Signs:

The subject incidence of >5% weight loss or gain on study will be provided, as well as whether weight loss was explained or unexplained.

6.9. Body Surface Area

Body surface area (BSA) affected by AD will be analyzed descriptively.

6.10. Local Tolerance Assessment

For Investigator's assessment, the numeric application site reaction scores will be summarized individually by using number and percentage of subjects by visit, as well as mean/median scores.

6.11. Medical History and Physical Examinations

Medical history for all subjects will be presented descriptively by parameter.

Physical examination for all subjects will be presented descriptively by parameter. Changes in physical examinations will be described in the text of the final report.

6.12. Clinical Laboratory Results and Vital Signs

All clinical laboratory results and vital signs measurements will be summarized descriptively by parameter, visit, and treatment group along with time point of collection.

A shift from baseline table describing out-of-normal range shifts will be provided for clinical laboratory results.

A shift from baseline table will identify subjects who gain or lose >5% body weight over the course of the study.

6.13. Prior and Concomitant Medications

Prior and concomitant medication information for all randomized subjects will be presented in a by-subject listing. Summary tables will be presented by World Health Organization-Anatomical Therapeutic Chemical Classification System (WHO-ATC) therapeutic category and product.

6.14. Subject Reported Outcomes Analyses

Weekly average WI-NRS scale will be summarized by treatment group and by visits.

6.14.1. Dermatology Life Quality Indexes, Children's Dermatology Life Quality Index, SCORAD and POEM

Dermatology Life Quality Index (DLQI), Children's Dermatology Life Quality Index (CDLQI), the Dermatitis Family Impact (DFI), the Scoring Atopic Dermatitis (SCORAD), and the

Patient-oriented Eczema Measure (POEM) will be analyzed by evaluation of the reduction in total score at Week 4. These efficacy endpoints will be analyzed descriptively.

6.15. Pharmacokinetic Analysis

Plasma drug concentrations will be summarized using descriptive statistics, reporting n, mean, standard deviation, median, minimum, and maximum. The PK population will be used for these analyses.

7. STUDY ADMINISTRATION

7.1. Ethics

7.1.1. Ethics Review Board

Before screening of subjects into the study, the current protocol, ICF, and any accompanying material to be provided to the subjects will be reviewed and approved by an appropriate IRB, as required by FDA (21 CFR § 56) and ICH GCP regulations. A letter documenting the IRB approval must be received by the Sponsor before the initiation of the study at a clinical site. Amendments to the protocol will be subject to the same requirements as the original protocol.

The Investigator, Sponsor, or designee will submit a progress report at least once yearly to the IRB. However, the frequency of these reports will depend on IRB requirements. As soon as possible after completion or termination of the study, the Investigator will submit a final report to the IRB per the IRB requirements, and in compliance with FDA and ICH GCP guidelines.

The Investigator, the Sponsor, or designee shall promptly notify the IRB of any SAEs, SUSARs, or any other information that may affect the safe use of the study drug(s) during the study, per the IRB local requirements, and in compliance with FDA regulations and ICH GCP guidelines.

7.1.2. Ethical Conduct of the Study

This research will be carried out in accordance with the protocol, the principles of the Tri-Council Policy Statement (TCPS), the ethical principles set forth in the Declaration of Helsinki, and the ICH harmonized tripartite guideline regarding GCP (E6 Consolidated Guidance, April 1996).

7.1.3. Subject Information and Consent/Assent

The investigator is responsible for obtaining written informed consent from each individual participating in this study after adequate explanation (in non-technical terms) of the aims, methods, objectives, and potential hazards of the study and before undertaking any study-related procedures. The investigator must use the most current IRB- or IEC-approved consent form for documenting written informed consent. Subjects will be assured that they may withdraw from the study at any time without jeopardizing their medical care. Each informed consent (or assent as applicable) will be read, appropriately signed and dated by the subject or the subject's legally authorized representative and the person conducting the consent discussion, and also by an impartial witness if required by IRB or IEC or local requirements.

Subjects will be given a signed copy of their ICF/assent.

7.2. Study Completion and Termination

7.2.1. Study Completion

The study is considered completed with the last visit of the last subject participating in the study. The final data from the investigational site will be sent to the Sponsor (or designee) after completion of the final subject visit at that site, in the time frame specified in the Clinical Trial Agreement.

7.2.2. Study Termination

The Sponsor reserves the right to close the investigational site or terminate the study at any time for any reason at the sole discretion of the Sponsor. Investigational sites will be closed upon study completion. An investigational site is considered closed when all required documents and study supplies have been collected and a site closure visit has been performed. The investigator may initiate site closure at any time, provided there is reasonable cause and sufficient notice is given in advance of the intended termination.

Reasons for the early closure of an investigational site by the Sponsor or investigator may include but are not limited to:

- Failure of the investigator to comply with the protocol, the Sponsor's procedures, or GCP guidelines.
- Inadequate recruitment of subjects by the investigator.
- Discontinuation of further drug development.

7.3. Study Monitoring

Prior to the initiation of the clinical investigation, Sponsor representatives or designees may visit the clinical site where the investigation is to be conducted. Sponsor representatives or designees shall ensure that the Investigator understands the investigational status of the investigational product, all requirements of the investigation to be undertaken, and all of his/her responsibilities as an Investigator. Sponsor representatives or designees will also visit the clinical site at appropriate intervals as required to ensure compliance with the protocol and to verify the accuracy and completeness of data reported on the CRFs. The Study Director or designees shall be available for consultation with the Investigator and serve as liaisons between the clinical site and the Sponsor.

The Sponsor or authorized designees may inspect all documents and records required to be maintained by the Investigator, including but not limited to medical records (office, clinic, or hospital) and investigational product dispensation logs for the subjects in this clinical investigation. The Investigator must permit access to such records. The Investigator must obtain, as part of informed consent, permission for an authorized representative of the Sponsor, or regulatory authorities, to review, in confidence, any records identifying subjects.

7.4. Data Quality Assurance

In order to ensure the collection of accurate, consistent, complete, and reliable data during this clinical investigation, Sponsor representatives or designees may conduct audits of participating sites at appropriate intervals throughout the study. The results of these periodic site audits may be subject to review by independent auditors at completion of the clinical investigation.

All clinical data will undergo a quality control check prior to clinical database lock. Edit checks are performed for appropriate databases as a validation routine using SAS version 9.4 to check for missing data, data inconsistencies, data ranges etc. Corrections are made prior to database lock.

7.5. Data Handling and Record Keeping

During the clinical study, the Investigator will maintain adequate source records, including medical records, records detailing the progress of the study for each subject, laboratory reports, signed informed consent forms, IP disposition records, correspondence with the IRB and Study Monitor/Sponsor, AE reports, and information regarding subject discontinuation and completion of the clinical investigation.

All required study data will be recorded on eCRFs. Any change of data will be recorded on the audit trail and a reason for the change will be entered.

The principal investigator must retain all documentation relating to the study for a period of at least 2 years after the last marketing application approval or, if not approved, 2 years following the discontinuance of the test article for investigation. If this requirement differs from any local regulations, the local regulations will take precedence unless the local retention policy is less than 2 years.

7.6. Protocol Amendments and Deviations

No change or amendment to this protocol may be made by the investigator or Sponsor after the protocol has been agreed to and signed by all parties unless such change(s) or amendment(s) has (have) been agreed upon by the investigator and Sponsor. Any change agreed upon will be recorded in writing, and the written amendment will be signed by the investigator and Sponsor. Institutional review board approval is required prior to the implementation of an amendment, unless overriding safety reasons warrant immediate action, in which case the IRB(s) will be promptly notified.

No deviation from the protocol will be made except to protect the life or physical well-being of a subject in an emergency. Except in such emergencies, prior approval of the Sponsor, and the IRB, is required before deviations from the planned protocol. All protocol deviations that occur during the study will be documented and reported to Sponsor and to the IRB(s), if applicable, according to regulations. Further details about the documentation, evaluation, and follow-up of protocol deviations are detailed in this study's clinical monitoring plan and Protocol Deviation Management Plan (PDMP).

No waivers to inclusion/exclusion criteria will be granted; subjects need to meet all criteria, exactly as specified, to be enrolled. Additionally, prospective deviations from the protocol or investigational plan are not permitted except to protect the life or physical well-being of a subject in an emergency. Deviations that occur unintentionally or are the result of action by the subject must be documented and reported to the IRB(s), if applicable, according to regulations. Further details about the documentation, evaluation, and follow-up of protocol deviations are detailed in this study's clinical monitoring plan and PDMP.

7.7. Confidentiality and Privacy

The investigator must assure that subjects' anonymity will be strictly maintained and that their identities are protected from unauthorized parties. Only an identification code and any other unique identifier(s) as allowed by local law (such as date of birth) will be recorded on any form or biological sample submitted to the Sponsor.The investigator agrees that all information received from Arcutis Biotherapeutics, Inc., including but not limited to the investigator brochure, this protocol, CRF/eCRF, the IP, and any other study information, remain the sole and exclusive property of Arcutis Biotherapeutics, Inc. during the conduct of the study and thereafter. This information is not to be disclosed to any third party (except employees or agents directly involved in the conduct of the study or as required by law) without prior written consent from Arcutis Biotherapeutics, Inc. The investigator further agrees to take all reasonable precautions to prevent the disclosure by any employee or agent of the study site to any third party or otherwise into the public domain.

7.8. Conflict of Interest

All study investigators will provide documentation of their financial interest or arrangements with Arcutis Biotherapeutics, Inc., or proprietary interests in the investigational drug under study. This documentation must be provided prior to the investigator's participation in the study. All investigators with reported conflict of interest will be required to have such conflicts managed in a way that is appropriate to their participation in the design and conduct of this study.

7.9. Report Format

According to the ICH Harmonized Tripartite Guideline (Organization of the Common Technical Document for the Registration of Pharmaceuticals for Human Use M4 and the ICH M2 Expert Working Group), the final report will be written according to the ICH E3 Guideline (Structure and Content of Clinical Study Reports).

7.10. Publication Policy

The Sponsor is supportive of publishing clinical trial findings. Any form of publication that is derived from this study must be submitted to Arcutis Biotherapeutics Inc. for review and approval. The process of coordinating publication efforts is detailed in the Clinical Trial Agreement.

8. REFERENCES

Baumer W, Hoppmann J, Rundfeldt C, et al. Highly selective phosphodiesterase 4 inhibitors for the treatment of allergic skin diseases and psoriasis. *Inflamm Allergy Drug Targets.* 2007;6: 17-26.

Berger RS, Bowman JP. A reappraisal of the 21-day Cumulative Irritation Test in man. *J. Toxicol Ot & Ocular Toxicol.* 1982:1(2);109-115.

Bethke TD, Lahu G. High absolute bioavailability of the new oral phosphodiesterase-4 inhibitor roflumilast. *Int. J Clin Pharmacol Ther.* 2011;49:51-57.

DALIRESP (roflumilast) tablets [package insert]. AstraZeneca Pharmaceuticals LP, Wilmington, DE USA; 2018.

Dastidar SG, Rajagopal D, Ray A. Therapeutic benefit of PDE4 inhibitors in inflammatory diseases. *Curr Opin Investig Drugs*. 2007;8:364–372.

Furue M, Kitahara Y, Akama H, et al. Safety and efficacy of topical E6005, a phosphodiesterase 4 inhibitor, in Japanese adult patients with atopic dermatitis: results of a randomized, vehicle-controlled, multicenter clinical trial. *J Dermatol.* 2014;41:577-585.

Grewe SR, Chan SC, Hanifin JM. Elevated leukocyte cyclic AMP-phosphodiesterase in atopic disease: a possible mechanism for cy-clic AMP-agonist hyporesponsiveness. *J Allergy Clin Immunol.* 1982;70:452–457.

Hanifin JM, Thurston M, Omoto M, et al. The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. *Exp Dermatol.* 2001;10:11-18.

Hanifin JM, Chan SC, Cheng JB, et al. Type 4 phosphodiesterase inhibitors have clinical and in vitro anti-inflammatory effects in atopic dermatitis. *J Invest Dermatol.* 1996;107:51–56.

Hanifin J.M. and Rajka G. Diagnostic features of atopic dermatitis. *Acta Derm Venereol Suppl (Stockh).* 1980;92: 44–47.

Hatzelmann A, Morcillo EJ, Lungarella G, et al. The preclinical pharmacology of roflumilast – a selective, oral phosphodiesterase 4 inhibitor in development for chronic obstructive pulmonary disease. *Pulm Pharmacol Ther.* 2010;23:235–256.

Investigator's Brochure, Edition 4.0, 22 January 2021.

Michalski JM, Golden G, Ikari J, et al. PDE4: A novel target in the treatment of chronic obstructive pulmonary disease. *Clin Pharmacol Ther.* 2012;91:134–142.

Murrell DF, Gebauer K, Spelman L, et al. Crisaborole topical ointment, 2% in adults with atopic dermatitis: a phase 2a, vehicle- controlled, proof-of-concept study. *J Drugs Dermatol.* 2015;14:1108–1112.

Nazarian R, Weinberg JM. AN-2728, a PDE4 inhibitor for the potential topical treatment of psoriasis and atopic dermatitis. *Curr Opin Investig Drugs.* 2009;10:1236–1242.

Newton et al., Exploring Content and Psychometric Validity of Newly Developed Assessment Tools for Itch and Skin Pain in Atopic Dermatitis. *Journal of Patient-Reported Outcomes*. 2019;3:42.

Nygaard U, Deleuran M, Vestergaard C. Emerging treatment options in atopic dermatitis: topical therapies. *Dermatology.* 2017;233:333-343.

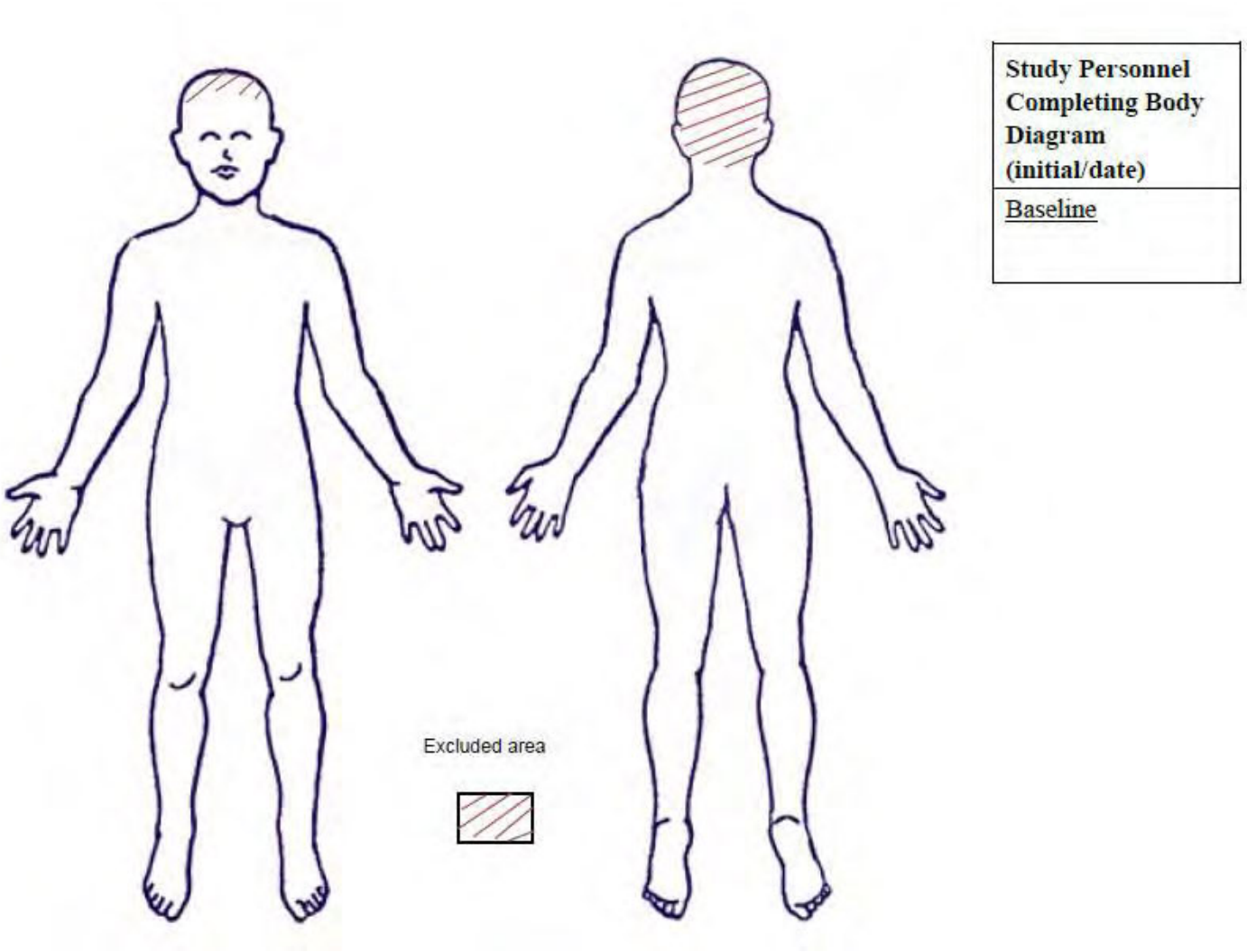
Wedzicha JA, Calverley PMA, Klaus FR. Roflumilast: a review of its use in the treatment of COPD. *Int J Chron Obstruct Pulmon Dis.* 2016;11:81-90.

9. APPENDICES

APPENDIX 1. BODY DIAGRAM

Site personnel to mark treatable areas identified by the Investigator.

(Reminder: Application will be all areas affected (except for the scalp). Continue to apply even if area(s) clears and treat new lesions (except scalp).



Site to photocopy this page after updating at the Baseline and retain the original in source. Provide the copy to the subject to refer to for study application at home.

APPENDIX 2. PATIENT HEALTH QUESTIONAIRE-8 (PHQ-8)

SMRC
Self-Management
Resource Center

Personal Health Questionnaire
Depression Scale (PHQ-8)

Over the **last 2 weeks**, how often have you been bothered by any of the following problems?
*(circle **one** number on each line)*

How often during the past 2 weeks were you bothered by...	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself, or that you are a failure, or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3

Scoring

If two consecutive numbers are circled, score the higher (more distress) number. If the numbers are not consecutive, do not score the item. Score is the sum of the 8 items. If more than 1 item missing, set the value of the scale to missing. A score of 10 or greater is considered major depression, 20 or more is severe major depression.

APPENDIX 3. PATIENT HEALTH QUESTIONNAIRE DEPRESSION SCALE (MODIFIED PHQ-A)

Instructions: How often have you been bothered by each of the following symptoms during the past **two weeks**? For each symptom put an **"X"** in the box beneath the answer that best describes how you have been feeling.

	(0) Not at all	(1) Several days	(2) More than half the days	(3) Nearly every day
1. Feeling down, depressed, irritable, or hopeless?				
2. Little interest or pleasure in doing things?				
3. Trouble falling asleep, staying asleep, or sleeping too much?				
4. Poor appetite, weight loss, or overeating?				
5. Feeling tired, or having little energy?				
6. Feeling bad about yourself – or feeling that you are a failure, or that you have let yourself or your family down?				
7. Trouble concentrating on things like school work, reading, or watching TV?				
8. Moving or speaking so slowly that other people could have noticed? Or the opposite – being so fidgety or restless that you were moving around a lot more than usual?				

APPENDIX 4. CHILDREN'S DEPRESSION INVENTORY 2 (PARENT REPORT)

By Maria Kovacs, Ph.D.

CDI 2 PARENT

Child's Name/ID: _____ Child's Sex: Male Female

Parent's Name/ID: _____ Date of Birth: ____/____/____

Relationship to Child: _____ Today's Date: ____/____/____

Child's Age: _____ Child's Grade: _____

Instructions:

For each of the statements below, select one response that best describes your observations of your child in the **past two weeks**.

Indicate your response for each item by **circling** the number that best corresponds to your choice. You may change an item response by drawing an X through your original choice and selecting a new response.

Remember, for each statement, pick **one** answer that best describes your observations of your child in the PAST TWO WEEKS.

My child	Not at all	Some of the time	Often	Much or most of the time
1. looks sad	0	1	2	3
2. has fun.	0	1	2	3
3. does not like himself or herself.	0	1	2	3
4. blames himself or herself for things.	0	1	2	3
5. cries or looks tearful.	0	1	2	3
6. is cranky or irritable.	0	1	2	3
7. enjoys being with people.	0	1	2	3
8. thinks that he or she is ugly.	0	1	2	3
9. has to push himself or herself to do schoolwork	0	1	2	3
10. has trouble sleeping at night	0	1	2	3
11. looks tired or fatigued.	0	1	2	3
12. seems lonely.	0	1	2	3
13. enjoys school.	0	1	2	3
14. spends time with friends.	0	1	2	3
15. is showing worse school performance than before.	0	1	2	3
16. does what he or she is told.	0	1	2	3
17. has disagreements and conflicts with others.	0	1	2	3

APPENDIX 5. COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS) "BASELINE/SCREENING" VERSION

COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS)

Baseline/Screening Version

Version 1/14/09

Posner, K.; Brent, D.; Lucas, C.; Gould, M.; Stanley, B.; Brown, G.; Fisher, P.; Zelazny, J.; Burke, A.; Oquendo, M.; Mann, J.

Disclaimer:

This scale is intended to be used by individuals who have received training in its administration. The questions contained in the Columbia-Suicide Severity Rating Scale are suggested probes. Ultimately, the determination of the presence of suicidal ideation or behavior depends on the judgment of the individual administering the scale.

Definitions of behavioral suicidal events in this scale are based on those used in ***The Columbia Suicide History Form****, developed by John Mann, MD and Maria Oquendo, MD, Conte Center for the Neuroscience of Mental Disorders (CCNMD), New York State Psychiatric Institute, 1051 Riverside Drive, New York, NY, 10032. (Oquendo M. A., Halberstam B. & Mann J. J., Risk factors for suicidal behavior: utility and limitations of research instruments. In M.B. First [Ed.]* Standardized Evaluation in Clinical Practice*, pp. 103 -130, 2003.)*

For reprints of the C-SSRS contact Kelly Posner, Ph.D., New York State Psychiatric Institute, 1051 Riverside Drive, New York, New York, 10032; inquiries and training requirements contact posnerk@nyspi.columbia.edu

SUICIDAL IDEATION		
Ask questions 1 and 2. If both are negative, proceed to "Suicidal Behavior" section. If the answer to question 2 is "yes", ask questions 3, 4 and 5. If the answer to question 1 and/or 2 is "yes", complete "Intensity of Ideation" section below.	**Lifetime: Time He/She Felt Most Suicidal**	**Past __ Months**
1. Wish to be Dead Subject endorses thoughts about a wish to be dead or not alive anymore, or wish to fall asleep and not wake up ***Have you wished you were dead or wished you could go to sleep and not wake up?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
2. Non-Specific Active Suicidal Thoughts General non-specific thoughts of wanting to end one's life/commit suicide (e g , *"I ve thought about killing myself"*) without thoughts of ways to kill oneself/associated methods, intent, or plan during the assessment period ***Have you actually had any thoughts of killing yourself?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
3. Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act Subject endorses thoughts of suicide and has thought of at least one method during the assessment period This is different than a specific plan with time, place or method details worked out (e g thought of method to kill self but not a specific plan) Includes person who would say, *"I thought about taking an overdose but I never made a specific plan as to when, where or how I would actually do it…and I would never go through with it."* ***Have you been thinking about how you might do this?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
4. Active Suicidal Ideation with Some Intent to Act, without Specific Plan Active suicidal thoughts of killing oneself and subject reports having <u>some intent to act on such thoughts</u> as opposed to *"I have the thoughts but I definitely will not do anything about them."* ***Have you had these thoughts and had some intention of acting on them?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
5. Active Suicidal Ideation with Specific Plan and Intent Thoughts of killing oneself with details of plan fully or partially worked out and subject has some intent to carry it out ***Have you started to work out or worked out the details of how to kill yourself? Do you intend to carry out this plan?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
INTENSITY OF IDEATION		
The following features should be rated with respect to the most severe type of ideation (i.e., 1-5 from above, with 1 being the least severe and 5 being the most severe). Ask about time he/she was feeling the most suicidal. <u>Lifetime</u> - ***Most Severe Ideation:*** _____ (*Type # (1-5)*) _____ (*Description of Ideation*) <u>Past X Months</u> - ***Most Severe Ideation:*** _____ (*Type # (1-5)*) _____ (*Description of Ideation*)	Most Severe	Most Severe
Frequency ***How many times have you had these thoughts?*** (1) Less than once a week (2) Once a week (3) 2-5 times in week (4) Daily or almost daily (5) Many times each day	____	____
Duration ***When you have the thoughts how long do they last?*** (1) Fleeting - few seconds or minutes (2) Less than 1 hour/some of the time (3) 1-4 hours/a lot of time (4) 4-8 hours/most of day (5) More than 8 hours/persistent or continuous	____	____
Controllability ***Could/can you stop thinking about killing yourself or wanting to die if you want to?*** (1) Easily able to control thoughts (2) Can control thoughts with little difficulty (3) Can control thoughts with some difficulty (4) Can control thoughts with a lot of difficulty (5) Unable to control thoughts (0) Does not attempt to control thoughts	____	____
Deterrents ***Are there things - anyone or anything (e.g., family, religion, pain of death) - that stopped you from wanting to die or acting on thoughts of committing suicide?*** (1) Deterrents definitely stopped you from attempting suicide (2) Deterrents probably stopped you (3) Uncertain that deterrents stopped you (4) Deterrents most likely did not stop you (5) Deterrents definitely did not stop you (0) Does not apply	____	____
Reasons for Ideation ***What sort of reasons did you have for thinking about wanting to die or killing yourself? Was it to end the pain or stop the way you were feeling (in other words you couldn't go on living with this pain or how you were feeling) or was it to get attention, revenge or a reaction from others? Or both?*** (1) Completely to get attention, revenge or a reaction from others (2) Mostly to get attention, revenge or a reaction from others (3) Equally to get attention, revenge or a reaction from others and to end/stop the pain (4) Mostly to end or stop the pain (you couldn't go on living with the pain or how you were feeling) (5) Completely to end or stop the pain (you couldn't go on living with the pain or how you were feeling) (0) Does not apply	____	____

SUICIDAL BEHAVIOR *(Check all that apply, so long as these are separate events; must ask about all types)*	**Lifetime**	**Past _ Years**
Actual Attempt: A potentially self-injurious act committed with at least some wish to die, *as a result of act.* Behavior was in part thought of as method to kill oneself Intent does not have to be 100% If there is ***any*** intent/desire to die associated with the act, then it can be considered an actual suicide attempt ***There does not have to be any injury or harm***, just the potential for injury or harm If person pulls trigger while gun is in mouth but gun is broken so no injury results, this is considered an attempt Inferring Intent: Even if an individual denies intent/wish to die, it may be inferred clinically from the behavior or circumstances For example, a highly lethal act that is clearly not an accident so no other intent but suicide can be inferred (e g , gunshot to head, jumping from window of a high floor/story) Also, if someone denies intent to die, but they thought that what they did could be lethal, intent may be inferred ***Have you made a suicide attempt?*** ***Have you done anything to harm yourself?*** ***Have you done anything dangerous where you could have died?*** ***What did you do?*** ***Did you_____ as a way to end your life?*** ***Did you want to die (even a little) when you_____?*** ***Were you trying to end your life when you _____?*** ***Or Did you think it was possible you could have died from_____?*** ***Or did you do it purely for other reasons / without ANY intention of killing yourself (like to relieve stress, feel better, get sympathy, or get something else to happen)?*** *(*Self-Injurious Behavior without suicidal intent) If yes, describe:	**Yes** ☐ **No** ☐ Total # of Attempts _____	**Yes** ☐ **No** ☐ Total # of Attempts _____
Has subject engaged in Non-Suicidal Self-Injurious Behavior?	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
Interrupted Attempt: When the person is interrupted (by an outside circumstance) from starting the potentially self-injurious act *(if not for that, actual attempt would have occurred).* Overdose: Person has pills in hand but is stopped from ingesting Once they ingest any pills, this becomes an attempt rather than an interrupted attempt Shooting: Person has gun pointed toward self, gun is taken away by someone else, or is somehow prevented from pulling trigger Once they pull the trigger, even if the gun fails to fire, it is an attempt Jumping: Person is poised to jump, is grabbed and taken down from ledge Hanging: Person has noose around neck but has not yet started to hang - is stopped from doing so ***Has there been a time when you started to do something to end your life but someone or something stopped you before you actually did anything?*** If yes, describe:	**Yes** ☐ **No** ☐ Total # of interrupted _____	**Yes** ☐ **No** ☐ Total # of interrupted _____
Aborted Attempt: When person begins to take steps toward making a suicide attempt, but stops themselves before they actually have engaged in any self-destructive behavior Examples are similar to interrupted attempts, except that the individual stops him/herself, instead of being stopped by something else ***Has there been a time when you started to do something to try to end your life but you stopped yourself before you actually did anything?*** If yes, describe:	**Yes** ☐ **No** ☐ Total # of aborted _____	**Yes** ☐ **No** ☐ Total # of aborted _____
Preparatory Acts or Behavior: Acts or preparation towards imminently making a suicide attempt This can include anything beyond a verbalization or thought, such as assembling a specific method (e g , buying pills, purchasing a gun) or preparing for one's death by suicide (e g , giving things away, writing a suicide note) ***Have you taken any steps towards making a suicide attempt or preparing to kill yourself (such as collecting pills, getting a gun, giving valuables away or writing a suicide note)?*** If yes, describe:	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐
Suicidal Behavior: Suicidal behavior was present during the assessment period?	**Yes** ☐ **No** ☐	**Yes** ☐ **No** ☐

Answer for Actual Attempts Only	Most Recent Attempt Date:	Most Lethal Attempt Date:	Initial/First Attempt Date:
Actual Lethality/Medical Damage: 0 No physical damage or very minor physical damage (e g , surface scratches) 1 Minor physical damage (e g , lethargic speech; first-degree burns; mild bleeding; sprains) 2 Moderate physical damage; medical attention needed (e g , conscious but sleepy, somewhat responsive; second-degree burns; bleeding of major vessel) 3 Moderately severe physical damage; *medical* hospitalization and likely intensive care required (e g , comatose with reflexes intact; third-degree burns less than 20% of body; extensive blood loss but can recover; major fractures) 4 Severe physical damage; *medical* hospitalization with intensive care required (e g , comatose without reflexes; third-degree burns over 20% of body; extensive blood loss with unstable vital signs; major damage to a vital area) 5 Death	*Enter Code* _____	*Enter Code* _____	*Enter Code* _____
Potential Lethality: Only Answer if Actual Lethality=0 Likely lethality of actual attempt if no medical damage (the following examples, while having no actual medical damage, had potential for very serious lethality: put gun in mouth and pulled the trigger but gun fails to fire so no medical damage; laying on train tracks with oncoming train but pulled away before run over) 0 = Behavior not likely to result in injury 1 = Behavior likely to result in injury but not likely to cause death 2 = Behavior likely to result in death despite available medical care	*Enter Code* _____	*Enter Code* _____	*Enter Code* _____

APPENDIX 6. COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS) "SINCE LAST VISIT" VERSION

COLUMBIA-SUICIDE SEVERITY RATING SCALE (C-SSRS)

Since Last Visit

Version 1/14/09

Posner, K.; Brent, D.; Lucas, C.; Gould, M.; Stanley, B.; Brown, G.; Fisher, P.; Zelazny, J.; Burke, A.; Oquendo, M.; Mann, J.

Disclaimer:

This scale is intended to be used by individuals who have received training in its administration. The questions contained in the Columbia-Suicide Severity Rating Scale are suggested probes. Ultimately, the determination of the presence of suicidal ideation or behavior depends on the judgment of the individual administering the scale.

Definitions of behavioral suicidal events in this scale are based on those used in ***The Columbia Suicide History Form****, developed by John Mann, MD and Maria Oquendo, MD, Conte Center for the Neuroscience of Mental Disorders (CCNMD), New York State Psychiatric Institute, 1051 Riverside Drive, New York, NY, 10032. (Oquendo M. A., Halberstam B. & Mann J. J., Risk factors for suicidal behavior: utility and limitations of research instruments. In M.B. First [Ed.]* Standardized Evaluation in Clinical Practice*, pp. 103 -130, 2003.)*

For reprints of the C-SSRS contact Kelly Posner, Ph.D., New York State Psychiatric Institute, 1051 Riverside Drive, New York, New York, 10032; inquiries and training requirements contact posnerk@nyspi.columbia.edu

SUICIDAL IDEATION

Ask questions 1 and 2. If both are negative, proceed to "Suicidal Behavior" section. If the answer to question 2 is "yes", ask questions 3, 4 and 5. If the answer to question 1 and/or 2 is "yes", complete "Intensity of Ideation" section below.

Since Last Visit

1. Wish to be Dead
Subject endorses thoughts about a wish to be dead or not alive anymore, or wish to fall asleep and not wake up
Have you wished you were dead or wished you could go to sleep and not wake up?
If yes, describe:
Yes ☐ **No** ☐

2. Non-Specific Active Suicidal Thoughts
General, non-specific thoughts of wanting to end one's life/commit suicide (e g , *"I ve thought about killing myself"*) without thoughts of ways to kill oneself/associated methods, intent, or plan during the assessment period
Have you actually had any thoughts of killing yourself?
If yes, describe:
Yes ☐ **No** ☐

3. Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
Subject endorses thoughts of suicide and has thought of at least one method during the assessment period This is different than a specific plan with time, place or method details worked out (e g , thought of method to kill self but not a specific plan) Includes person who would say, *"I thought about taking an overdose but I never made a specific plan as to when, where or how I would actually do it…and I would never go through with it."*
Have you been thinking about how you might do this?
If yes, describe:
Yes ☐ **No** ☐

4. Active Suicidal Ideation with Some Intent to Act, without Specific Plan
Active suicidal thoughts of killing oneself and subject reports having some intent to act on such thoughts as opposed to *"I have the thoughts but I definitely will not do anything about them."*
Have you had these thoughts and had some intention of acting on them?
If yes, describe:
Yes ☐ **No** ☐

5. Active Suicidal Ideation with Specific Plan and Intent
Thoughts of killing oneself with details of plan fully or partially worked out and subject has some intent to carry it out
Have you started to work out or worked out the details of how to kill yourself? Do you intend to carry out this plan?
If yes, describe:
Yes ☐ **No** ☐

INTENSITY OF IDEATION

The following features should be rated with respect to the most severe type of ideation (i.e., 1-5 from above, with 1 being the least severe and 5 being the most severe).

Most Severe Ideation: _____ _____

Type # (1-5) ***Description of Ideation***

Most Severe

Frequency
How many times have you had these thoughts?
(1) Less than once a week (2) Once a week (3) 2-5 times in week (4) Daily or almost daily (5) Many times each day
____

Duration
When you have the thoughts, how long do they last?
(1) Fleeting - few seconds or minutes
(2) Less than 1 hour/some of the time
(3) 1-4 hours/a lot of time
(4) 4-8 hours/most of day
(5) More than 8 hours/persistent or continuous
____

Controllability
Could/can you stop thinking about killing yourself or wanting to die if you want to?
(1) Easily able to control thoughts
(2) Can control thoughts with little difficulty
(3) Can control thoughts with some difficulty
(4) Can control thoughts with a lot of difficulty
(5) Unable to control thoughts
(0) Does not attempt to control thoughts
____

Deterrents
Are there things - anyone or anything (e.g., family, religion, pain of death) - that stopped you from wanting to die or acting on thoughts of committing suicide?
(1) Deterrents definitely stopped you from attempting suicide
(2) Deterrents probably stopped you
(3) Uncertain that deterrents stopped you
(4) Deterrents most likely did not stop you
(5) Deterrents definitely did not stop you
(0) Does not apply
____

Reasons for Ideation
What sort of reasons did you have for thinking about wanting to die or killing yourself? Was it to end the pain or stop the way you were feeling (in other words you couldn't go on living with this pain or how you were feeling) or was it to get attention, revenge or a reaction from others? Or both?
(1) Completely to get attention, revenge or a reaction from others
(2) Mostly to get attention, revenge or a reaction from others
(3) Equally to get attention, revenge or a reaction from others and to end/stop the pain
(4) Mostly to end or stop the pain (you couldn't go on living with the pain or how you were feeling)
(5) Completely to end or stop the pain (you couldn't go on living with the pain or how you were feeling)
(0) Does not apply
____

SUICIDAL BEHAVIOR *(Check all that apply, so long as these are separate events; must ask about all types)*	**Since Last Visit**
Actual Attempt: A potentially self-injurious act committed with at least some wish to die, *as a result of act.* Behavior was in part thought of as method to kill oneself Intent does not have to be 100% If there is ***any*** intent/desire to die associated with the act, then it can be considered an actual suicide attempt ***There does not have to be any injury or harm***, just the potential for injury or harm If person pulls trigger while gun is in mouth but gun is broken so no injury results, this is considered an attempt Inferring Intent: Even if an individual denies intent/wish to die, it may be inferred clinically from the behavior or circumstances For example, a highly lethal act that is clearly not an accident so no other intent but suicide can be inferred (e g , gunshot to head, jumping from window of a high floor/story) Also, if someone denies intent to die, but they thought that what they did could be lethal, intent may be inferred ***Have you made a suicide attempt?*** ***Have you done anything to harm yourself?*** ***Have you done anything dangerous where you could have died?*** ***What did you do?*** ***Did you_____ as a way to end your life?*** ***Did you want to die (even a little) when you_____?*** ***Were you trying to end your life when you _____?*** ***Or did you think it was possible you could have died from_____?*** ***Or did you do it purely for other reasons / without ANY intention of killing yourself (like to relieve stress, feel better, get sympathy, or get something else to happen)?*** *(*Self-Injurious Behavior without suicidal intent) If yes, describe: **Has subject engaged in Non-Suicidal Self-Injurious Behavior?**	**Yes** ☐ **No** ☐ Total # of Attempts _____ **Yes** ☐ **No** ☐
Interrupted Attempt: When the person is interrupted (by an outside circumstance) from starting the potentially self-injurious act *(if not for that, actual attempt would have occurred).* Overdose: Person has pills in hand but is stopped from ingesting Once they ingest any pills, this becomes an attempt rather than an interrupted attempt Shooting: Person has gun pointed toward self, gun is taken away by someone else, or is somehow prevented from pulling trigger Once they pull the trigger, even if the gun fails to fire, it is an attempt Jumping: Person is poised to jump, is grabbed and taken down from ledge Hanging: Person has noose around neck but has not yet started to hang - is stopped from doing so ***Has there been a time when you started to do something to end your life but someone or something stopped you before you actually did anything?*** If yes, describe:	**Yes** ☐ **No** ☐ Total # of interrupted _____
Aborted Attempt: When person begins to take steps toward making a suicide attempt, but stops themselves before they actually have engaged in any self-destructive behavior Examples are similar to interrupted attempts, except that the individual stops him/herself, instead of being stopped by something else ***Has there been a time when you started to do something to try to end your life but you stopped yourself before you actually did anything?*** If yes, describe:	**Yes** ☐ **No** ☐ Total # of aborted _____
Preparatory Acts or Behavior: Acts or preparation towards imminently making a suicide attempt This can include anything beyond a verbalization or thought, such as assembling a specific method (e g , buying pills, purchasing a gun) or preparing for one's death by suicide (e g , giving things away, writing a suicide note) ***Have you taken any steps towards making a suicide attempt or preparing to kill yourself (such as collecting pills, getting a gun, giving valuables away or writing a suicide note)?*** If yes, describe:	**Yes** ☐ **No** ☐
Suicidal Behavior: Suicidal behavior was present during the assessment period?	**Yes** ☐ **No** ☐
Suicide:	**Yes** ☐ **No** ☐
Answer for Actual Attempts Only	Most Lethal Attempt Date:
Actual Lethality/Medical Damage: 0 No physical damage or very minor physical damage (e g , surface scratches) 1 Minor physical damage (e g , lethargic speech; first-degree burns; mild bleeding; sprains) 2 Moderate physical damage; medical attention needed (e g , conscious but sleepy, somewhat responsive; second-degree burns; bleeding of major vessel) 3 Moderately severe physical damage; *medical* hospitalization and likely intensive care required (e g , comatose with reflexes intact; third-degree burns less than 20% of body; extensive blood loss but can recover; major fractures) 4 Severe physical damage; *medical* hospitalization with intensive care required (e g , comatose without reflexes; third-degree burns over 20% of body; extensive blood loss with unstable vital signs; major damage to a vital area) 5 Death	*Enter Code* _____
Potential Lethality: Only Answer if Actual Lethality=0 Likely lethality of actual attempt if no medical damage (the following examples, while having no actual medical damage, had potential for very serious lethality: put gun in mouth and pulled the trigger but gun fails to fire so no medical damage; laying on train tracks with oncoming train but pulled away before run over) 0 = Behavior not likely to result in injury 1 = Behavior likely to result in injury but not likely to cause death 2 = Behavior likely to result in death despite available medical care	*Enter Code* _____

APPENDIX 7. VALIDATED INVESTIGATOR GLOBAL ASSESSMENT SCALE FOR ATOPIC DERMATITIS

Validated Investigator Global Assessment scale for Atopic Dermatitis

vIGA-AD™

Instructions:

The IGA score is selected using the descriptors below that best describe the overall appearance of the lesions at a given time point. It is not necessary that all characteristics under Morphological Description be present.

Score	Morphological Description
0 – Clear	No inflammatory signs of atopic dermatitis (no erythema, no induration/papulation, no lichenification, no oozing/crusting). Post-inflammatory hyperpigmentation and/or hypopigmentation may be present.
1 – Almost clear	Barely perceptible erythema, barely perceptible induration/papulation, and/or minimal lichenification. No oozing or crusting.
2 – Mild	Slight but definite erythema (pink), slight but definite induration/papulation, and/or slight but definite lichenification. No oozing or crusting.
3 – Moderate	Clearly perceptible erythema (dull red), clearly perceptible induration/papulation, and/or clearly perceptible lichenification. Oozing and crusting may be present.
4 – Severe	Marked erythema (deep or bright red), marked induration/papulation, and/or marked lichenification. Disease is widespread in extent. Oozing or crusting may be present.

Notes:

1. In indeterminate cases, please use extent to differentiate between scores.

For example:

- *Patient with marked erythema (deep or bright red), marked papulation and/or marked lichenification that is limited in extent, will be considered "3 – Moderate".*

2. Excoriations should not be considered when assessing disease severity.

APPENDIX 8. DERMATOLOGY LIFE QUALITY INDEX

DLQI

Site No: Date: Score:
Name:
Address: Diagnosis:

The aim of this questionnaire is to measure how much your skin problem has affected your life OVER THE LAST WEEK. Please tick ☑ one box for each question.

1.	Over the last week, how **itchy**, **sore**, **painful** or **stinging** has your skin been?	Very much A lot A little Not at all	❒ ❒ ❒ ❒
2.	Over the last week, how **embarrassed** or **self conscious** have you been because of your skin?	Very much A lot A little Not at all	❒ ❒ ❒ ❒
3.	Over the last week, how much has your skin interfered with you going **shopping** or looking after your **home** or **garden**?	Very much A lot A little Not at all Not relevant	❒ ❒ ❒ ❒ ❒
4.	Over the last week, how much has your skin influenced the **clothes** you wear?	Very much A lot A little Not at all Not relevant	❒ ❒ ❒ ❒ ❒
5.	Over the last week, how much has your skin affected any **social** or **leisure** activities?	Very much A lot A little Not at all Not relevant	❒ ❒ ❒ ❒ ❒
6.	Over the last week, how much has your skin made it difficult for you to do any **sport**?	Very much A lot A little Not at all Not relevant	❒ ❒ ❒ ❒ ❒
7.	Over the last week, has your skin prevented you from **working** or **studying**?	Yes No Not relevant	❒ ❒ ❒
	If "No", over the last week how much has your skin been a problem at **work** or **studying**?	A lot A little Not at all	❒ ❒ ❒

8.	Over the last week, how much has your skin created problems with your **partner** or any of your **close friends** or **relatives**?	Very much ❒ A lot ❒ A little ❒ Not at all ❒ Not relevant ❒
9.	Over the last week, how much has your skin caused any **sexual difficulties**?	Very much ❒ A lot ❒ A little ❒ Not at all ❒ Not relevant ❒
10.	Over the last week, how much of a problem has the **treatment** for your skin been, for example by making your home messy, or by taking up time?	Very much ❒ A lot ❒ A little ❒ Not at all ❒ Not relevant ❒

Please check you have answered EVERY question. Thank you.

APPENDIX 9. CHILDREN'S DERMATOLOGY LIFE QUALITY INDEX

Site No.:
Name: Diagnosis: CDLQI
Age: SCORE:
Address: Date:

The aim of this questionnaire is to measure how much your skin problem has affected you OVER THE LAST WEEK. Please tick ✓ one box for each question.

1.	Over the last week, how **itchy**, **"scratchy"**, **sore** or **painful** has your skin been?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
2.	Over the last week, how **embarrassed** or **self conscious**, **upset** or **sad** have you been because of your skin?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
3.	Over the last week, how much has your skin affected your **friendships**?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
4.	Over the last week, how much have you changed or worn **different** or **special clothes/shoes** because of your skin?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
5.	Over the last week, how much has your skin trouble affected **going out**, **playing**, or **doing hobbies**?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
6.	Over the last week, how much have you avoided **swimming** or **other sports** because of your skin trouble?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
7.	<u>Last week</u>, was it **school time?** **OR** was it **holiday time?**	**If school time**: Over the last week, how much did your skin problem affect your **school work**? **If holiday time**: How much over the last week, has your skin problem interfered with your enjoyment of the **holiday**?	Prevented school Very much Quite a lot Only a little Not at all Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

8.	Over the last week, how much trouble have you had because of your skin with other people **calling you names**, **teasing**, **bullying**, **asking questions** or **avoiding you**?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
9.	Over the last week, how much has your **sleep** been affected by your skin problem?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
10.	Over the last week, how much of a problem has the **treatment** for your skin been?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐

APPENDIX 10. DERMATITIS FAMILY IMPACT QUESTIONNAIRE (DFI)

Child's Name: **Mother/Father/Carer** **Date:** **Score**

The aim of this questionnaire is to measure how much your child's skin problem has affected you and your family OVER THE LAST WEEK. Please tick ☑ one box for each question.

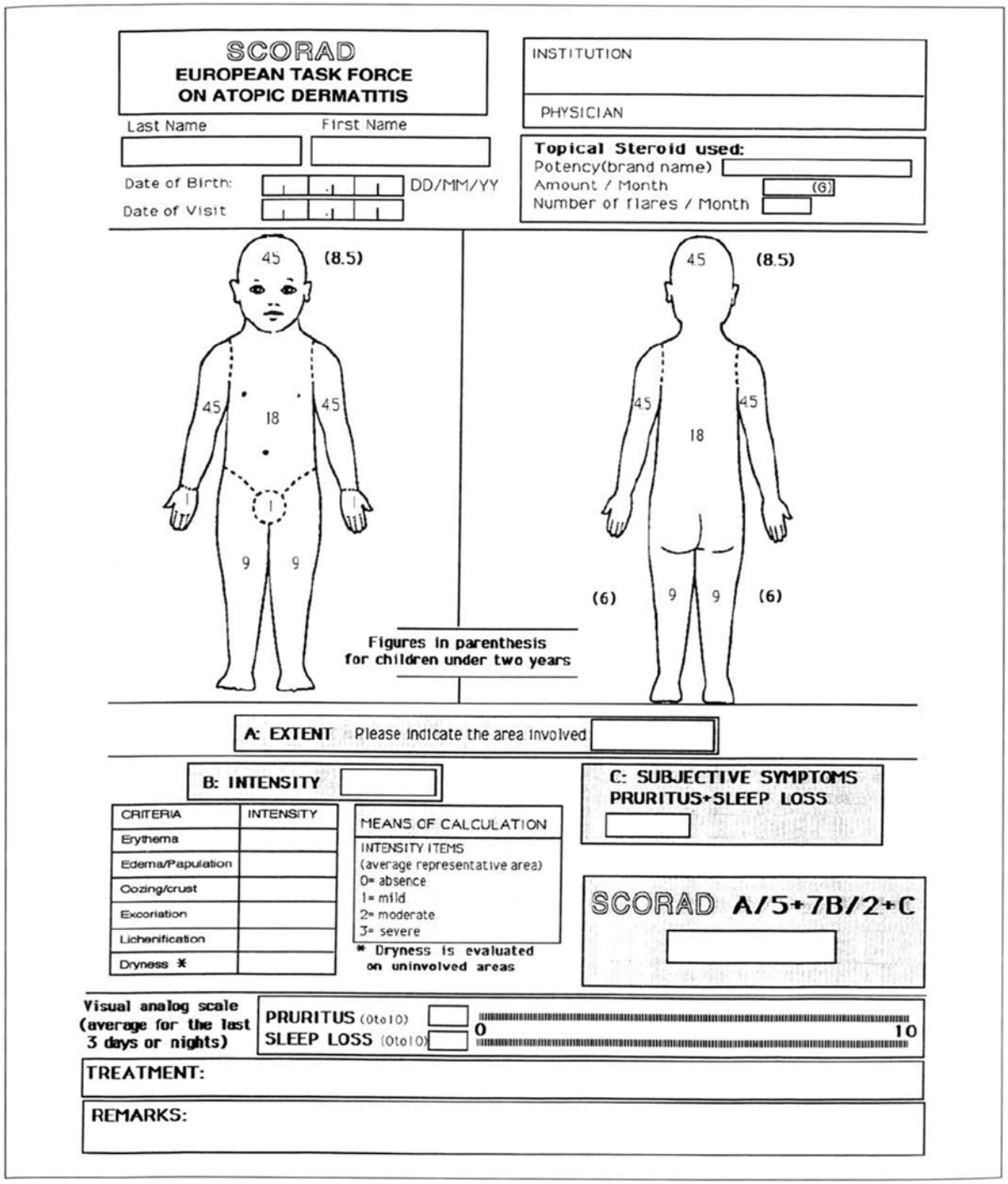
1.	Over the last week, how much effect has your child having eczema had on **housework**, e.g. washing, cleaning.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
2.	Over the last week, how much effect has your child having eczema had on **food preparation** and **feeding**.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
3.	Over the last week, how much effect has your child having eczema had on the **sleep of others in family**.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
4.	Over the last week, how much effect has your child having eczema had on **family leisure activities**, eg swimming.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
5.	Over the last week, how much effect has your child having eczema had on **time spent on shopping for the family.**	Very much A lot A little Not at all	❒ ❒ ❒ ❒
6.	Over the last week, how much effect has your child having eczema had on your **expenditure**, eg costs related to treatment, clothes, etc.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
7.	Over the last week, how much effect has your child having eczema had on causing **tiredness** or **exhaustion** in your child's parents/carers.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
8.	Over the last week, how much effect has your child having eczema had on causing **emotional distress** such as depression, frustration or guilt in your child's parents/carers.	Very much A lot A little Not at all	❒ ❒ ❒ ❒

9.	Over the last week, how much effect has your child having eczema had on **relationships** between the **main carer and partner** or between the **main carer and other children** in the family.	Very much A lot A little Not at all	❒ ❒ ❒ ❒
10.	Over the last week, how much effect has **helping with your child's treatment** had on the main carer's life.	Very much A lot A little Not at all	❒ ❒ ❒ ❒

Please check you have answered EVERY question. Thank you

APPENDIX 11. SCORAD

SCORAD https://www.ncbi.nlm.nih.gov/pubmed/8435513



APPENDIX 12. PATIENT-ORIENTED ECZEMA MEASURE (POEM)

POEM
Patient-Oriented Eczema Measure

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

POEM for self-completion and/or proxy completion

Patient Details: _____

_____

_____

_____ Date: _____

Please circle one response for each of the seven questions below about your/your child's eczema. If your child is old enough to understand the questions then please fill in the questionnaire together. Please leave blank any questions you feel unable to answer.

1. Over the last week, on how many days has your/your child's skin been itchy because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

2. Over the last week, on how many nights has your/your child's sleep been disturbed because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

3. Over the last week, on how many days has your/your child's skin been bleeding because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

4. Over the last week, on how many days has your/your child's skin been weeping or oozing clear fluid because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

5. Over the last week, on how many days has your/your child's skin been cracked because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

6. Over the last week, on how many days has your/your child's skin been flaking off because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

7. Over the last week, on how many days has your/your child's skin felt dry or rough because of the eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

Total POEM Score (Maximum 28): .....

POEM
Patient-Oriented Eczema Measure

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

POEM for self-completion and/or proxy completion

How is the scoring done?

Each of the seven questions carries equal weight and is scored from 0 to 4 as follows:

No days	= 0
1-2 days	= 1
3-4 days	= 2
5-6 days	= 3
Every day	= 4

Note:

- If one question is left unanswered this is scored 0 and the scores are summed and expressed as usual out of a maximum of 28
- If two or more questions are left unanswered the questionnaire is not scored
- If two or more response options are selected, the response option with the highest score should be recorded

What does a poem score mean?

To help patients and clinicians to understand their POEM scores, the following bandings have been established (see references below):

• 0 to 2	= Clear or almost clear
• 3 to 7	= Mild eczema
• 8 to 16	= Moderate eczema
• 17 to 24	= Severe eczema
• 25 to 28	= Very severe eczema

Do I need permission to use the scale?

Whilst the POEM scale is protected by copyright, it is freely available for use and can be downloaded from: www.nottingham.ac.uk/dermatology
We do however ask that you register your use of the POEM by e-mailing cebd@nottingham.ac.uk with details of how you would like to use the scale, and which countries the scale will be used in.

References

Charman CR, Venn AJ, Williams HC. The Patient-Oriented Eczema Measure: Development and Initial Validation of a New Tool for Measuring Atopic Eczema Severity From the Patients' Perspective. Arch Dermatol. 2004;140:1513-1519

Charman CR, Venn AJ, Ravenscroft JC, Williams HC. Translating Patient-Oriented Eczema Measure (POEM) scores into clinical practice by suggesting severity strata derived using anchor-based methods. Br J Dermatol. Dec 2013; 169(6): 1326–1332.

APPENDIX 13. NIAID DMID TOXICITY TABLE

NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DISEASES, DIVISION OF MICROBIOLOGY AND INFECTIOUS DISEASES (DMID) Toxicity Table for Use in Trials Enrolling Healthy Adults (2014) Modified

ABBREVIATIONS USED IN FOLLOWING TABLES:

Abbreviation/ Term	Definition/Explanation
ALT	alanine aminotransferase
aPTT	activated partial thromboplastin time
AST	aspartate aminotransferase
AV block	atrioventricular block
bpm	beats per minute
BUN	blood urea nitrogen
CK	creatine kinase
CPK	creatine phosphokinase
FEV_1	forced expiratory volume in 1 second
g	Gram
HI	High
HPF	high power field
IU	international unit
IV	Intravenous
K/CUMM	$x10^3/mm^3$
LLN	lower limit of normal

Abbreviation/ Term	Definition/Explanation
LO	Low
mEq	Milliequivalent
mmHg	millimeter of mercury
Ms	Millisecond
N	Normal
PT	prothrombin time
PTT	partial thromboplastin time
QTc	QT-interval corrected for heart rate
QTcB	Bazett's corrected QT interval
QTcF	Fridericia's corrected QT interval
RBC	red blood cell
Rx	Therapy
S	Second
U	Unit
ULN	upper limit of normal

ESTIMATING SEVERITY GRADE

For abnormalities NOT found elsewhere in the Toxicity Tables use the scale below to estimate grade of severity:

GRADE 1	**Mild**:	Transient or mild discomfort (<48 hours); no medical intervention/therapy required
GRADE 2	**Moderate**:	Mild to moderate limitation in activity - some assistance may be needed; no or minimal medical intervention/therapy required
GRADE 3	**Severe**:	Marked limitation in activity, some assistance usually required; medical intervention/therapy required hospitalizations possible.

LABORATORY RANGES

Where discrepancies in the ULN and LLN of laboratory ranges occur between those included in this document and those of the laboratory that performs the assays, the values provided by the laboratory will be used for assignment of severity grade.

CLINICAL ADVERSE EVENTS

Cardiovascular	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Arrhythmia	-	Asymptomatic, transient signs, no Rx required	Recurrent/persistent; symptomatic Rx required
Hemorrhage, blood loss	Estimated blood loss ≤100 mL	Estimated blood loss >100 mL, no transfusion required	Transfusion required
QTcF (Fridericia's correction)[a] or QTcB (Bazett's correction)	Asymptomatic, QTc interval 450-479 ms, *OR* Increase in interval<30 ms above baseline	Asymptomatic, QTc interval 480-499 ms, *OR* Increase in interval 30-59 ms above baseline	Asymptomatic, QTc interval ≥500 ms, *OR* Increase in interval ≥60 ms above baseline
PR interval (prolonged)	PR interval 0.20-0.25 s	PR interval >0.25 s	Type II 2nd degree AV block *OR* Ventricular pause >3.0 s
Respiratory	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Cough	Transient-no treatment	Persistent cough	Interferes with daily activities
Bronchospasm, acute	Transient wheeze; no treatment	Requires treatment; normalizes with bronchodilator and FEV_1 <80% predicted before bronchodilator	Minimal normalization with bronchodilator andFEV_1 <80% predicted after bronchodilator
Dyspnea	Does not interfere with usual and social activities	Interferes with usual and social activities, no treatment	Prevents daily and usual social activity or requires treatment
Nasal discharge (rhinitis infective per CTCAE 4.0)	-	Localized; local intervention indicated (eg, topical antibiotic, antifungal, or antiviral)	-
Pharyngitis **(CTCAE 4.0)**	-	Localized; local intervention indicated (eg, topical antibiotic, antifungal, or antiviral)	IV antibiotic, antifungal, or antiviral intervention indicated; radiologic or operative intervention indicated

a Inclusion dependent upon protocol requirements

Respiratory (Cont.)	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Pneumonitis (rales or rhonchi) (**CTCAE 4.0**)	Asymptomatic; clinical or diagnostic observations only; intervention not indicated	Symptomatic; medical intervention indicated; limiting instrumental ADL	Severe symptoms; limiting self-care ADL; oxygen indicated
Lung infection (**CTCAE 4.0**)	-	Moderate symptoms; oral intervention indicated (eg, antibiotic, antifungal, antiviral)	IV antibiotic, antifungal, or antiviral intervention indicated; radiologic, endoscopic, or operative intervention indicated
Gastrointestinal	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Nausea	No interference with activity	Some interference with activity	Prevents daily activities
Vomiting	No interference with activity or 1-2 episodes/24 hours	Some interference with activity or >2 episodes/24 hours	Prevents daily activity or requires IV hydration
Diarrhea	2-3 loose or watery stools or <400 g/24 hours	4-5 loose or watery stools or 400-800 g/24 hours	6 or more loose or watery stools or >800 g/24 hours or requires IV hydration
Urinary Tract	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Urinary tract infection (CTCAE 4.0)	-	Localized; local intervention indicated (eg, oral or topical antibiotic, antifungal, antiviral)	IV antibiotic, antifungal, or antiviral intervention indicated; radiologic or operative intervention indicated
Reactogenicity	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Local reactions			
Pain	Does not interfere with activity	Repeated use of non-narcotic pain reliever >24 hours or interferes with activity	Any use of narcotic pain reliever or prevents daily activity
Tenderness	Discomfort only to touch	Discomfort with movement	Significant discomfort at rest
Erythema/redness [b]	2.5-5 cm	5.1-10 cm	>10 cm

b In addition to grading the measured local reaction at the greatest single diameter, the measurement should be recorded as a continuous variable.

Reactogenicity (Cont.)	Mild (Grade 1)	Moderate (Grade 2)	Severe (Grade 3)
Induration/swelling [c]	2.5-5 cm and does not interfere with activity	5.1-10 cm or interferes with activity	>10 cm or prevents daily activity
Systemic reactions			
Allergic reaction	Pruritus without rash	Localized urticaria	Generalized urticaria; angioedema or anaphylaxis
Headache	No interference with activity	Repeated use of non-narcotic pain reliever >24 hours or some interference with activity	Significant; any use of narcotic pain reliever or prevents daily activity
Fatigue	No interference with activity	Some interference with activity	Significant; prevents daily activity
Myalgia	No interference with activity	Some interference with activity	Significant; prevents daily activity
All other conditions	**Mild (Grade 1)**	**Moderate (Grade 2)**	**Severe (Grade 3)**
Illness or clinical adverse event (as defined according to applicable regulations)	No interference with activity	Some interference with activity not requiring medical intervention	Prevents daily activity and requires medical intervention

c Induration/swelling should be evaluated and graded using the functional scale as well as the actual measurement.

LABORATORY AND VITAL SIGNS TOXICITY GRADING (Some laboratory values have been modified to be consistent with the normal ranges of the ACM laboratory used in the present study)

Blood, Serum, or Plasma Chemistries [a]	LO/HI/N [b]	Mild (Grade 1) [c]	Moderate (Grade 2)	Severe (Grade 3)
Sodium (mEq/L or mmol/L)	LO	131-<LLN	130	<130
	HI	>ULN-148	149-150	>150
Potassium (mEq/L or mmol/L)	LO	<LLN-3.2	<3.2-3.1	<3.1
	HI	>ULN-5.6	>5.6-5.7	>5.7
Glucose (mg/dL)	LO mmol/L	<LLN-3.0	<3.0-2.2	<2.2
	HI mmol/L	>ULN-8.9	>8.9-13.9	>13.9
Blood urea nitrogen	HI mmol/L	>8.9-17.8	>17.8-35.5	>35.5
Creatinine	N	115-151 (µmol/L)	152-177 (µmol/L)	>177 (µmol/L)
Calcium **(CTCAE 4.0)**	LO mmol/L	<LLN-2.0	<2.0-1.75	<1.75
	HI mmol/L	>ULN-2.9	>2.9-3.1	>3.1
Magnesium **(CTCAE 4.0)**	LO mmol/L	<LLN-0.5	<0.5-0.4	<0.4
Phosphorous **(CTCAE 4.0)**	LO mmol/L	<LLN-0.8	<0.8-0.6	<0.6
Creatine kinase (CPK or CK) **(CTCAE 4.0)**	HI	>ULN-2.5xULN	>2.5xULN-5xULN	>5xULN
Albumin	LO g/L	<30-28	<28-25	<25
Total protein	LO g/L	<LLN-52	<52-50	<50
Alkaline phosphatase (U/L) **(CTCAE 4.0)**	HI	>ULN-2.5xULN	>2.5xULN-5xULN	>5xULN
AST (U/L) **(CTCAE 4.0)**	HI	>ULN-3xULN	>3xULN-5xULN	>5xULN
ALT (U/L) **(CTCAE 4.0)**	HI	>ULN-3xULN	>3xULN-5xULN	>5xULN
Bilirubin, serum total (mmol/L) **(CTCAE 4.0)**	HI mmol/L	>ULN-1.5xULN	>1.5xULN-3xULN	>3xULN
Bilirubin, serum total (mg/dL) when ALT ≥105 (Hy's law)	HI	1.3-1.5	1.6-2.0	>2.0

Blood, Serum, or Plasma Chemistries [a]	LO/HI/N [b]	Mild (Grade 1) [c]	Moderate (Grade 2)	Severe (Grade 3)
Bilirubin, serum direct (mmol/L) (CTCAE 4.0)	HI mmol/L	>ULN-1.5xULN	>1.5xULN-3xULN	>3xULN
Amylase (U/L) **(CTCAE 4.0)**	HI	>ULN-1.5xULN	>1.5xULN-2xULN	>2xULN
Lipase (U/L) **(CTCAE 4.0)**	HI	>ULN-1.5xULN	>1.5xULN-2xULN	>2xULN
Uric acid (mg/dL/mmol/L) **(CTCAE 4.0)**	HI	>ULN – 10 mg/dL (0.59 mmol/L) without physiologic consequences	-	>ULN – 10 mg/dL (0.59 mmol/L) with physiologic consequences

a Depending upon the laboratory used, references ranges, eligibility ranges and grading may be split out by sex and/or age.

b Low, High, Not Graded (N).

c If initial boundary of grade 1 has gap from reference range or eligibility range, calculations based on the New England Journal of Medicine (NEJM) reference ranges.

Hematology	LO/HI/N [a]	Mild (Grade 1) [b]	Moderate (Grade 2)	Severe (Grade 3)
Hemoglobin (women) (g/dL)	LO	10.8-11.3	9.2-10.7	<9.2
Hemoglobin (men) (g/dL)	LO	12.0-12.5	10.0-11.9	<10.0
White blood cell count (K/CUMM)	HI	11.00-15.00	15.00-20.00	>20.00
	LO	2.50-3.50	1.50-2.49	<1.50
Lymphocytes (K/CUMM)	LO	0.76-0.90	0.50-0.75	<0.5
Neutrophils (K/CUMM)	LO	1.50-1.95	1.00-1.49	<1.00
Eosinophils (K/CUMM)	HI	0.58-0.74	0.75-1.50	>1.50
Platelets (K/CUMM)	LO	120-130	100-120	<100
Coagulation				
Prothrombin time (PT, seconds)	HI	> ULN-14.4	14.5-15.7	>15.7
Partial thromboplastin time (PTT or aPTT, seconds)	HI	>ULN-42.1	42.2-50.0	>50.0
Fibrinogen (mg/dL) **(CTCAE 4.0)**	HI	>ULN-500	501-600	>600
	LO	<LLN-0.75xLLN	<0.75xLLN-0.5xLLN	<0.5xLLN
Urine				
Protein (dipstick)	HI	1+	2+	>2+
Glucose (dipstick)	HI	1+	2+	>2+
Blood (microscopic) - red blood cells per high power field (RBC/HPF)	HI	5-10 for males 9-10 for females	11-50	>50 and/or gross blood

a Low, High, Not Graded.

b If initial boundary of grade 1 has gap from reference range or eligibility range, calculations based on the New England Journal of Medicine (NEJM) reference ranges.

Vital Signs	LO/HI/N [a]	Mild (Grade 1) [b]	Moderate (Grade 2)	Severe (Grade 3)
Fever (°C) [c]	HI	38.0-38.4	38.5-38.9	>38.9
Fever (°F)	HI	100.4-101.1	101.2-102.0	>102.1
Tachycardia - beats per minute	HI	101-115	116-130	>130 or ventricular dysrhythmias
Bradycardia - beats per minute	LO	40-45	35-40	<35
Hypertension (systolic) - mm Hg [d]	HI	141-150	151-160	>160
Hypertension (diastolic) - mm Hg	HI	91-95	96-100	>100
Hypotension (systolic) - mm Hg	LO	85-89	80-84	<80
Tachypnea - breaths per minute	HI	23-25	26-30	>30

a Low, High, Not Graded.

b If initial boundary of grade 1 has gap from reference range or eligibility range, calculations based on the New England Journal of Medicine (NEJM) reference ranges.

c Oral temperature; no recent hot or cold beverages or smoking. A protocol should select either °C or °F for inclusion.

d Assuming subject is awake, resting, and supine; for adverse events, 3 measurements on the same arm with concordant results.

[redacted]

Statistical Analysis Plan

Study Title:	**A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis**
Protocol Number and Version:	**ARQ-151-312, Original dated 31 October 2020** **Amendment 1 dated 08 June 2021** **Amendment 2 dated 18 July 2021** **Amendment 3 dated 23 May 2022** **Amendment 4 dated 28 September 2022**
Product:	**ARQ-151 Cream 0.15%**
Sponsor:	**Arcutis Biotherapeutics, Inc.** **3027 Townsgate Road, Suite 300** **Westlake Village, CA 91361**
Date:	**10 November 2022**
Version:	**Final 3.0**
Prepared by:	[redacted]

STATISTICAL ANALYSIS PLAN REVISION SUMMARY			
Version	**Version Date**	**Author**	**Summary of Changes**
Original V1.0	05-JUL-2022	[redacted]	Initial version
Final V2.0	12-Oct-2022	[redacted]	Primary estimand and multiple imputation procedure were updated as per FDA's request discussed at the Type B Pre-NDA meeting on [redacted] for a similar program. A baseline definition was added for randomized subjects who never received any study treatment (if any). Analysis day, and number of days in study derivations were updated accordingly.
Final V3.0	10-Nov-2022	[redacted]	Clarify that for subjects who discontinued early from the study due to adverse event or lack of efficacy, a subject will be considered as non-responder (for MI and non-responder imputation analyses) or missing (for observed case analyses) for all analysis visits (refer to Section 5.4) for which the subject's last dose day falls within the analysis visit window or is prior to the start of the analysis visit windows.

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(S):

Name	Title	Signature	Date (DD-MMM-YYYY)
[redacted]	Principal Biostatistician [redacted]	[redacted]	

APPROVALS:

Name	Title	Signature	Date (DD-MMM-YYYY)
[redacted]	Vice President, Biometrics [redacted]	[redacted]	
[redacted]	Vice President, Biometrics Arcutis Biotherapeutics, Inc.	[redacted]	
[redacted]	Vice President, Clinical Development Arcutis Biotherapeutics, Inc.	[redacted]	

TABLE OF CONTENT

ABBREVIATIONS

AD	Atopic Dermatitis
AE	Adverse Event
ANCOVA	Analysis of Covariance
ATC	Anatomical Therapeutic Chemical
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
EASI	Eczema Area and Severity Index
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EDC	Electronic Data Capture
EM	Expectation-Maximization
ET	Early Termination
HR	Heart Rate
IP	Investigational Product
ITT	Intent to Treat
IWRS	Interactive Web Response System
LOCF	Last Observation Carried Forward
MCMC	Markov-Chain Monte-Carlo
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent to Treat
PHQ-8	Patient Health Questionnaire-8
PHQ-A	Modified PHQ-9 for Adolescents
PMM	Predictive Mean Matching
POEM	Patient-Oriented Eczema Measure
PK	Pharmacokinetic
PP	Per Protocol
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
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-151-312. The analyses described in the SAP are based upon the protocol Amendment 4 dated 28 September 2022. In case of changes (note that any such changes are described in section 3.5 below) between the protocol and the SAP, the SAP will be used to guide the statistical analysis. Any deviations from the SAP will be described and justified in the final Clinical Study Report (CSR), as appropriate.

On [redacted] FDA provided the following advice in preliminary feedback prior to a pre-NDA meeting held with Arcutis for a similar program on [redacted].

> *For the primary estimand, you [Arcutis] proposed a treatment policy strategy to handle all intercurrent events, including treatment discontinuation due to adverse event or lack of efficacy. For intercurrent events of treatment discontinuation due to adverse events or lack of efficacy, we [FDA] recommend a composite strategy policy where subjects will be defined as non-responders, as we consider this to be the appropriate approach for handling such events. Your proposal to handle intercurrent events of treatment discontinuation using the treatment policy strategy can be used as part of a supportive estimand.*

The purpose of this version of the SAP is to document the change to the primary estimand and related multiple imputation strategies as requested by the FDA for a similar program.

This SAP has been developed prior to database lock, unblinding, and associated analyses. All final analyses will be performed after approval of the SAP, the clinical trial data are entered into the database, any discrepancies in the data are resolved, determination of the inclusion/exclusion of each subject from each analysis population, the database is locked, and the unblinding request form is signed.

2 STUDY OBJECTIVES AND ENDPOINTS

OBJECTIVES	ENDPOINTS
Efficacy	
To assess the efficacy of Roflumilast (also known as ARQ-151) cream 0.15% vs vehicle administered once daily	Primary efficacy endpoint:
	• Validated Investigator Global Assessment scale for Atopic Dermatitis (vIGA-AD) Success, defined as a vIGA-AD score of 'clear' (0) or 'almost clear' (1)

OBJECTIVES	ENDPOINTS
(QD) x 4 weeks to individuals 6 years of age and older with atopic dermatitis (AD).	plus at least a 2-grade improvement from Baseline at Week 4
	Secondary efficacy endpoints:
	• In subjects with a vIGA-AD score of 'Moderate' at randomization, vIGA-AD Success at Week 4 • In subjects ≥12 years old with baseline Worst Itch - Numeric Rating Score (WI-NRS) ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 4 • In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 2 • In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction on the WI-NRS at Week 1 • Achievement of at least a 75% reduction in the Eczema Area and Severity Index (EASI-75) at Week 4 • vIGA-AD of 'clear' or 'almost clear' at Week 4 • vIGA-AD Success at Week 2 • vIGA-AD Success at Week 1 • vIGA-AD of 'clear' or 'almost clear' at Week 2 • vIGA-AD of 'clear' or 'almost clear' at Week 1
	Exploratory efficacy endpoints:
	• Other continuous efficacy endpoints include change and percent change in average weekly WI-NRS and daily WI-NRS, EASI, % body surface area (BSA) affected by AD, Dermatology Life Quality Index (CDLQI/DLQI), the Dermatitis Family Impact (DFI), Scoring Atopic Dermatitis (SCORAD), and Patient-oriented Eczema Measure (POEM) at Week 1, Week 2 and Week 4. • Other categorical efficacy endpoints include EASI-50, EASI-90, EASI-100, and vIGA-AD of 'clear' at Week 4
Safety	

OBJECTIVES	ENDPOINTS
To assess the safety of Roflumilast cream 0.15% vs vehicle administered once daily (QD) x 4 weeks to individuals 6 years of age and older with AD	• Incidence and severity of adverse events (AEs) • Changes and percent change in clinical laboratory results • Changes and percent change in vital signs • The subject incidence of >5% weight loss or gain on study • Local tolerability assessments • Patient Health Questionnaire depression scale (PHQ-8) and Modified PHQ-9 for Adolescents (PHQ-A) • Children's Depression Inventory 2nd Edition (CDI-2) • Columbia-Suicide Severity Rating Scale (C-SSRS)
Pharmacokinetic	
To assess the systemic exposure of roflumilast and its N-oxide metabolite following Roflumilast cream 0.15% QD application x 4 weeks	• Plasma concentrations of roflumilast and its N-oxide metabolite

3 STUDY DESIGN

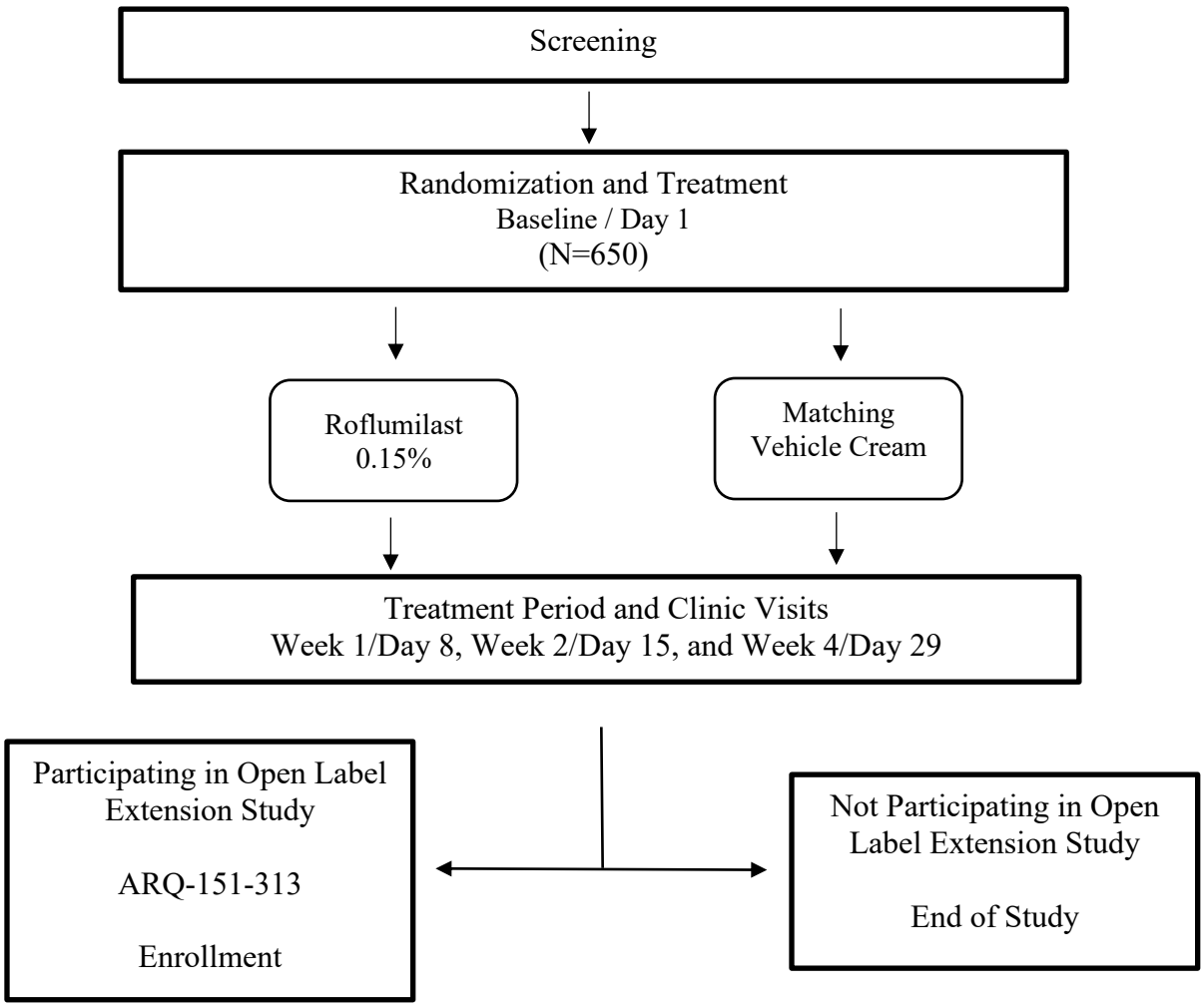
3.1 Overall Design

This is a Phase 3, parallel group, double blind, vehicle-controlled study in which Roflumilast cream 0.15% or vehicle is applied QD x 4 weeks to subjects with mild to moderate atopic dermatitis.

- Upon determination of eligibility, subjects will be randomized 2:1 to either Roflumilast cream 0.15% or matching vehicle cream. The randomization will be stratified by vIGA-AD score at baseline ('Mild' vs. 'Moderate') and by study site.
- At the Week 4 visit, subjects may be eligible to enroll in an open label extension study (ARQ-151-313) in which they will receive Roflumilast cream 0.15% QD.

The trial design is represented schematically in Figure 1.

Figure 1 **Study Schema**



A Phase 3, 4-Week, Parallel Group, Double Blind, Vehicle-Controlled Study of the Safety and Efficacy of ARQ-151 Cream 0.15% Administered QD in Subjects with Atopic Dermatitis

Approximately 650 subjects with atopic dermatitis will be randomized 2:1 to receive either:

- Roflumilast cream 0.15% or Vehicle cream

Subjects will have ≥3% BSA involvement (excluding the scalp, palms, soles) with a vIGA-AD score of '2' (mild) or '3' (moderate) for study entry

Up to 50% of the subjects will be ≥ 18 years old

3.2 Schedule of Events

Table 1 Schedule of Visits and Assessments

Study Procedure	Screen	Baseline Day 1	Wk 1 Day 8	Wk 2 Day 15	Wk 4 Day 29 / ET
Visit	**1**	**2**	**3**	**4**	**5**
Visit Window	**-30 days**	**N/A**	**+/- 3 days**	**+/- 3 days**	**+/- 3 days**
Informed consent/assent	X				
Demographics	X				
Medical and surgical history	X				
Physical examination[a]	X	X			X
I/E criteria	X	X			
Hematology, Serum Chemistries, and Urine Analysis[b]	X[b]	X[b]			X[b]
Vital signs, height, weight[c]	X	X	X	X	X
vIGA-AD, EASI, BSA, SCORAD[d]	X	X	X	X	X
WI-NRS pruritus[e]	X	X	X	X	X
POEM[f]	X	X	X	X	X
Local Tolerability Assessment[g]		X	X	X	X
CDI-2, PHQ-8, PHQ-A, C-SSRS[h]	X	X	X	X	X
DLQI, CDLQI, DFI[i]	X	X	X	X	X
Medical Photography[j]		X	X		X
Serum pregnancy test (FOCBP only)	X				
Urine pregnancy test[k]		X	X	X	X
PK draws[l]					X
Drug/vehicle application in clinic[m]		X	X	X	
Dispense/Re-dispense study medication kit[n]		X	X[o]	X[o]	X[o]
Dispense/review diary	X	X	X	X	X
Weigh study medication kit[p]		X	X	X	X
Compliance determination[q]			X	X	X

Study Procedure	Screen	Baseline Day 1	Wk 1 Day 8	Wk 2 Day 15	Wk 4 Day 29 / ET
Adverse event assessment[r]	X	X	X	X	X
Concomitant medications	X	X	X	X	X
Study Exit [s]					X

a Limited physical examination: skin (including assessment of Fitzpatrick skin type at Screening only), lungs, and heart only

[b] For all subjects entering this study under Amendment 2, to be collected at Screening, but subsequent samples will be collected only for subjects ≥12 years old (Baseline/Day 1 and Week 4/Day 29/ET). For subjects 12 to 18 years of age, if Baseline/Day 1 is within 3 weeks of Screening, the Screening results may be used.

[c] Height will be collected at Screening only. 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 other 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. **The vIGA-AD assessment should be completed prior to other physician assessments**. SCORAD total score will range between 0 and 103.

[e] Subjects will self-assess their pruritus at home on a daily basis starting 7 days prior to the Baseline/Day 1 visit, and then every day thereafter. WI-NRS score will be determined by the subject assessing worst itch over the past 24 hours. The scale is from 0 (no itch) to 10 (worst itch) and this value will be recorded by the subject each day. Subjects will be trained at the Screening visit in the accurate completion of the WI-NRS. In addition, parents/caregivers of children and adolescent subjects will be trained at the Screening visit by study staff on how to assist the subject, if needed, in completing the WI-NRS.

[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] Local tolerability assessments should be recorded prior to study drug application for the Investigator assessment of skin irritation (Berger and Bowman skin irritation score). **Note for investigator tolerability assessments: reactions at the site of product application, which may occur post-Baseline, should be differentiated from the preexisting inflammation associated with the subject's atopic dermatitis.** The subject will assess burning/stinging (0-3 score) 10-15 minutes post drug application. **Note subject burning stinging assessment: at Day 29, subjects will provide a recall assessment of burning/stinging experienced post drug application on the previous day (Day 28).**

[h] Adolescents and adults will complete the C-SSRS (12 years of age and older). Adults will complete the PHQ-8. Adolescents (ages 12 to 17, inclusive) will complete the PHQ-A (PHQ-9 modified). Parents/caregivers will complete CDI-2 (parent report) for children 6-11 years of age, inclusive.

[i] The DLQI will be completed by subjects ≥17 years of age. The CDLQI will be completed for subjects 6 to 16 years old, inclusive. The Dermatitis Family Impact Questionnaire (DFI) will be completed by parents/caregivers for all subjects 6 to ≤17 years of age.

[j] Photography of AD lesion(s) selected by the Investigator will be performed at all investigational sites. All efforts will be made to de-identify the subjects. Canfield equipment will be used to capture photographs. Subjects who are unwilling to participate in the medical photography will be allowed to opt out of this procedure, as documented on the Informed Consent Form.

[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 entering this study under Amendment 2, a single PK trough draw will be collected at Day 29 only for subjects ≥12 years old. Ensure study medication was not applied in the area where PK will be drawn.

[m] Subjects to apply assigned IP during clinic visits, except for the Day 29/ET visit.

[n] It is expected that kits will be dispensed based on %BSA affected. See IP Handling Manual for details.

[o] On Day 8 and 15, dispensing of IP is optional. Site should review IP kit to ensure sufficient IP is available until the next visit and only dispense additional IP if needed. On Day 29, if the subject is unable to perform the Day 29 clinic visit due to COVID-19 restrictions (isolation, quarantine, etc.) then additional IP may need to be dispensed so IP can continue to be applied at home until the subject is able to return to the clinic to complete the Day 29 assessments (see IP Handling Manual for the process to dispense additional IP at or after Day 29).

[p] Every tube should be weighed and recorded when dispensed and returned. See IP Handling Manual for details.

[q] Compliance determination is described in the IP Handling Manual

[r] All AEs should be collected starting after the first application of the investigational product through the end of the study. All SAEs should be collected starting after the signing of the informed consent through 30 days after the last day of the application of the investigational product or the end of the study (whichever is later). Any AEs (whether serious or non-serious) and clinically abnormal laboratory test values(s) will be evaluated by the PI and treated and/or followed up for up to 30 days after end of treatment or until symptoms or value(s) return to normal, or acceptable level, as judged by the PI (if the subject is continuing into the ARQ-151-313 OLE study, then AEs from this study (ARQ-151-312) will only be followed until exit from this study).

[s] Subjects who enroll into the open label extension study (ARQ-151-313) must complete the ARQ-151-312 visit requirements at Week 4.

3.3 Treatment

Roflumilast cream 0.15% or vehicle cream will be administered QD for 28 days (+/- 3 days).

- Roflumilast cream 0.15%
- Vehicle cream

3.4 Randomization, Replacement, and Unblinding Procedures

Randomization will take place at the Baseline visit prior to first dosing. Subjects who meet all eligibility criteria will be randomized at a 2:1 ratio (drug:vehicle) to receive Roflumilast cream 0.15% QD or matching vehicle QD. The randomization will be stratified by vIGA-AD score at baseline ('Mild' vs. 'Moderate') and by study site according to a computer-generated randomization list. Kits containing tubes of study medication will be assigned to each subject using an internet-based response system (IWRS). A subject may receive more than one kit for the treatment period. The kits and tubes are blinded and each kit is numbered with a unique kit number.

The study is double-blinded, therefore neither the subjects nor the Investigator, sponsor and clinical personnel will be aware of which treatment an individual subject receives.

3.5 Changes to the Analysis from the Protocol

Rationale for change	Description of the change
A categorical summary 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.
To separate COVID patient from Per Protocol analysis set.	Original text: Per protocol (PP) population will include all subjects in the ITT population, who are at least 80% compliant with study medication application, and show no major deviations from the study protocol that would affect the interpretation of efficacy. In addition, subjects who miss the Week 4 vIGA-AD assessment specifically due to novel coronavirus disease-19 (COVID-19) disruptions will be excluded from per protocol population. Changed to: Per protocol population will include all subjects who are randomized, at least 80% compliant with study medication application, have a vIGA-AD assessment within the Week 4 visit window, and show no "major deviations" from the study protocol that would affect the interpretation of efficacy.
Added mITT population to address COVID-19 impact.	The mITT population includes all randomized subjects with the exception of subjects who missed the week 4 vIGA-AD assessment specifically due to COVID-19 disruption. This population will be used for sensitivity analysis for the primary endpoint.

4 POPULATIONS FOR ANALYSIS

4.1 Intent-to-Treat (ITT) Population

Intent-to-Treat population will include all subjects who are randomized, and all subjects will be analyzed according to the treatment group and stratum to which they were randomized.

4.2 Modified Intent-to-Treat (mITT) Population

The mITT population includes all randomized subjects with the exception of subjects who missed the week 4 vIGA-AD assessment specifically due to COVID-19 disruption. This population will be used for sensitivity analysis for the primary endpoint. All subjects will be analyzed according to the treatment group and stratum to which they were randomized.

4.3 Per Protocol (PP) Population

Per protocol population will include all subjects who are randomized, at least 80% compliant with study medication application, have a vIGA-AD assessment within the Week 4 visit window and show no "major deviations" from the study protocol that would affect the interpretation of efficacy. A complete list of major deviations from the study protocol will be created prior to unblinding and include a list of all subjects who will be excluded due to those major deviations. See section 7.2 for more details.

All subjects will be analyzed according to the actual treatment group they received and the stratum they belong to. Actual and randomized treatment will only differ if the subject received the wrong treatment throughout their participation in the study.

4.4 vIGA-AD Moderate ITT Population

vIGA-AD Moderate ITT population will be a subset of the ITT population with vIGA-AD score (randomized score) 'Moderate' at randomization.

All subjects will be analyzed according to the treatment group and the stratum to which they were randomized.

4.5 vIGA-AD Moderate PP Population

vIGA-AD Moderate PP population will be a subset of the PP population with vIGA-AD score (actual score) 'Moderate' at randomization.

All subjects will be analyzed according to the actual treatment group they received and the stratum they belong to. Actual and randomized treatment will only differ if the subject received the wrong treatment throughout their participation in the study.

4.6 WI-NRS Population

WI-NRS population will be a subset of the ITT population among subjects ≥12 years old with average weekly baseline WI-NRS score ≥ 4.

The subjects in WI-NRS population are those:

1. Completed at least 4 of 7 evaluable daily WI-NRS questionnaires during the last 7 days of the Screening period;
2. Have a mean baseline WI-NRS score ≥ 4.0, defined as the average of all non-missing scores reported during the last 7 days of the Screening period if at least 4 of 7 evaluable daily WINRS questionnaires available. If 4 or more evaluable daily questionnaires are missing, then the data will be treated as missing.

All subjects will be analyzed according to the treatment group and stratum to which they were randomized.

4.7 Safety Population

Safety population will include all subjects who are enrolled and received at least one confirmed dose of study medication.

Subjects will be analyzed based on the treatment group received and the stratum they belong to. Actual and randomized treatment will only differ if the subject received the wrong treatment throughout their participation in the study.

4.8 Pharmacokinetic (PK) Population

Pharmacokinetic population will include all subjects receiving the active drug with quantifiable plasma concentrations of roflumilast.

5 GENERAL CONSIDERATIONS

Formats and layouts of tables, listings, and figures (TLF) will be provided in a separate document (output general layout is described in Appendices

Appendix **1**).

5.1 Sample Size

There are approximately 650 subjects planned for this study. In order to have the desired power for the secondary endpoint of vIGA-AD Success in subjects with a vIGA-AD score of 'Moderate' at randomization, approximately 490 of the subjects to be accrued will have vIGA-AD score of 'Moderate' at randomization. Randomization will be stratified by vIGA-AD score ('Mild' vs. 'Moderate') and by study site.

The sample size of 650 subjects provides approximately 95% power to detect an overall 15% difference (Odds Ratio = 2.1) between treatment groups on vIGA-AD success at Week 4 at α=0.05 using a 2-sided stratified Cochran-Mantel-Haenszel (CMH) test. The results from a recent Phase 2 study (ARQ-151-212) of Roflumilast cream 0.15% compared to vehicle treatment were used to estimate the treatment difference. Specifically, in the Phase 2 trial, approximately 37% of subjects demonstrated vIGA-AD Success at Week 4 in the ARQ-151 0.15% group compared to 22% in the vehicle group. The sample size of 490 also provides approximately 90% power to detect an overall 17% difference (Odds Ratio = 2.1) between treatment groups (28% of vIGA-AD Success at Week 4 in vehicle treatment) on vIGA-AD Success at Week 4 among subjects with vIGA-AD score 'moderate' at randomization. The same testing method, the stratified CMH test, will be used as for the primary endpoint.

To control for familywise type I error at level of 0.05, the secondary endpoint of vIGA-AD Success at Week 4 in subjects with vIGA-AD of 'Moderate' at randomization will only be tested if the primary endpoint demonstrates statistical significance. In addition, the remaining secondary endpoints will be inferentially tested only if the primary and secondary endpoint (vIGA-AD Success at Week 4 for subjects with vIGA-AD score of 'Moderate' at randomization) comparisons are statistically significant using the hierarchical testing procedure by partitioning of alpha (see Section 6.5 for more details).

5.2 Baseline

Unless otherwise specified, baseline value will be defined as the last non-missing assessment prior to or concurrently with the first study treatment dosing* (including unscheduled/retest assessments). If the last non-missing assessment is performed on the same date as the first study treatment administration* and time is not available, it is assumed that the assessment took place prior to IP application*, per study site training, and the assessment will be considered as baseline, except for adverse events (AEs) and medications starting on the first study treatment dose administration date which will be considered postbaseline.

Average weekly baseline WI-NRS is defined as the average of all non-missing scores reported during the last 7 days prior to treatment* if at least 4 of 7 evaluable daily WI-NRS questionnaires

available. Daily baseline WI-NRS is defined as the last non-missing assessment prior to or concurrently with the first study treatment dosing*. Day 1 WI-NRS score will be used to calculate the value for week 1 only when it is collected after the application of the first study drug*. If Day 1 WI-NRS score is collected prior to or concurrently with the application of the first study drug*, then the Day 1 WI-NRS score will be included in baseline calculation.

For investigator/subject tolerability assessments, baseline is derived as the measurement taken on the day of first application of study drug*.

* For randomized subjects who discontinued study before the first application of study treatment, the date of randomization will be considered instead of the date of the first application of study treatment.

5.3 Reference Start Date and Analysis Day

Analysis day will be calculated from the first study treatment administration date* and will be used to derive start/end day of assessments or events.

Analysis day = (Date of event – Date of first dose administration*) + 1 if date of event is on or after the date of first dose administration of study treatment*;

= (Date of event – Date of first dose administration*) if date of event is before the date of first dose administration of study treatment*.

In the situation where the assessment/event date is partial or missing, analysis day will be missing.

* For randomized subjects who discontinued study before the first application of study treatment, the date of randomization will be considered instead of the date of the first application of study treatment.

5.4 Windowing Conventions

Visits will be analysed as scheduled. Unscheduled, early termination visits, and/or retest measurements will only be included if a scheduled measurement is not available and the early termination or unscheduled/retest measurement falls within the analysis visit windows as described in Table 2, Table 3, and Table 4 when appropriate. Unscheduled/retest measurements will be listed.

If there is more than one assessment for a given timepoint and analysis visit when a scheduled measurement is not available, the assessment closest to the target day will be considered.

Table 2 Analysis Visit Windows for Efficacy Endpoints, Vital Signs, CDI-2, PHQ-8, PHQ-A, C-SSRS and Local Tolerability Assessment

Analysis Visit	Target Day	Lower Limit	Upper Limit
Week 1	8	2	11
Week 2	15	12	22
Week 4	29	23	43

Table 3 Analysis Visit Windows for Clinical Laboratory, Physical Examination and PK assessment

Analysis Visit	Target Day	Lower Limit	Upper Limit
Week 4	29	23	43

Table 4 Windows for the derivation of Average Weekly WI-NRS

Days for calculation of weekly average	Week (Derived)
(-7, -1)*	Baseline
(1, 7)*	Week 1
(8, 14)	Week 2
(15, 21)	Week 3
(22, 28)	Week 4

* Day 1 WI-NRS score will be used to calculate the value for week 1 only when it is collected after the application of the first study drug (randomization for randomized subjects who were never treated with study drug). If Day 1 WI-NRS score is collected prior to or concurrently with the application of the first study drug (randomization for randomized subjects who were never treated with study drug), then the Day 1 WI-NRS score will be included in baseline calculation.

Note: With the caveat described in footnote '*', if more than one WI-NRS score is available on the same day, the worst score of the day will be considered in the analyses.

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 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 MI and non-responder imputation analyses) or missing (for observed case analyses) for all pre-specified analysis visits (refer to Section 5.4) for which the subject's last dose day falls within analysis visit window or is prior to the start of the analysis visit window.

- Average weekly WI-NRS = Average weekly WI-NRS pruritus score will be calculated as the sum of the daily WI-NRS scores reported during a specific week (in a 7-day period; refer to Section 5.4) of the study divided by the number of days with non-missing scores for that week. A minimum of 4 days of observations are needed to calculate an average weekly WI-NRS pruritus score. Otherwise, the corresponding average weekly WI-NRS pruritus score will be considered missing.

- With the following exception, a WI-NRS 4-point reduction = achievement of a 4- point reduction in average weekly WI-NRS pruritus score compared to average weekly WI-NRS baseline, calculated only for the subjects ≥12 years old with average weekly WI-NRS score of ≥ 4 at 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 WI-NRS 4-point reduction (for MI and non-responder imputation analyses) or missing (for observed case analyses) for all pre-specified analysis visits (refer to Section 5.4) for which the subject's last dose day falls within the analysis visit window or is prior to the start of the analysis visit window.

- EASI total score = $0.1\ (E_h + I_h + E_{Xh} + L_h)\ A_h + 0.2\ (E_u + I_u + E_{Xu} + L_u)\ A_u + 0.3\ (E_t + I_t + E_{Xt} + L_t)\ A_t + 0.4\ (E_l + I_l + E_{Xl} + L_l)\ A_l$ for subjects '≥ 8 years old and $0.2\ (E_h + I_h + E_{Xh} + L_h)\ A_h + 0.2\ (E_u + I_u + E_{Xu} + L_u)\ A_u + 0.3\ (E_t + I_t + E_{Xt} + L_t)\ A_t + 0.3\ (E_l + I_l + E_{Xl} + L_l)\ A_l$ for subjects < 8 years old

  where E, I, Ex, 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. Scalp, palms, and soles may be treated with investigational product in this study but will be excluded from the EASI assessment. If a subject turns 8 years old during the study, the formula used at Screening will continue to be used through the study duration of the subject's participation in the study.

- With the following exception, EASI-50, EASI-75, EASI-90, and EASI-100 = Achievement of at least a 50%, 75%, 90%, or 100% reduction from baseline in EASI total score, respectively. 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 reached EASI-50, EASI-75, EASI-90 and EASI-100 (for MI and non-responder imputation analyses) or missing (for observed case analyses) for all pre-specified analysis visits (refer to Section 5.4) for which the subject's last dose day falls within the analysis visit window or is prior to the start of the analysis visit window.

- DLQI Score = sum of the 10 questions (individual questions scored as Very much=3, A lot=2, A little=1, Not at all=0, Not relevant=0, Question 7: Yes=3, 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.

- CDLQI Score = sum of the 10 questions (individual questions scored as Very much=3, Quite a lot=2, Only a little=1, Not at all=0; Question 7: if the last week was school time, the question was scored as Very much=3, Quite a lot=2, Only a little=1, Not at all=0, with Prevented school recoded to 3, and if the last week was 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.

- DFI 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.

- PHQ-8 = sum of the 8 questions (individual questions scored as Not at all=0, Several days=1, More than half the days=2, and Nearly every day=3), 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 (=7).

- Modified PHQ-A = sum of the 8 questions (individual questions scored as Not at all=0, Several days=1, More than half the days=2, and Nearly every day=3), 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 (=7).

- CDI-2 total score is a sum of the 17 questions based on the scoring grid (individual questions scored as much or most of the time=0, often=1, some of the time=2, Not at all=3 for questions Q2, Q7, Q13, Q14, and Q16; individual questions scored as much or most of the time=3, often=2, some of the time=1, Not at all=0 for all other questions), ranging from

0 to 51. All CDI-2 raw scores can be prorated using the following formula, rounding to the nearest whole number:

- Prorated score = (Obtained raw score for scale) * (Total # of items on scale/subscale) / Total # items on scale/subscale with responses
- CDI-2 emotional problem scale is a sum of 9 questions (Q1, Q3-6, Q8, Q10-12). 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*9)/number of answered items (=8).
- CDI-2 functional problem scale is a sum of 8 questions (Q2, Q7, Q9, Q13-17). 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 (=7).
- CDI-2 total score is a sum of 17 questions. If more than 2 items are missing the score should not be calculated. If 1 item is missing the score is calculated as (sum of answered items*17)/number of answered items (=16). If 2 items are missing the score is calculated as (sum of answered items*17)/number of answered items (=15).

- POEM = 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 is scored 0 and the scores are summarized and expressed as usual out of a maximum of 28. If 2 or more questions are left unanswered the questionnaire is not scored.
- SCORAD = [Overall BSA affected by AD / 5] + [Intensity score*7/2] + subjective symptoms score (pruritus + sleep loss); SCORAD score will be set to missing if information for any of the three measures is missing.

5.6 Descriptive Statistics

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

Categorical variables will be presented as frequencies and percentages.

Summary tables will be presented by visit, when applicable.

Change from baseline will be calculated as:

Assessment value at postbaseline visit X – baseline value.

Percent change from baseline will be calculated as:

(Assessment value at postbaseline visit X – baseline value) × 100% / (baseline value)

Percent change from baseline will be missing in situation where baseline value equals to 0.

5.7 Statistical Tests

Unless otherwise specified, all statistical tests will be two-sided and will be performed with a significant level of 0.05. Confidence intervals (CIs) will be two-sided with 95% coverage.

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.

All data from retest, unscheduled measurements and ET visit assessments will be listed.

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

Covariates for this study include pooled study site and baseline vIGA-AD (vIGA-AD=2 - Mild vs. vIGA-AD=3 - Moderate at randomization). Subgroup analyses will be generated for the baseline covariates.

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

All subjects, regardless of completion status, will have available data assigned to the pre-specified analysis visit using the analysis windows defined in Section 5.4, including the last available data

of subjects who prematurely withdraws from the study. To comply with the definition of the primary estimand (Section 12.1.1), for subjects who discontinue due to lack of efficacy or adverse event, 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. Similarly, WI-NRS weekly averages will be removed from the source data used for multiple imputation if subject's last dose day falls within the analysis visit window or is prior to the start of the interval used to compute the weekly average as defined in Section 5.4. This procedure will ensure that the data collected on or after intercurrent events are not used in the imputation process.

For the primary efficacy endpoint of vIGA-AD success at Week 4 and the secondary endpoint of vIGA-AD success at Week 4 among subjects with a 'Moderate' randomized vIGA-AD score, the primary analysis will impute missing values using a Predictive Mean Matching (PMM) sequential-regression multiple imputation model for the ITT population. This is a three-step process.

1. The first step is to understand the pattern of missingness. In order to perform the multiple imputation, a monotone missing pattern has to be achieved. For example, if there exist values for baseline and Week 4 visits, but missing values for the Week 1 or 2 visits, the Markov-Chain Monte-Carlo (MCMC) method will be used to impute the small amount of missing data that may be missing at the intermediate visits that is required to make the missing data pattern monotone before applying the multiple imputation algorithm. This method uses a non-informative Jeffreys prior to derive the posterior mode from the expectation-maximization (EM) algorithm as the starting values for the MCMC method. The MCMC method will use the seed 6457149. The vIGA-AD score will be treated as a continuous variable for this step and the model will include the vIGA-AD scores at baseline, Week 1, Week 2, and Week 4. 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:

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

   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 treatment group, pooled study site, and vIGA-AD score outcomes at previous visits, using a seed of 482371. For other scales/questionnaires, the actual baseline vIGA score will also be included as an independent variable. 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 of the 25 to 250 completed dataset, the necessary derived variable will be computed as defined in Section 5.5 and analyzed using a CMH analysis, adjusted for the pooled study site and randomized vIGA-AD score for the primary efficacy endpoint and adjusted for the pooled site for the secondary efficacy endpoint. Results will be combined into one multiple imputation inference as follows:
    a. Common proportion of success, common Mantel-Haenszel (MH) proportion difference (and associated 95% CI), and common MH odds ratio (and associated common 95% CI) will be combined using PROC MIANALYZE based on Rubin's rule. Common MH odds ratios and associated common 95% CI will first be normalized using a log-transformation before being combined using PROC MIANALYZE. The resulting combined common MH odds ratio and associated combined common 95% CI will be back-transformed to the arithmetic scale before being presented in a table.
    b. For the combined common proportion of success, the associated combined common 95% CI will be calculated as per Lott and Reiter multiple imputation Wilson interval method[1].
    c. Two p-values will be produced for each analysis:
        i. The primary p-value will be obtained from a multiple imputation CMH test, where CMH general association statistics and their standard errors obtained

from the analysis of each multiply-imputed dataset will be transformed as per Wilson-Hilferty[2] before being combined using PROC MIANALYZE based on Rubin's rule. Because the Wilson-Hilferty transformation is a monotone transformation, the p-value for the CMH test is the one-sided p-value from the t distribution. This p-value will be the primary p-value used to evaluate the result according to the multiple testing strategy described in Section 6.5.

i. Should the common MH odds ratio from the analysis of at least one of the multiply-imputed datasets be not estimable, combined common MH odds ratio and associated combined common 95% CI and p-value will not be presented. Under such circumstance, conclusions will be based on the p-value obtained from a multiple imputation test of the proportion difference, where the common proportion difference using MH weights and associated common standard errors based on the Sato variance estimator obtained from the analysis of each multiply-imputed dataset will be combined using PROC MIANALYZE based on Rubin's rule.

Similar multiple imputation method will be used for the average weekly WI-NRS and EASI total score.

For the average weekly WI-NRS (refer to Section 5.5), it's the missing average weekly WI-NRS data (i.e., those that cannot be computed because only 3 or less non-missing assessments are available during a given week) at Weeks 1, 2, 3, and 4 that will be multiply imputed, not the missing daily WI-NRS data. Since the missing average weekly WI-NRS values have a precision of 1 decimal place, the MCMC imputation for the non-monotone missing data will be restricted to values between 0 and 10, rounded to the 1st decimal (i.e., 0.1). Imputation of missing data for WI-NRS will be based on the ITT population, i.e., all randomized subjects will be included in the imputation process.

For the EASI total score (refer to Section 5.5), it's the missing EASI total score data at Weeks 1, 2, and 4 that will be imputed, not the missing EASI question score data. Since the missing EASI total scores have a precision of 1 decimal place, the MCMC imputation for the non-monotone missing data will be restricted to values between 0 and 72, rounded to the 1st decimal (i.e., 0.1). Imputation of missing data for EASI total score will be based on the ITT population.

6.2.2 Non-responder Imputation Analysis

If assessment of vIGA-AD after baseline is missing, the subject will be considered as non-responder (for example, no vIGA-AD success in the analysis of vIGA-AD success). For the vIGA-

AD success endpoints, subjects with missing baseline vIGA-AD score will be considered as non-responder for these endpoints only. That is, for the secondary efficacy endpoints of vIGA-AD score of clear (0) or almost clear (1) and vIGA-AD score of clear (0), the post-baseline assessment at a specific visit will be established based on the vIGA-AD score at that visit only, regardless of the availability of the baseline vIGA-AD score.

Similar imputation method will be used for 4-point reduction on the average weekly WI-NRS and EASI-75. For EASI-75, subjects with missing baseline EASI total score will be considered as non-responder. For 4-point reduction on the average weekly WI-NRS, subjects with missing baseline average weekly WI-NRS will not be imputed since the analysis of WI-NRS endpoints will be performed based on the WI-NRS population from which subjects with a missing baseline average weekly WI-NRS are excluded.

6.2.3 Tipping Point Analysis

As a sensitivity analysis to the multiple imputation analysis as described in Section 6.2.1 for the vIGA-AD success primary endpoint and secondary endpoint of vIGA-AD success among subjects with Moderate disease at baseline, a tipping point analysis will be performed in order to determine the inflection point at which the inference under the missing not at random (MNAR) assumption changes substantially.

The sensitivity analysis for the primary endpoint will be performed by using a specified sequence of shift parameters. The range of shift parameters to be included in this analysis are 0 to 2 by 0.2 for active and -2 to 0 by 0.2 for Vehicle. The values at which the results of the primary analysis are shifted from significant (i.e., $\alpha \leq 0.05$) to non-significant (i.e., $\alpha > 0.05$) will be determined. Steps 1 and 3 of the analysis will be the same as for the multiple imputation analysis as described in Section 6.2.1. However, Step 2 of the analysis is where the shift parameters will be applied.

Imputed values for subjects who discontinue due to lack of efficacy or adverse event will be handled as described in Step 2 of Section 6.2.1 to ensure that these subjects are analyzed as non-responders at all visits on or after discontinuation of treatment.

6.3 Interim Analysis

No interim analysis is planned for this study.

6.4 Multicenter Studies

The study will be conducted at approximately 60 study sites in the US, Canada, and Poland. During the conduct of the study, additional countries and/or sites may be added if necessary.

Sites with less than 16 subjects in the ITT population will be pooled within a country:

a. Sites with less than 16 subjects will be ordered from lowest to highest in terms of number of ITT subjects. In case of ties, the ordering for tied sites will be determined according to the site identification number (from smallest to largest).
b. Sites will be combined beginning at the smallest until the resulting pooled site contains at least 16 ITT subjects with at least 1 subject in each treatment group. The sites pooled in this way will be considered as a single site in the statistical analyses.
c. The process described above will resume for the remaining sites not meeting the criterion of at least 16 ITT subjects with at least 1 subject in each treatment group. If the final set of pooled sites does not meet the criterion of at least 16 ITT subjects with at least 1 subject in each treatment group, the final set will be pooled with the preceding pooled site.
d. If there is only one site with less than 16 ITT subjects with at least 1 subject in each treatment group, then this site will be combined with the site with the second lowest number of subjects. As above, in the case of ties, the ordering for tied sites will be determined according to the site identification number (from smallest to largest).

As a sensitivity analysis of different pooling strategy, sites that have randomized less than 50% of the number of randomized subjects at the site with the largest number of randomized subjects, those sites will be pooled within each country. The sensitivity analysis of pooling strategy will be applied for the primary endpoint only.

6.5 Multiple Comparisons/Multiplicity

To control for familywise type I error at level of 0.05, the secondary endpoint of vIGA-AD Success at Week 4 for subjects with vIGA-AD score of 'Moderate' at randomization will only be tested if the primary endpoint demonstrates statistical significance. In addition, the remaining secondary endpoints will be inferentially tested only if the primary and the secondary endpoint (vIGA-AD Success at Week 4 for subjects with vIGA-AD score of 'Moderate' at randomization) comparisons are statistically significant using hierarchical testing procedure by partitioning of alpha.

Upon successful demonstration of statistical significance for the primary and above secondary endpoint, the remaining endpoints will be grouped into secondary endpoint family 1, comprised of the 4-point reduction on the WI-NRS endpoint, at Week 4, Week 2 and Week 1, and secondary endpoint family 2, comprised of the endpoints of EASI-75 at Week 4, vIGA-AD of 'clear' or 'almost clear' at Week 4, vIGA-AD of success at Week 2 and Week 1, vIGA-AD of 'clear' or 'almost clear' at Week 2 and Week 1. An alpha level of 0.03 will be used to test the endpoints in the secondary endpoint family 1 sequentially. An alpha level of 0.02 will be used to test the endpoints in secondary endpoint family 2 sequentially.

In addition to the partitioning of the overall 0.05 alpha into two families, the Fallback Method will be applied. The fallback method is a modification of the fixed-sequence method, providing opportunity to test an endpoint later in the sequence even if an endpoint tested early in the sequence has failed to show statistical significance. The order of the endpoints remains important. The appeal of the fallback method is that if an endpoint later in the sequence has a robust treatment effect while the preceding endpoint is unsuccessful, there is a modest amount of alpha retained as a fallback to allow interpretation of that endpoint without inflating the Type I error rate. Applying the fallback method begins by dividing the total alpha (not necessarily equally) among the endpoints and maintains a fixed sequence for the testing. In this study, the Fallback Method will be applied to the fixed sequence of testing Family 1, and then Family 2.

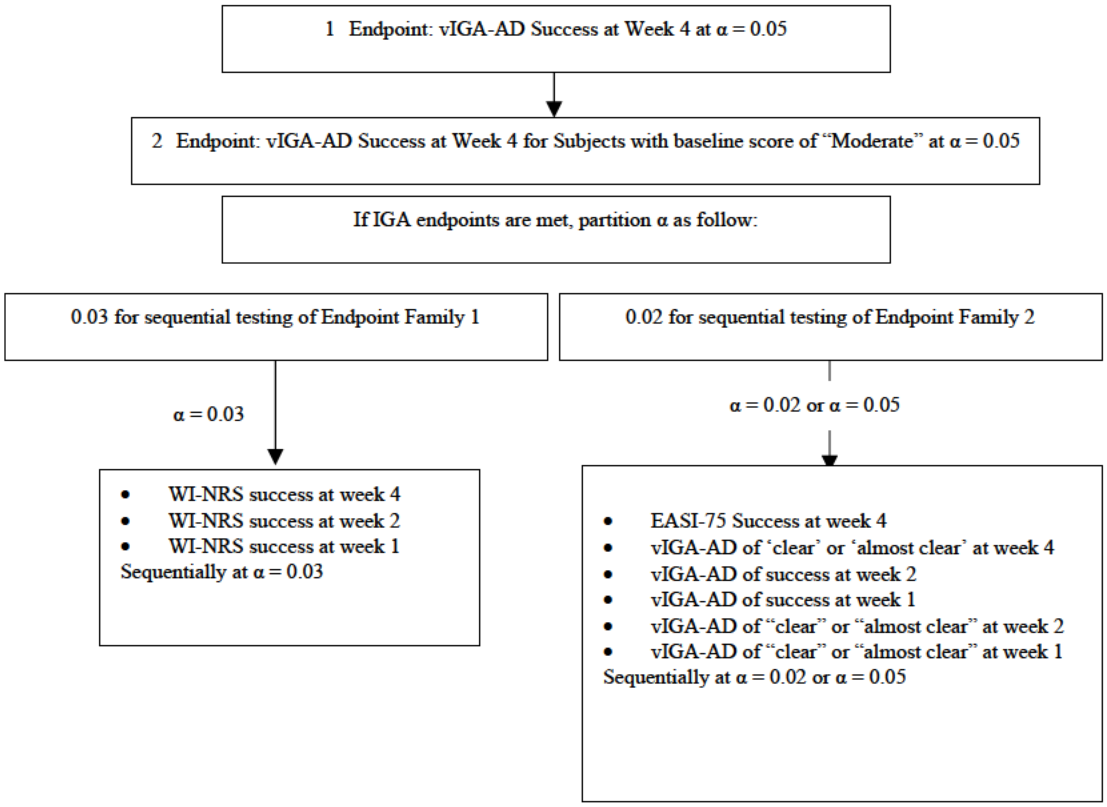
As the testing sequence progresses, a successful test preserves its assigned alpha as "saved" ("unused" or "accumulated") alpha that is passed along to the next test in the sequence, as is the case for the sequential method. This accumulated alpha is added to the prospectively assigned alpha (if any) of that next endpoint and the summed alpha is used for testing that endpoint. Thus, as sequential tests are successful, the alpha accumulates for the endpoints later in the sequence; these endpoints are then tested with progressively larger alphas.

In this study, the Fallback Method will be applied following this sequence:

Family 1: Testing will proceed at the 0.03 level sequentially within Family 1. Should all 3 endpoints in Family 1 be statistically significant at the 0.03 level, then the full 0.03 alpha will be carried to Family 2. Family 2 would then be tested at the full (α=0.02+0.03=0.05).

Should, anywhere during the sequential testing of Family 1, there is a p-value >0.03, the testing within Family 1 will stop, and no additional alpha can be carried over to Family 2.

Figure 2 Multiple Testing Scheme



Achievement of vIGA-AD success is a score of "clear" or "almost clear" plus a 2-grade improvement from baseline.

WI-WRS Success is a 4-point reduction in WI-NRS among subjects ≥12 years old with WI-NRS ≥ 4 at baseline.

EASI-75: achievement of at least a 75% reduction in the Eczema Area and Severity Index

6.6 Examination of Subgroups

Subset analysis for the following subgroups will be performed for the primary and secondary efficacy endpoints:

- Age group (6 – 11 years vs. 12 - 17 years vs. ≥18 years),
- Gender (male vs. female),
- Race (White vs. Black or African American vs. Asian vs. other),
- Ethnicity (Hispanic vs. Non-Hispanic),
- Randomized vIGA-AD score (mild (2) vs. moderate (3)),

- Actual baseline vIGA-AD score (mild (2) vs. moderate (3)),
- Baseline % BSA (<10% vs. ≥ 10%),
- %BSA categories based on tertiles,
- Baseline EASI total score (≤7 vs >7),
- Baseline EASI total score based on tertiles
- Fitzpatrick skin type at Screening (Type I, II and III vs. Type IV, V, and VI),
- Prior inadequate response, intolerance, or contraindication to Topical Corticosteroids (yes vs. no),
- Facial Involvement (yes vs. no)

For subgroup based on tertiles, tertiles will be derived using pooled data from both treatment group based on the ITT population.

Details on these analyses are described in Section 12.4.

7 STUDY SUBJECTS

7.1 Disposition of Subjects

All subjects who provide informed consent will be accounted for in this study. The number of subjects who were screened and who failed screening (screen failures) will be presented. The reasons for screen failure will be presented for all screened subjects who failed screening.

The number of subjects randomized will be presented by treatment group. The number and percentage of the subjects included in each analysis population will be provided by treatment group. The number and percentage of the subjects who completed the study, who discontinued the study, the reasons for study discontinuation, and early termination due to COVID-19 disruption will be presented by treatment group. The percentages will be calculated using the number of the randomized subjects as denominator.

Number of days in the study will be summarized with descriptive statistics by treatment group and overall. For each subject, the number of days in the study will be calculated as following:

Number of days in study = Date of completion/discontinuation – 1st dose date* +1

* For randomized subjects who discontinued study before the first application of study treatment, the date of randomization will be considered instead of the date of the first application of study treatment.

A listing of subject's disposition and randomization will be provided. Information on first screening for subjects who were rescreened, including the rescreened subject identifier and the reason for first screening failure, will be presented under the first screening subject identifier. The reason for screening failure will be listed as well.

A table of randomized strata vs. actual strata will be provided if there is any mis-randomization discrepancy.

7.2 Protocol Deviations

A data review will be conducted before database lock by the Medical Monitor and the Sponsor to classify protocol deviations as minor or major.

The number and percentage of subjects with at least one important protocol deviation (including important protocol deviations associated with COVID-19) will be summarized by deviation category and treatment group using the safety analysis set.

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

8 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Demographics and baseline characteristics will be summarized with descriptive statistics using the ITT and safety population. The list of demographics and baseline characteristics to be summarized will include:

- Age (years)
- Age Group: 6-11, 12-17, 6-17 and ≥18, 18-64 and ≥65
- Childbearing potential
- Sex at birth
- Ethnicity
- Race*
- Baseline Height (cm)

- Baseline Weight (kg)
- Baseline Body Mass Index (BMI) (kg/m^2): Adult, Child/Adolescent
- Fitzpatrick Skin Type
- Prior failure of Topical Corticosteroids, Topical Calcineurin Inhibitors, Eucrisa
- Atopic Dermatitis involvement on the face, on the eyelids
- Baseline vIGA-AD
- Average weekly baseline WI-NRS
- Daily baseline WI-NRS
- Baseline BSA (%)
- Baseline BSA (%) Group - <10% and ≥10%, tertile groups
- Baseline EASI total score
- Baseline EASI score group - ≤7 and >7, tertile groups
- Baseline SCORAD
- Baseline DLQI/CDLQI
- 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.

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

Child and Adolescent (6 to 17 years): After BMI is calculated using the same formula above for children and teens, it is expressed as a percentile obtained from either a graph or a percentile calculator. Percentiles will be calculated using data files and instructions provided by the CDC (https://www.cdc.gov/growthcharts/percentile_data_files.htm). Baseline height will be used to derive BMI for each visit since height is not collected at all visits.

A listing of all demographics, analysis population flag, reason not included in the efficacy analysis 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.

Surgical and medical history will be summarized by system organ class (SOC) and preferred term (PT) using the safety analysis set. 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 analysis set.

10 PRIOR AND CONCOMITANT MEDICATIONS

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

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

Incidence of prior and concomitant medications will be tabulated by ATC level 3 and PT using the safety analysis set. A subject with the same medication taken multiple times will be counted only once for the corresponding PT. 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. Prior and concomitant medications will be sorted alphabetically by ATC level and within each ATC level, the PT will be presented by descending order.

11 STUDY TREATMENT EXPOSURE AND TREATMENT COMPLIANCE

A summary of exposure related to Roflumilast cream and the vehicle will be presented using the safety population by treatment group. It will include descriptive statistics on the number of days on IP, as well as the number of investigational product applications based on diary, for each treatment group. The number of days on IP, will be calculated as follows:

[(last treatment date- first treatment date) + 1].

For each subject, investigational product application compliance (%) will be calculated as follows:

$$\frac{\text{Number of investigational product applications}}{\text{Number of expected investigational product applications}} \times 100$$

Number of investigational product applications will be calculated as number of expected investigational product applications minus number of doses missed. Number of doses missed and the date that the dose was missed were collected in eCRF.

Number of expected investigational product applications will be calculated as calculated as last treatment/interruption date – first treatment date + 1. If latest treatment date ≥ latest interruption date, then the latest treatment date will be used; otherwise, latest interruption date will be used in deriving the expected number of IP applications.

Descriptive statistics for the compliance as well as the number of missed applications, subjects with < 80%, [80% - 100%], and >100% compliance will be presented by treatment. Furthermore, the incidence of subjects who missed more than 3 consecutive doses and compliant subjects will be presented by treatment.

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

Total weight of study medication applied (determined by weighing the study medication before and after use) will be summarized by treatment using descriptive statistics. Weight of study medication used will be documented in source documents and in eCRF. Total weight of IP used is determined by subtracting minimum of (returned, remained) tube weight from the maximum of

(dispensed, prepared) tube weight for each tube that was dispensed and summing the weights. Calculate Actual IP Used: maximum of (dispensed, prepared) tube weight minus minimum of (returned, remained) tube weight = ___ grams

If a tube is not returned at the end of the study or a tube weight is missing, actual IP used for the tube and total weight of IP applied during study will be missing.

Number of days on IP and compliance, including compliance collected at each clinic visit, will be displayed in a listing of study treatment administration by subject, for each treatment. A listing of drug accountability including the kit number, tube number, dispensed and returned weight will also be provided.

12 EFFICACY ANALYSIS

12.1 Primary Efficacy Endpoint Analysis

12.1.1 Primary Efficacy Endpoint and Estimand

The vIGA-AD 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). 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 primary efficacy endpoint is vIGA-AD success, defined as an vIGA-AD score of ‘clear’ (0) or ‘almost clear’ (1) plus at least a 2-grade improvement from Baseline at Week 4.

The primary estimand is described by the following attributes:

Population: Patients with Atopic Dermatitis

Endpoint: vIGA-AD success at Week 4

Intercurrent events: In the course of the 4-week randomized treatment period, subjects may be exposed to possible known or unknown intercurrent events that could possibly impact the estimates of the estimand, such as treatment discontinuation due to a specific adverse effect or perhaps a lack of effect. A composite strategy will be implemented that handles subjects who discontinue due to lack of efficacy or adverse event as missing not at random differently than all other subjects. That is, subjects who discontinue due to lack of efficacy or adverse event will be treated as non-responders for all pre-specified analysis visits (refer to Section 5.4) for which the subject’s last dose day falls within the analysis window or is prior to the start of the analysis window (refer to

Section 5.5) while the "Treatment Policy Strategy" will be adopted for handling intercurrent events in this study other than discontinuation due to lack of efficacy or adverse event.

Population-level summary: ratio of the odds of achieving vIGA-AD success after 4 weeks of using roflumilast cream 0.15%, relative to the odds of success after 4 weeks using a matching vehicle cream in the ITT population.

The supportive population-level summary: the proportion difference between Roflumilast cream 0.15% and vehicle groups will be provided for the patients who achieve vIGA-AD success at week 4 in the ITT population.

12.1.2 Hypothesis Testing

Primary hypothesis testing on the odds ratio: The null hypothesis is that the vIGA-AD success does not differ between roflumilast cream 0.15% and matching vehicle cream. The alternative hypothesis is that the vIGA-AD success does differ between roflumilast cream 0.15% and matching vehicle cream.

Null Hypothesis (H_0): $P_R\ Q_V\ /\ P_V\ Q_R = 1.0$,
Alternative Hypothesis (H_A): $P_R\ Q_V\ /\ P_V\ Q_R \neq 1.0$, where

P_R = the proportion of vIGA-AD success in roflumilast cream 0.15%
P_V = the proportion of vIGA-AD success in matching vehicle cream
$Q_R = 1 - P_R$
$Q_V = 1 - P_V$.

12.1.3 Primary Endpoint Analysis

For the primary analysis, missing vIGA-AD scores will be imputed using multiple imputation as described in Section 6.2.1. These imputations will result in a minimum of 25 to a maximum of 250 complete analysis datasets, depending on the number of imputed monotone datasets that are required.

Percentages of subjects having a vIGA-AD Success (refer to Section 12.1.1) will be presented by visit and treatment group based on multiply imputed data in the ITT population along with a 95% Wilson CI. The common MH odds ratio and common MH proportion difference, adjusted for the randomization factors (i.e., randomized vIGA-AD score and pooled study site) will also be provided along with their common associated 95% CIs. Additionally, count and percentage of subjects having a vIGA-AD success, count and percentage of subjects in each category of the vIGA-AD scale, and descriptive statistics for the vIGA-AD scores, change in vIGA-AD score from

baseline and percent change in vIGA-AD score from baseline will be presented by visit and treatment group based on observed data in the ITT population.

The primary endpoint (vIGA-AD Success at Week 4) will then be analyzed using multiple imputation CMH test stratified by the randomization factors. To do so, a CMH analysis will be performed separately for each of the complete multiply imputed analysis data sets, and results will be combined into one multiple imputation inference using the methodology described in Section 6.2.1. Statistical significance will be concluded at the 5% significance level (2-sided). Should the odds ratio be not estimable for at least one multiply imputed dataset, the conclusion of the study will be based on the p-value obtained from a MH test, stratified by the randomization factors, for the common MH proportion difference at Week 4.

The following sensitivity analyses to the primary analysis of the primary endpoint will be performed:

- Multiply-imputed data (refer to Section 6.2.1) on mITT population (refer to Section 4.2) and
- Tipping Point analysis (refer to Section 6.2.3) on ITT population.
- Non-responder imputation (refer to Section 6.2.2) on ITT population.
- Observed data on ITT population (refer to Section 4.1).
- Observed data on PP population (refer to Section 4.3).

For the last three sensitivity analyses, count of subjects having vIGA-AD success will also be presented by visit and treatment group in addition to the percentage of subjects having vIGA-AD success.

To assess the impact of the pooling of the study sites on the primary analysis of the primary endpoint, the following analyses will be performed:

- The primary analysis of the primary efficacy endpoint will be repeated but with a different site pooling strategy (refer to Section 6.4).
- To assess the impact of site on the primary analysis endpoint, the proportion of subjects achieving vIGA-AD success and 95% CI within each site will be tabulated. No p-values will be provided. Forest plots of the proportions (and associated 95% CI) for each site and treatment will also be provided.

- An additional analysis to examine the impact of study site will examine the changes in p-values that occur after removal of subjects from a site. To do so, the primary analysis of the primary efficacy endpoint will be repeated but removing a different pooled study site for each iteration. Forest plots of the odds ratio (and associated 95% CI) for each pooled site removed will also be provided.

12.2 Secondary Endpoints Analysis

Secondary efficacy endpoints are based on the vIGA-AD scale, WI-NRS scale or EASI questionnaire. The list of secondary efficacy endpoints can be found in Section 2 and their derivation in Section 5.5.

- For more details about the vIGA-AD scale refers to Section 12.1.1.
- The WI-NRS scale is a single item scale assessing the subject-reported worst itch severity during the previous 24-hour period. The scale is from 0 to 10 ('no itch' to 'worst itch imaginable').
- For the EASI questionnaire, 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): 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%).

For each secondary efficacy endpoint, missing data will be imputed and data will be summarized as for the primary efficacy endpoint (refer to Section 12.1.3).

Upon successful demonstration of statistical significance for the primary efficacy endpoint, the secondary efficacy endpoint of vIGA-AD success among subjects with vIGA-AD score of 'Moderate' at randomization will be analyzed as described for the primary efficacy endpoint (refer to Section 12.1.3) but based on the vIGA-AD Moderate ITT population (refer to Section 4.4) and stratifying by pooled study site only. That is, as per the vIGA-AD Moderate ITT population definition all subjects included in this analysis will have a randomized vIGA-AD score of Moderate (3) and so, the CMH test cannot be stratified by randomized vIGA-AD score.

Upon successful demonstration of statistical significance for the primary and above secondary efficacy endpoints, the remaining secondary efficacy endpoints will be grouped into secondary

endpoint family 1, comprised of the 4-point reduction on the average weekly WI-NRS at Week 4, Week 2 and Week 1, and secondary endpoint family 2, comprised of the EASI-75 at Week 4, vIGA-AD of 'clear' or 'almost clear' at Week 4, vIGA-AD of success at week 2 and week 1, vIGA-AD of 'clear' or 'almost clear' at Week 2 and Week 1. The analysis of secondary endpoint family 1 will be performed as described for the primary efficacy endpoint (refer to Section 12.1.3) but based on the WI-NRS population (refer to Section 4.6) and an alpha level of 0.03 will be used to test these endpoints sequentially as described in Section 6.5. Similarly, the analysis of secondary endpoint family 2 will be performed as described for the primary efficacy endpoint (refer to Section 6.2.1) based on the ITT population (refer to Section 4.1) but an alpha level of 0.02 will be used to test these endpoints sequentially as described in Section 6.5.

Secondary Endpoint Family 1 (α=0.03, hierarchical testing)

- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the average weekly WI-NRS at Week 4
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the average weekly WI-NRS at Week 2
- In subjects ≥12 years old with baseline WI-NRS ≥ 4, achievement of at least a 4-point reduction in the average weekly WI-NRS at Week 1

Secondary Endpoint Family 2 (α = 0.02, hierarchical testing)

- Achievement of at least a 75% reduction in the Eczema Area and Severity Index at Week 4 (EASI-75)
- vIGA-AD of 'clear' or 'almost clear' at Week 4
- vIGA-AD Success at Week 2
- vIGA-AD Success at Week 1
- vIGA-AD of 'clear' or 'almost clear' at Week 2
- vIGA-AD of 'clear' or 'almost clear' at Week 1

For the secondary efficacy endpoint of vIGA-AD Success at Week 4 based on the vIGA-AD moderate ITT population, the following sensitivity analyses to the primary analysis of this efficacy endpoint will be performed if and only if the hierarchical testing procedure (refer to Section 6.5) allows it:

- Tipping Point analysis (refer to Section 6.2.3) on vIGA-AD moderate ITT population.

- Non-responder imputation (refer to Section 6.2.2) on vIGA-AD moderate ITT population.
- Observed data on vIGA-AD moderate ITT population.
- Observed data on vIGA-AD moderate PP population (refer to Section 4.5).

For the last three sensitivity analyses, count of subjects having vIGA-AD success will also be presented at Week 4 by treatment group in addition to the percentage of subjects having vIGA-AD success.

For all other secondary efficacy endpoints, the following sensitivity analyses to the primary analysis of these secondary endpoints will be performed if and only if the hierarchical testing procedure (refer to Section 6.5) allows it:

- Non-responder imputation on WI-NRS population for the WI-NRS secondary efficacy endpoints and ITT population for all other secondary efficacy endpoints.
- Observed data on WI-NRS population for the WI-NRS secondary efficacy endpoints and ITT population for all other secondary efficacy endpoints.

The daily WI-NRS score, average weekly WI-NRS, and WI-NRS success flag will be listed.

The EASI total score and EASI-75 flag at each visit will be listed.

12.3 Exploratory Efficacy Endpoints

Exploratory efficacy endpoints are based on the vIGA-AD scale, WI-NRS scale, EASI questionnaire, %BSA affected by AD, CDLQI/DLQI questionnaires, DFI questionnaire, SCORAD tool or POEM tool. The list of exploratory efficacy endpoints can be found in Section 2 and their derivation in Section 5.5.

- For more details about the vIGA-AD scale, refer to Sections 12.1.1.
- For more details about the WI-NRS scale, refer to Section 12.2.
- For more details about the EASI questionnaire, refer to Section 12.2.
- The % 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, and soles).
- The CDLQI/DLQI is a self-administered validated questionnaire designed to measure the health-related quality of life of children/adult subjects suffering from a skin disease. It

consists of 10 questions concerning subjects' perception of the impact of skin disease on different aspects of their health-related quality of life over the last week. Questions 1 to 6 and 8 to 10 are rated from 0 (Not at all) to 3 (Very much). For children, if last week was a school time, question 7 is rated from 0 (Not at all) to (Prevented school) but if the last week was a holiday time, question 7 is rated from 0 (Not at all) to 3 (Very much). For adults, question 7 is rated from 0 (Not at all/Not relevant/No) to 3 (Yes).

- The DFI questionnaire measures how much having a child with AD affects the quality of life of other (adult) members of the family over the last week. It is designed to be completed by caregivers of subjects ≤ 17 years of age and consists of 10 questions rated from 0 (Not at all) to 3 (Very much).

- The SCORAD is a clinical tool to assess 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).

- The POEM is a tool used for monitoring atopic eczema severity. If focus on the illness as experience 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).

For the continuous exploratory efficacy endpoints of change and percent from baseline in average weekly WI-NRS, EASI total score, % BSA, DLQI/CDLQI score, DFI score, SCORAD score, and POEM score at Week 1, Week 2, and Week 4, descriptive statistics for the score, change from baseline and percent change from baseline will be presented by visit and treatment group based on observed data in the ITT population. Descriptive statistics will also be presented similarly for the average weekly WI-NRS at Week 3. For the continuous efficacy exploratory endpoint of daily WI-NRS, descriptive statistics will be presented similarly but on a daily basis instead of a weekly basis. Additionally, a plot of the mean (and standard error) daily WI-NRS scores over time for each treatment group will also be provided based on observed data. A similar plot will be provided for the percent change from baseline in daily WI-NRS.

For each categorical exploratory efficacy endpoint, count and percentages of subjects meeting the criteria for an exploratory efficacy endpoint will be presented by visit and treatment group based on observed data in the ITT population along with a 95% Wilson CI. Additionally, count and percentage of subjects meeting the criteria for vIGA-AD of Clear (0), EASI-50, EASI-90, and EASI-100 at Week 1 and Week 2 will also be presented by visit and treatment group based on observed data in the ITT population along with a 95% Wilson CI.

Continuous exploratory efficacy endpoints will be analyzed at Week 1, Week 2, and Week 4 and for daily WI-NRS assessments using an analysis of covariance (ANCOVA) with the factors of treatment, two stratification variables (pooled study site and randomized vIGA-AD score), and baseline of the variable under analysis as covariate. Statistical comparisons between the treatment groups will be obtained using contrasts. The Least Square (LS) mean and its standard error, difference in LS means between treatment group (i.e., active - vehicle), its standard error and associated 95% confidence interval, and p-value for difference from vehicle will be presented at each visit. These analyses will be performed based on observed data in the ITT population.

Categorical exploratory efficacy endpoints at Weeks 1, 2, and 4 will be analyzed using a CMH test adjusted for the two stratification variables (pooled study site and randomized vIGA-AD score). Common MH odds ratio, common MH proportion difference and their common associated 95% CI, adjusted for the randomization factors, will be provided. The p-value will be from a CMH test for the common MH odds ratio, unless the common MH odds ratio is not estimable. In such circumstances, the p-value will be from a MH test for the common MH proportion difference.

The analysis of these endpoints will be performed on the observed data with no imputation. The p-values will be nominal as no formal inferential testing will be done on exploratory efficacy endpoints.

12.4 Subgroup Analysis

With the following exception, analyses of the primary and secondary efficacy endpoints (refer to Sections 12.1 and 12.2, respectively) will be repeated by subgroups (refer to Section 6.6) based on the multiple imputation data using mITT, ITT or WI-NRS population, as applicable (refer to Section 12.2). The exception is that the subgroup analyses by randomized vIGA-AD score and by actual baseline vIGA-AD score will not be performed for the secondary efficacy endpoint of vIGA-AD Success based on the vIGA-AD Moderate ITT population.

The following alternatives could be implemented if the odds ratio and/or proportion difference are not estimable:

1. An unstratified model can be used if odds ratios are not estimable due to over-stratification, i.e., too few events for the number of strata.
2. If the between-imputation variance of odds ratios and/or proportion difference is 0, i.e., the estimates are equal across all imputations, the estimates from any imputation will be reported.

Should a specific subgroup have less than 10 subjects across both treatment groups, no statistical inference will be performed. Forest plots of subgroup analysis for primary efficacy endpoint and secondary efficacy endpoints will also be provided.

No subgroup analyses will be presented for the exploratory efficacy endpoint (refer to Section 2).

12.5 Summary of Primary and Secondary Efficacy Analysis

Table 5 provides a summary of the primary and sensitivity analyses that will be provided for primary and secondary efficacy endpoints.

Table 6 Summary of Primary and Secondary Efficacy Analyses

Efficacy Endpoint	**Primary Analysis**	**Sensitivity Analysis**
Primary (refer to Section 12.1)		
vIGA-AD Success (a score of '0' or '1' plus at least a 2-grade improvement) at week 4	ITT, multiple imputation (CMH)	#1 mITT, multiple imputation (CMH) #2 ITT, Tipping point (CMH)) #3 ITT, non-responder imputation (CMH) #4 ITT, observed data (CMH) #5 PP, observed data (CMH) #6 ITT, multiple imputation (CMH) and a different site pooling strategy (CMH) #7 ITT, multiple imputation (CMH) by pooled study site #8 ITT, multiple imputation (CMH) and removing one pooled study site at the time
Secondary (refer to Section 12.2)		

vIGA-AD Success (a score of '0' or '1' plus at least a 2-grade improvement) at week 4	vIGA-AD Moderate ITT, multiple imputation (CMH)	#1 vIGA-AD Moderate ITT, Tipping point (CMH) #2 vIGA-AD Moderate ITT, non-responder imputation (CMH) #3 vIGA-AD Moderate ITT, observed data (CMH) #4 vIGA-AD Moderate PP, observed data (CMH)
Family 1 (α=0.03): Average weekly WI-NRS Success (achievement of at least a 4-point reduction) at weeks 1,2, and 4	WI-NRS population, multiple imputation (CMH)	#1 WI-NRS, non-responder imputation (CMH) #2 WI-NRS, observed data (CMH)
Family 2 (α=0.02 or α=0.05): EASI-75 at week 4	ITT, multiple imputation (CMH)	#1 ITT, non-responder imputation (CMH) #2 ITT, observed data (CMH)
Family 2 (α=0.02 or α=0.05): vIGA-AD Success (a score of '0' or '1' plus at least a 2-grade improvement) at weeks 1 and 2	ITT, multiple imputation (CMH)	#1 ITT, non-responder imputation (CMH) #2 ITT, observed data (CMH)
Family 2 (α=0.02 or α=0.05): vIGA-AD score of '0' or '1' at weeks 1, 2 and 4	ITT, multiple imputation (CMH)	#1 ITT, non-responder imputation (CMH) #2 ITT, observed data (CMH)

13 SAFETY ANALYSIS

Safety analyses will be conducted using the safety population. Subjects will be analyzed based on the treatment received and the stratum they belong to.

No formal inferential statistics will be performed on safety assessments.

Also, treatment-emergent adverse events by SOC and PT table and summary of weight change will be presented by treatment groups for age 6 to 17 years and ≥ 18 years separately.

13.1 Adverse Events

Adverse events (AEs) will be coded according to the latest available version of Medical Dictionary for Regulatory Activities (MedDRA) (Version 25.0).

Treatment emergent adverse events (TEAEs) are defined as any AEs with onset on or after the first study drug application. 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. All reported TEAEs will be summarized by treatment group.

Overall summary will be presented, which will include the total number of events, and the number and percentage of subjects who experienced TEAE, TEAE by the strongest relationship, TEAE by the maximum severity, treatment-related TEAE by maximum severity, treatment-emergent serious AE (TESAE), 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.

The number and percentage of subjects who experience TEAE will be summarized by SOC and PT within SOC. In addition, similar table will be presented by treatment groups for age 6 to 17 years and ≥ 18 years separately. Unless otherwise specified, a subject experiencing the same TEAE multiple times will be counted only once for the corresponding PT. Similarly, if a subject experiences multiple TEAEs within the same SOC, the subject 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 the safety analysis set. 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 be summarized by SOC, PT and the maximum severity (mild/moderate/severe/life threatening/death related to AE). 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 TESAE, treatment-emergent Non-SAE, will be summarized by SOC and PT within SOC.

Frequency and percentage of subjects who experience TEAE on an application site will be summarized by SOC and PT.

A table and plot of most frequent TEAE (‘≥1%) by PT will be provided by treatment arms (overall TEAE, overall TESAE will be included in the same plot).

All the AEs will be listed. Any TEAE leading to death will also be included in the AE listing (if there is any). The TEAE related to application site will be flagged in the AE listing.

13.2 Clinical Laboratory

Descriptive statistics for the observed values in chemistry, hematology, and quantitative urinalysis, change from baseline and percent change from baseline values will be presented by treatment group at each scheduled visit. For qualitative urinalysis data, the number and percentage of the subjects for each level of result by treatment at each scheduled visit will be provided.

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

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

Laboratory data will be presented in SI units.

13.3 Vital Signs

Descriptive statistics will be presented for vital signs (systolic blood pressure, diastolic blood pressure, heart rate, body temperature, and weight). Observed values, change from baseline and percent change from baseline values will be presented by treatment group at each scheduled visit.

The number and percentage of subjects with gain or lose >5% from baseline in body weight over the course of the study will be summarized for overall and separately for children and adolescents (6 to 17 years) and adults (≥ 18 years). BMI will be summarized using descriptive statistics (n, mean, standard deviation, median, Q1, Q3, minimum, and maximum) on observed values, change from baseline and percent change from baseline values separately for children and adolescents (6 to 17 years) and adults (≥ 18 years). The BMI percentile rather than BMI kg/m^2 will be summarized for children and adolescents (6 to 17 years).

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

13.4 Local Tolerability Assessments

The investigator's assessment of the application site reaction will be summarized by visit using both categorical methods (number and percentage of subject with each score) as well as continuous methods (e.g., mean, standard deviation, etc.)

Local tolerability (burning/stinging sensation) assessed by the subject will be summarized using number and percentage similarly.

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

The Modified PHQ-A Assessment will be performed in adolescent subjects (12-17 years old, inclusive; question 9 has been removed since that is better evaluated by use of the C-SSRS tool).

The PHQ-8/Modified PHQ-A score is the sum of the responses for the questions, each question ranging from 0 (Not at all) to 3 – (Nearly every day). Five severity categories of depression are defined 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 score will be set to missing in case of at least one missing value. The number and percentage of subjects in each category will be summarized by treatment and visit. A summary of the shifts in depression category from baseline to each study visit will also be provided.

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 baseline will be calculated for CDI-2 total score and the 2 scales, i.e., emotional problems and functional problems, and will be summarized descriptively by treatment group and visit.

The CDI-2 total score will be categorized as follows:

1. Normal: Male Score < 22, Female Score < 21
2. Elevated: $22 \leq$ Male Score < 32, $21 \leq$ Female Score < 32
3. Very Elevated: Male or Female Score ≥ 32.

The proportion of subjects that meeting the criteria will be summarized by treatment group and visit.

A summary of the shifts in CDI-2 category from baseline to each study visit will also be provided.

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

The Columbia-Suicide Severity Rating Scale (C-SSRS) is a questionnaire used (for adolescents and adults 12 years old and older) for suicide assessment developed by multiple institutions, including Columbia University. The C-SSRS prospectively assesses Suicidal Ideation and Suicidal Behavior. At the Screening study visit, "Baseline/Screening" version of the C-SSRS will be used. This version assesses Suicidal Ideation and Suicidal Behavior during the subject's lifetime and during the past 6 months. For the Screening visit, "lifetime" experience of the subject with Suicidal Ideation and Suicidal Behavior will be summarized. On all subsequent visits, the "Since Last Visit" version will be used (Baseline/Day 1, Week 1/Day 8, Week 2/Day 15 or Week 4/Day 29).

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.

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 and treatment group at each study visit.

14 PHARMACOKINETICS ANALYSIS

Concentration data will be summarized by visit by age group (<18 vs. >=18) and overall, for active treatment group using descriptive statistics, reporting n, mean, standard deviation, median, Q1, Q3, minimum, and maximum, and geometric statistics including geometric mean and coefficient of variation. For computation of mean plasma concentrations, data that are below the limit of quantification (BLQ) will be set to 0.0001. The PK population will be used for these analyses.

PK data will be presented in the listing.

15 REFERENCES

1. Anne Lott & Jerome P. Reiter (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

2. Ratitch, B., Lipkovich, I., & O'Kelly, M. (2013). *Combining Analysis Results from Multiply Imputed Categorical Data*. PharmaSUG. https://www.pharmasug.org/proceedings/2013/SP/PharmaSUG-2013-SP03.pdf

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.

P-values equal to or above 0.0001 will be reported to 4 decimal places; p-values less than 0.0001 will be reported as "<0.0001"; p-values greater than 0.9999 will be reported as ">0.9999".

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 group, 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:

Study Treatment Full Name	Study Treatment Output Name
Roflumilast cream 0.15%	Roflumilast Cream 0.15%
Vehicle cream	Vehicle

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 day 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 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 date is the same as the first dose date and event occurred prior to study drug application (as flagged in CRF), impute to 00:00.
- If the event date is the same as the first dose date and event did not occur prior to study drug application (as flagged in CRF), 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)

Impute to 23:59.

[redacted]

Signer Events	Signature	Timestamp

In Person Signer Events	Signature	Timestamp
Editor Delivery Events	**Status**	**Timestamp**
Agent Delivery Events	**Status**	**Timestamp**
Intermediary Delivery Events	**Status**	**Timestamp**
Certified Delivery Events	**Status**	**Timestamp**
Carbon Copy Events	**Status**	**Timestamp**
Witness Events	**Signature**	**Timestamp**
Notary Events	**Signature**	**Timestamp**

Envelope Summary Events	Status	Timestamps
Envelope Sent	Hashed/Encrypted	11/11/2022 8:41:13 AM
Certified Delivered	Security Checked	11/11/2022 1:53:29 PM
Signing Complete	Security Checked	11/11/2022 1:54:04 PM
Completed	Security Checked	11/11/2022 1:54:04 PM

Payment Events	Status	Timestamps
Electronic Record and Signature Disclosure		

ELECTRONIC RECORD AND SIGNATURE DISCLOSURE

From time to time, Innovaderm Research Inc. (we, us or Company) may be required by law to provide to you certain written notices or disclosures. Described below are the terms and conditions for providing to you such notices and disclosures electronically through the DocuSign system. Please read the information below carefully and thoroughly, and if you can access this information electronically to your satisfaction and agree to this Electronic Record and Signature Disclosure (ERSD), please confirm your agreement by selecting the check-box next to 'I agree to use electronic records and signatures' before clicking 'CONTINUE' within the DocuSign system.

Getting paper copies

At any time, you may request from us a paper copy of any record provided or made available electronically to you by us. You will have the ability to download and print documents we send to you through the DocuSign system during and immediately after the signing session and, if you elect to create a DocuSign account, you may access the documents for a limited period of time (usually 30 days) after such documents are first sent to you. After such time, if you wish for us to send you paper copies of any such documents from our office to you, you will be charged a $0.00 per-page fee. You may request delivery of such paper copies from us by following the procedure described below.

Withdrawing your consent

If you decide to receive notices and disclosures from us electronically, you may at any time change your mind and tell us that thereafter you want to receive required notices and disclosures only in paper format. How you must inform us of your decision to receive future notices and disclosure in paper format and withdraw your consent to receive notices and disclosures electronically is described below.

Consequences of changing your mind

If you elect to receive required notices and disclosures only in paper format, it will slow the speed at which we can complete certain steps in transactions with you and delivering services to you because we will need first to send the required notices or disclosures to you in paper format, and then wait until we receive back from you your acknowledgment of your receipt of such paper notices or disclosures. Further, you will no longer be able to use the DocuSign system to receive required notices and consents electronically from us or to sign electronically documents from us.

All notices and disclosures will be sent to you electronically

Unless you tell us otherwise in accordance with the procedures described herein, we will provide electronically to you through the DocuSign system all required notices, disclosures, authorizations, acknowledgements, and other documents that are required to be provided or made available to you during the course of our relationship with you. To reduce the chance of you inadvertently not receiving any notice or disclosure, we prefer to provide all of the required notices and disclosures to you by the same method and to the same address that you have given us. Thus, you can receive all the disclosures and notices electronically or in paper format through the paper mail delivery system. If you do not agree with this process, please let us know as described below. Please also see the paragraph immediately above that describes the consequences of your electing not to receive delivery of the notices and disclosures electronically from us.

How to contact Innovaderm Research Inc.:

You may contact us to let us know of your changes as to how we may contact you electronically, to request paper copies of certain information from us, and to withdraw your prior consent to receive notices and disclosures electronically as follows:
To contact us by email send messages to: [redacted]

To advise Innovaderm Research Inc. of your new email address

To let us know of a change in your email address where we should send notices and disclosures electronically to you, you must send an email message to us at [redacted] and in the body of such request you must state: your previous email address, your new email address. We do not require any other information from you to change your email address.

If you created a DocuSign account, you may update it with your new email address through your account preferences.

To request paper copies from Innovaderm Research Inc.

To request delivery from us of paper copies of the notices and disclosures previously provided by us to you electronically, you must send us an email to [redacted] and in the body of such request you must state your email address, full name, mailing address, and telephone number.

To withdraw your consent with Innovaderm Research Inc.

To inform us that you no longer wish to receive future notices and disclosures in electronic format you may:

i. decline to sign a document from within your signing session, and on the subsequent page, select the check-box indicating you wish to withdraw your consent, or you may;

ii. send us an email to [redacted] and in the body of such request you must state your email, full name, mailing address, and telephone number. We do not need any other information from you to withdraw consent.. The consequences of your withdrawing consent for online documents will be that transactions may take a longer time to process..

Required hardware and software

The minimum system requirements for using the DocuSign system may change over time. The current system requirements are found here: https://support.docusign.com/guides/signer-guide-signing-system-requirements.

Acknowledging your access and consent to receive and sign documents electronically

To confirm to us that you can access this information electronically, which will be similar to other electronic notices and disclosures that we will provide to you, please confirm that you have read this ERSD, and (i) that you are able to print on paper or electronically save this ERSD for your future reference and access; or (ii) that you are able to email this ERSD to an email address where you will be able to print on paper or save it for your future reference and access. Further, if you consent to receiving notices and disclosures exclusively in electronic format as described herein, then select the check-box next to 'I agree to use electronic records and signatures' before clicking 'CONTINUE' within the DocuSign system.

By selecting the check-box next to 'I agree to use electronic records and signatures', you confirm that:

- You can access and read this Electronic Record and Signature Disclosure; and
- You can print on paper this Electronic Record and Signature Disclosure, or save or send this Electronic Record and Disclosure to a location where you can print it, for future reference and access; and

[redacted]