STATISTICAL ANALYSIS PLAN

ARC003
Peanut Allergy Oral Immunotherapy Study of AR101 for Desensitization in Children and Adults (PALISADE)

Final Version 2.0 – 24 Jan 2018
Reference Numbers: NCT02635776, EudraCT 2015-004257-41

Aimmune Therapeutics, Inc.
8000 Marina Blvd, Suite 300
Brisbane, CA 94005
United States

Sponsor Name:	**Aimmune Therapeutics, Inc.** **(formerly Allergen Research Corporation)**
Protocol Number and Title:	**ARC003** **Peanut Allergy Oral Immunotherapy Study of AR101 for Desensitization in Children and Adults (PALISADE)**
Protocol Version and Date:	Protocol Amendment 4 (Global); 31 July 2017
Author:	Brian Beus Principal Biostatistician Agility Clinical, Inc.
SAP Version:	Final Version 2.0
SAP Version Date:	24Jan2018

Version: Final Version 2.0 **Version Date: 24Jan2018**

I confirm that I have reviewed this document and agree with the content.

Agility Clinical, Inc.

25 JAN 2018

Brian Beus, MS
Principal Biostatistician
Date

Aimmune Therapeutics, Inc

25 January 2018

Daniel Adelman, MD
Chief Medical Officer
Date

25 JAN 2018

Rezi Zawadzki, DrPH
Senior Director, Biometrics
Date

TABLE OF CONTENTS

1. PURPOSE 9
1.1. Document History 9
1.2. Responsibilities 10
1.3. Timing of Analyses 10
2. STUDY OBJECTIVES 11
2.1. Primary Objective 11
2.2. Secondary Objectives 11
2.3. Brief Description of Study Design 11
2.4. Determination of Sample Size 14
2.5. Treatment Assignment and Blinding Considerations 16
2.6. Administration of Study Treatment 19
2.7. Double Blind Placebo Controlled Food Challenge 20
3. ENDPOINTS 22
3.1. Primary Efficacy Endpoint 22
3.2. Secondary Endpoints 22
3.2.1. Key Secondary Efficacy Endpoints 22
3.2.2. Other Secondary Endpoints 23
3.3. Secondary Safety Endpoints 23
3.4. Exploratory Endpoints 24
4. ANALYSIS POPULATIONS 25
4.1. Safety Population 25
4.2. Intent-to-Treat (ITT) Population 25
4.3. Completer Population 25
4.4. Per Protocol (PP) Population 25
5. PROTOCOL DEVIATIONS 27
6. GENERAL ASPECTS FOR STATISTICAL ANALYSIS 28
6.1. Strata and Covariates 28
6.2. Subgroup Analyses 28
6.3. Multiple Comparisons and Multiplicity 29
6.4. Significance Level 29

6.5. Statistical Notation and Methodology 29
7. DATA HANDLING METHODS 31
7.1. Visit Windows 31
7.2. Data Presentation 31
7.3. Maximum Tolerated Dose at DBPCFC 31
7.4. Classification of Responder Status 32
7.4.1. Classification as a responder 32
7.4.2. Classification as a non-responder 32
7.4.3. Exit DBPCFC imputation rules for non-responders 32
7.4.4. Classification of Escalation and OIT Maintenance Failures 32
7.5. Data Derivations and Definitions 33
7.6. Missing Data 35
7.7. Pooling 36
8. DEMOGRAPHIC, OTHER BASELINE CHARACTERISTICS 37
8.1. Subject Disposition and Withdrawals 37
8.2. Protocol Deviations 37
8.3. Demographic and Other Baseline Characteristics 37
8.4. Allergy History 38
8.5. Non-Peanut Allergy History 38
8.6. Other Medical History 38
9. EFFICACY 39
9.1. Primary Endpoint 39
9.2. Additional Analyses of the Primary Efficacy Endpoint 39
9.3. Key Secondary Efficacy Endpoints 41
9.3.1. Desensitization Response Rates 42
9.3.2. Maximum Severity of Symptoms 42
9.4. Additional Analysis of Key Secondary Endpoints 43
9.5. Analysis of Other Secondary Endpoints 44
9.5.1. Maximum Tolerated Dose at Screening and Exit DBPCFCs 45
9.5.2. Change from Baseline in MTD at DBPCFC 46
9.5.3. Use of Epinephrine as a Rescue Medication at Exit DBPCFC 47
9.5.4. Peanut-specific IgE and IgG_4 48
9.5.5. Peanut Skin Prick Test 49
9.5.6. Quality of Life Assessments 50

9.6. Exploratory Analyses 51
9.6.1. Treatment Satisfaction 51
9.6.2. Secondary Efficacy Endpoints Analyzed in Adults and Overall 52
9.6.3. Baseline Factors that may Influence Desensitization 52
9.6.4. Effect on Desensitization Response Rate 53
9.6.5. Effect on Maximum Tolerated Dose at Exit DBPCFC 54
9.6.6. Imputation of MTD based on Maximum AR101 Dose Achieved 54
9.6.7. Cutoff Predictive of MTD ≤ 30 55
9.6.8. Relationship of ps-IgE and SPT on Premature Discontinuations 55
9.7. Immune Tolerance Network (ITN) Samples 55
9.8. Interim Analysis 55
9.9. Data Safety Monitoring Committee 55
10. SAFETY 57
10.1. Study Treatment Exposure 57
10.2. Prior, Concomitant, And Rescue Medications and Therapies 59
10.3. Adverse Events 61
10.4. Food Allergy Episodes 63
10.5. Symptoms During DBPCFC 63
10.6. Symptoms During In-Clinic Study Drug Dosing 63
10.7. Adverse Events During At-Home Dosing 64
10.8. Pregnancy Test Results 64
10.9. Spirometry and PEFR 64
10.10. Vital Signs 64
10.11. Physical Examination 65
10.12. Assessment of asthma control 65
10.13. Assessment of GI Symptoms by PEESS 66
10.14. Epinephrine Use as Rescue Medication 66
10.15. Anaphylaxis Episodes 67
10.16. Algorithmic Search of Systemic Hypersensitivity Events 68
11. PRE-DATABASE LOCK BLINDED DATA REVIEWS 72
12. CHANGE FROM ANALYSIS PLANNED IN PROTOCOL 73
13. REFERENCE LIST 74
14. PROGRAMMING CONSIDERATIONS 76

14.1. General Considerations 76
14.2. Reporting Conventions 76
15. QUALITY CONTROL 78
16. STUDY SCHEDULE 79
17. INDEX OF TABLES, LISTINGS AND FIGURES 80

GLOSSARY OF ABBREVIATIONS

Abbreviation	Description
ACT	Asthma Control Test
AE	Adverse Event
ANCOVA	Analysis of Covariance
ATC	Anatomic Therapeutic Class
CI	Confidence Interval
AR101	Characterized Peanut Allergen
CRF	Case Report Form
CRO	Contract Research Organization
DBPCFC	Double-Blind Placebo-Controlled Food Challenge
DLS	Dose-Limiting Symptom
DSMC	Data Safety Monitoring Committee
eCRF	electronic Case Report Form
ECG	Electrocardiogram
EU	Europe
FAIM	Food Allergy Independent Measure
FAQLQ	Food Allergy Related Quality of Life Questionnaire
FDA	Food and Drug Administration
FEV_1	Forced Expiratory Volume in one second
FVC	Forced Vital Capacity
GI	Gastrointestinal
ICH	International Conference on Harmonization
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IND	Investigational New Drug Application
IRB	Institutional Review Board
ITN	Immune Tolerance Network
ITT	Intent-to-Treat
IXRS	Interactive Voice or Web-Based Response System

Abbreviation	Description
LLOQ	Lower Limit of Quantification
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified Intent-to-Treat
MTD	Maximum Tolerated Dose
NA	North America
OIT	Oral Immunotherapy
PEFR	Peak Expiratory Flow Rate
PPV	Positive Predictive Value
PRACTALL	Practical Issues in Allergology, Joint United States/European Union Initiative
PRN	As needed (pro re nata)
PP	Per Protocol
ps	Peanut-specific
PT	Preferred Term
QC	Quality Control
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SMQ	Standard MedDRA Query
SOC	System Organ Class
SOP	Standard Operating Procedure
SPT	Skin Prick Test
TEAE	Treatment Emergent Adverse Event
TLF	Table, Listing and Figure
TSQM-9	Treatment Satisfaction Questionnaire for Medication
ULOQ	Upper Limit of Quantification
WHODRUG	World Health Organization Drug Dictionary

1. PURPOSE

The purpose of this statistical analysis plan (SAP) is to ensure that the data listings, summary tables and figures which will be produced, and the statistical methodologies that will be used, are complete and appropriate to allow valid conclusions regarding the study objectives.

1.1. DOCUMENT HISTORY

SAP Final Version 1 was approved prior to the first ARC003 subject unblinding (Dec2016). Updates for Version 2 of the SAP are being made post-unblinding of several subjects. It is important to note that all Aimmune and Agility Clinical, Inc. (Agility) staff involved in SAP development and study conduct responsibilities remain blinded to subjects' treatment assignments. The staff will remain blinded until the full study is unblinded post database lock.

Version 2 of the SAP incorporates feedback received by Aimmune from regulatory authorities on the ARC003 study protocol. The FDA correspondence was dated 13Jan2017 and notes the FDA's concern about the number of adult subjects in the study. Per the FDA, licensure of AR101 will likely be possible only in pediatric subjects ages 4-17 years. As such, the efficacy and safety endpoints have been modified in this version of the SAP and in the study protocol to note the primary population for analysis of all endpoints is in the pediatric population. The definition of EU primary and secondary endpoints were integrated in this SAP.

Other notable changes from Version 1 to Version 2 of the SAP include:

- The proportion of subjects aged 18 to 55 years who tolerate a single highest dose of at least 600 mg (for North America) and of at least 1000 mg (for EU) of peanut protein with no more than mild symptoms at the Exit DBPCFC was added as a key secondary efficacy endpoint.
- A tipping point sensitivity analysis of the primary efficiacy endpoint was added based on FDA feedback.
- The mITT population was removed as an analysis population since it was decided that sufficient sensistivity analyses of efficacy endpoints are defined in the SAP without including an mITT population.
- An SMQ search analysis of hypersensitivity events occurring with 2 hours of each other will be presented.
- Adverse event summaries combining adverse events collected on the adverse event page with including in-clinic dosing symptoms will be presented

Other changes have been made but do not materially alter analytical methods or planned summaries detailed in Final Version 1 of the SAP (dated 07Dec2016).

1.2. RESPONSIBILITIES

Upon commencement of this study, INC Research was to provide all statistical and programming services for ARC003. Array Biostatistics, LLC (Array) was providing consulting and review services on behalf of Aimmune Therapeutics, Inc. During the course of the trial, the statistical and programming responsibilities for the study were transferred to to Agility . This included further development and finalization of the SAP that was initially drafted by INC Research.

Agility will perform the final statistical analyses and be responsible for the production and quality control of all tables, figures and listings.

INC Research will perform analyses prepared for the Data Safety Monitoring Committee (DSMC).

1.3. TIMING OF ANALYSES

Interim Analysis No interim analysis of efficacy is planned for this study.

DSMC (Data Safety Monitoring Committee). The DSMC will meet approximately every 3 months to monitor the study for safety. An unblinded team from INC Research biostatistics will perform the safety analyses. Details on the DSMC and unblinding will be described in the DSMC Charter.

Final Analysis The final analysis of safety and efficacy is planned after all subjects complete Exit/Early Discontinuation Visit assessments. The final analysis will include all data collected through the time of database lock.

2. STUDY OBJECTIVES

2.1. PRIMARY OBJECTIVE

The primary objective is to demonstrate the efficacy of AR101, a pharmaceutical-grade peanut allergen formulation, through reduction in clinical reactivity to limited amounts of peanut allergen in peanut-allergic children ages 4-17 years, inclusive.

2.2. SECONDARY OBJECTIVES

The secondary objectives of the study are:

- To demonstrate the safety of AR101 as measured by the incidence of adverse events, including serious adverse events in children ages 4-17 years, inclusive.

- To evaluate the immunological effects of peanut oral immunotherapy (OIT) in children ages 4-17 years, inclusive.

2.3. BRIEF DESCRIPTION OF STUDY DESIGN

This is an international, multicenter, randomized, double-blind, placebo-controlled study of the efficacy and safety of AR101 in a characterized desensitization OIT regimen in peanut-allergic individuals. The study will consist of a screening phase, that includes a Screening double-blind, placebo-controlled food challenge (DBPCFC), and a double-blind OIT Treatment Phase that includes an initial escalation period, an up-dosing period, and a maintenance period, followed by an Exit DBPCFC. The DBPCFC at screening and at exit is to be performed under double-blind conditions so that neither the subject, nor the subject's caregiver, nor any of the clinic staff (save for the unblinded preparer of the challenge foods) knows which challenge contains the peanut or the placebo. See SAP section 2.7 for a summary and protocol section 6.6 for additional details of the DBPCFC.

Peanut-allergic children and adults (ages 4-55 inclusive) will have an initial screening DBPCFC of up to 100 mg (144 mg cumulative) peanut protein or placebo. Those experiencing dose-limiting symptoms at or before the 100 mg dose of peanut protein (measured as 200 mg of peanut flour) will be eligible. Those who develop dose limiting symptoms (DLSs) in reaction to the placebo part of the screening DBPCFC will not be eligible. See Protocol Section 4 for a complete list of inclusion and exclusion criteria.

Approximately 500 peanut-allergic subjects will be randomized 3:1 to peanut OIT versus placebo. At least 80% of the subjects randomized will be children. All eligible subjects will receive escalating doses of either AR101 or placebo. The Treatment Phase comprises 3 periods:

- **Initial Escalation Period** – 2 days in duration;
- **Up-dosing Period** – This is defined as the time from the subject's first home dose of study product at 3 mg to the subject's first in-clinic dose at 300 mg, ideally 20

weeks in duration, but may be extended to a maximum of 40 weeks to accommodate dose reductions and re-escalations, if necessary;

- **Maintenance Period** – The Maintenance Period starts with the first home dose of 300 mg. Ideally, it will be 24 weeks in duration, but it may be extended by up to an additional 4 weeks (for a maximum Maintenance Period duration of 28 weeks), or to a total Treatment Phase duration of 68 weeks, whichever occurs first, to accommodate dose reductions and re-escalations that may occur during the final weeks of the Maintenance Period.

Subjects who reach the targeted dose of 300 mg/d and maintain that dose for approximately 24 weeks will undergo an Exit DBPCFC of up to 1000 mg (2043 mg cumulative) peanut protein or placebo (see protocol section 6.6). Subjects who do not reach 300 mg/d will be considered escalation failures and non-responders for the primary analysis.

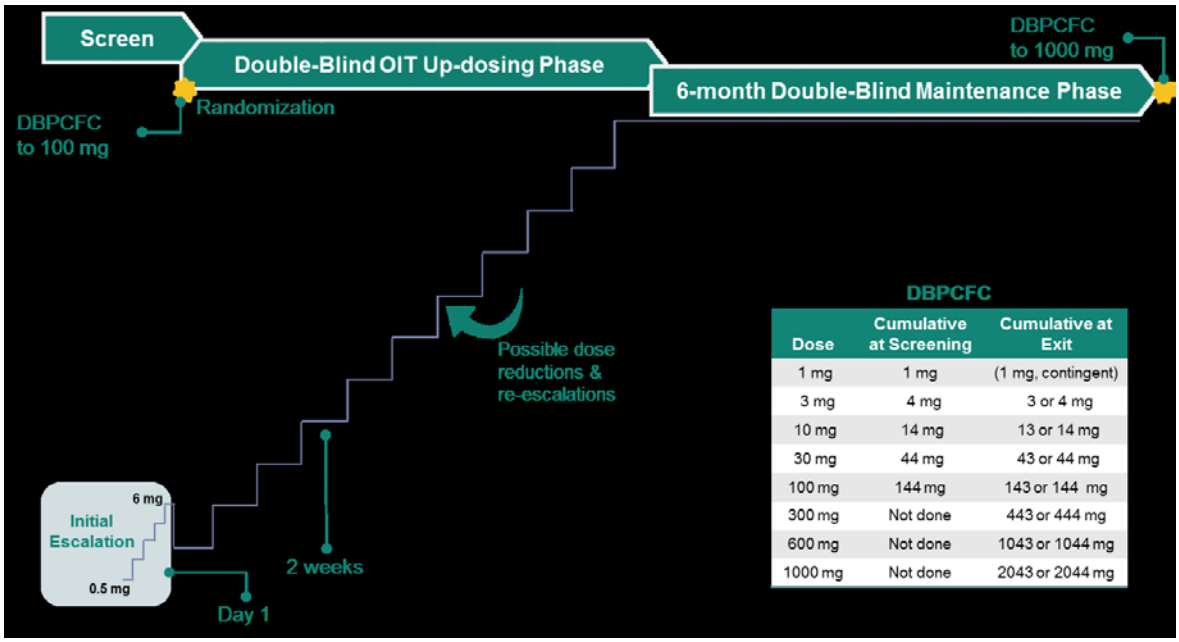
Each subject will be unblinded when he/she completes the DBPCFC at the end of the ~24-week Maintenance Period and all major data queries for the subject have been resolved.

Total duration of the study will be approximately 12 months (where the Treatment Phase ranges from 44 to 68 weeks).

Assessments will be performed according to the study schedule shown in Appendix 1 of the study protocol.

The study design information, treatment regimens and illustration of the study design are shown in Table 2.1 and the Figure below.

Figure 2.1 Illustration of study design



Ref. protocol: **Figure 3–1. Study Design**

Table 2.1 Study design information, treatment regimens

Phase of Development	3
Randomized Trial	Yes
Stratification Factors	Age Group: LT: Less than 18 years of age, GT: Equal to or more than 18 years of age Geographic Region: NA: North America, EU: Europe
No. Treatment Arms	2
No. of Subjects	500 planned per the protocol; actual enrollment is 555
Treatment Groups and Ratio	AR101 (peanut protein), placebo = 3:1

Investigational Product, Dosage, and Administration	AR101 (peanut protein) The initial dose escalation (first 2 days) includes doses of 0.5 mg up to 6 mg followed by 2 weeks at 3 mg. Future dose escalations will occur every 2 weeks with the escalation administered in the clinical research center. Doses will be escalated up to the subject's first in-clinic dose at 300 mg, ideally 20 weeks in duration, but may be extended to a maximum of 40 weeks to accommodate dose reductions and re-escalations, if necessary. The Maintenance Period starts with the first home dose of 300 mg. Ideally, it will be 24 weeks in duration, but it may be extended by up to an additional 4 weeks (for a maximum Maintenance Period duration of 28 weeks), or to a total Treatment Phase duration of 68 weeks, whichever occurs first, to accommodate dose reductions and re-escalations that may occur during the final weeks of the Maintenance Period.
Reference Therapy, Dose, and Administration or Comparative Intervention	Placebo: AR101 matching placebo on the same schedule

2.4. DETERMINATION OF SAMPLE SIZE

The sample size for the study, approximately 500 subjects randomized 3:1 (active:placebo), has been selected to provide sufficient power to detect a treatment effect for the primary efficacy analysis. As of the writing of Protocol Amendment 4 (31Jul2017), 555 subjects, 499 of which are between the ages of 4 and 17 years have been enrolled and enrollment has been completed.

The set of assumptions and calculations that follow are provided to show that this sample size is adequate to demonstrate the AR101 response rate is significantly higher than placebo with at least a 15% margin for the primary efficacy endpoint analysis in North America, and the sample size is adequate to demonstrate the AR101 response rate is significantly higher than placebo for the primary efficacy endpoint analysis in Europe. Additionally, in a trial of this size, there would be an 80% probability of observing at least one adverse event (AE) among 375 subjects assigned to AR101 when the background rate of the AE is 4.3 per 1,000 subjects.

While natural history of peanut allergy desensitization is not fully understood, significant short-term improvements in consumption amounts are believed to be uncommon. Nevertheless, the sensitivity thresholds to peanut allergen are known to vary day to day based on numerous intrinsic and extrinsic factors. The inherent variability in the sensitivity threshold for allergic reaction to peanut as measured by oral food challenge has been demonstrated in peanut-allergic subjects not undergoing immunotherapy (Glaumann et al., 2013), as well as in the placebo arms of several peanut immunotherapy trials (Sampson et al., 2011; Varshney et al., 2011; Fleischer et al., 2013; Sampson et al., 2015). The publication by Glaumann et al. (2013) showed that the threshold for responding in oral

food challenge can vary up or down by two orders of magnitude. The placebo response rates reported across therapeutic trials also vary widely, ranging anywhere from 11 to 55%, and are dependent on multiple factors, including the level of peanut protein set as a target response, the specific procedures for oral food challenge employed, the dose level of maintenance therapy, and the duration of immunotherapy.

The literature on peanut OIT, though sparse, also suggests that there could be a high degree of variability in the magnitude of the treatment effect. Only 2 randomized controlled trials of peanut OIT (Varshney et al., 2011; Anagnostou, et al., 2014) have been reported to date, and these employed significantly different study designs. In the Varshney study, 16 of 19 (84%) subjects randomized to active treatment demonstrated a high degree of desensitization (although baseline sensitivity to peanut had not been established). This stands in contrast to the Anagnostou study, wherein only 24 of 49 (49%) subjects randomized to active treatment achieved the predefined level of desensitization.

North America: In the Phase 2 ARC001 study the primary endpoint was determined at the 300 mg (443 mg cumulative) level of peanut protein in the Exit DBPCFC, whereas in the current Phase 3 study, ARC003, the primary endpoint will be determined at the 600 mg (1043 mg cumulative) dose level of peanut protein in the Exit DBPCFC, a level that was studied as an additional endpoint in ARC001. For the purpose of calculating statistical power for the ARC003 study, the response rates at both the 300 mg (primary) and 600 mg (additional) endpoints from ARC001 have been taken into consideration.

In the Phase 2 ARC001 study, the placebo response rate for the primary endpoint in the intent-to-treat (ITT) population of tolerating 300 mg (443 mg cumulative) of peanut protein at Exit DBPCFC was 19% (95% CI of 7% to 39%), and it was 0% (95% CI of 0% to 13%) in the post-hoc analysis of tolerating 600 mg (1043 mg cumulative) of peanut protein at Exit DBPCFC. Thus, the Phase 2 results suggest that the point estimate for the placebo response rate in the primary endpoint of ARC003 is likely to fall between 0 and 19% (based on point estimates), although it could be as high as 13 to 39% (based on upper bound of the 95% CIs). Taking into account the wide range of placebo response rates evident from the literature, a placebo response rate of 20% for the primary endpoint is assumed for power calculations in this study.

The response rate in the AR101 arm of the ARC001 study was 79% (95% CI of 60% to 92%) and 62% (95% CI of 42% to 79%) for endpoints based on tolerating 300 mg (443 mg cumulative) and 600 mg (1043 mg cumulative), respectively, at Exit DBPCFC in the ITT population. Thus, the Phase 2 results suggest that the point estimate for the active treatment response rate at the primary endpoint in ARC003 is likely to fall between 62 and 79% (based on point estimates), although it could be as low as 42 to 60% (based on lower bound of 95% CIs). Given that dropouts accounted for the majority of desensitization failures in ARC001, a 6-month study, and the opportunity for dropout increases as the duration of the trial increases, a 50% AR101 response rate at the primary endpoint is assumed for the purpose of power calculation for ARC003, a 12-month study. This approach is consistent with published recommendations for adjusting assumptions about Phase 3 response rates based on Phase 2 results (Wang et al., 2006; Kirby et al., 2014).

An additional 15% margin in the separation between the placebo response and the response to AR101 will be included in defining the success criteria for the primary endpoint. This will help to ensure that the number of peanut-allergic subjects spared moderate or severe allergic reactions at Exit DBPCFC (a model for accidental exposure) after treatment with AR101 represents a clinically meaningful benefit, and not just a statistically significant non-zero difference from placebo. The primary efficacy analysis is based on the Farrington-Manning test with a two-sided alternative hypothesis at the 5% level of significance (Farrington and Manning, 1990) and is conducted in the ITT population. Given placebo response rates of at most 20%, a 2-tailed 5% level test would have 89% power to detect an AR101 response rate of at least 50% and rule out treatment differences (AR101 minus placebo) that are 15% or less for the primary endpoint.

Europe: Based on preliminary data from ARC002, 40 out of 55 subjects treated in ARC001 and ARC002 with AR101 underwent an Exit DBPCFC with a single dose of 1000 mg (2043 mg cumulative) of peanut protein after approximately 12 or more weeks of maintenance dosing at 300 mg/d. Of these, 24 subjects tolerated a single dose of 1000 mg (2043 mg cumulative) of peanut with no more than mild symptoms. This corresponds to an AR101 response rate of 44% (95% CI of 30% to 58%). We assume a placebo response rate of 15% at a single dose of 1000 mg (2043 mg cumulative) of peanut protein (two-thirds of what we assumed for the placebo response rate at a single dose of 600 mg [1043 mg cumulative]). A sample size of 495 children, aged 4 to 17 years, provides at least 93% power to detect an AR101 response rate of at least 30% at a single dose of 1000 mg (2043 mg cumulative) of peanut protein when the placebo response rate is at most 15%.

2.5. TREATMENT ASSIGNMENT AND BLINDING CONSIDERATIONS

During the double blind study period, subjects are randomized in a 3:1 ratio to study treatment (AR101 or placebo) using the interactive voice or web-based response system (IXRS), with region (North America, Europe) and age group (children from 4 to 17, inclusive, and adults to age 55, inclusive) as stratification factors. A central randomization schedule of randomly permuted blocks with random block sizes was prepared by an independent unblinded statistician for use in the IXRS system.

This is a double-blind study. The study as a whole will not be unblinded until after the last subject exits ARC003 and the database is locked.

Based on ARC003 study design, the blind for the subject and investigator is maintained up to completion of the Exit DBPCFC. This will occur at different times for individual subjects. Consequently, decisions regarding rollover to the ARC004 follow-on study will need to be made at different times for different subjects. The outcome of Exit DBPCFC is the basis of the primary efficacy endpoints for ARC003. In accordance with the protocol, the blinded assessor is responsible for conducting the Exit DBPCFC. After completion of the Exit DBPCFC, the subject and investigator are, by virtue of the outcome of the DBPCFC, potentially unblinded. Following individual subject unblinding, site staff,

excluding the blinded assessor, will be unblinded to the subject's treatment in order to manage the subject's future treatment. Similar to site personnel, monitors responsible for monitoring ARC003 and ARC004 will require access to the subject's medical record in order to perform source data verification for both studies. Aimmune clinical project staff are responsible for overseeing the conduct and data quality of ARC003 study and as such will have access to ARC003 data throughout the study.

In order to maximize data quality and minimize bias, Aimmune will institute a policy (as described in the protocol) that no subject is to be formally unblinded until the known major data queries for that subject are addressed. Since subjects are expected to rollover to ARC004 on the same day as the ARC003 Exit DBPCFC, all factors critical to a subject exiting ARC003 and entry into ARC004 will be addressed prior to unblinding. The major data queries are those that have or may have bearing on the efficacy and/or safety aspects of the study are detailed in the ARC003 Treatment Masking Plan (Version 3.0, 23Nov2016). Subjects who discontinue the study prior to the Exit DBPCFC are not to be unblinded until database lock has occurred.

The potential for planned and unplanned unblinding exists in this study. Both situations are briefly described below. Further details are provided in the ARC003 Treatment Masking Plan and study protocol. For either type of unblinding, the date of unblinding, type (planned or unplanned), and person who requested the unblinding will be processed and documented in the IXRS. The actual treatment assignment will be made available only to the person requesting unblinding. Other study team personnel will receive notification that the treatment assignment has been provided, but will not be notified of the actual treatment assignment. If it is specifically necessary to provide a treatment assignment to the Sponsor, this information will be provided to an Unmasked Medical Monitor (e.g., ARC004 Medical Monitor) and/or Unmasked Clinical Operations Designee.

Masking procedures for outputs provided only to DSMC will be described in the DSMC Charter.

Unplanned Unblinding

Prior to the Exit DBPCFC assessment, a subject can be unblinded only when needed for making medical decisions regarding the care of a subject. The decision to unblind will be made in collaboration with the sponsor's Unmasked Medical Monitor. If a life-threatening event occurs, the subject should be treated as if the subject received active study product. For all unscheduled events that require unblinding, the investigator will contact the clinical monitor who will coordinate with the sponsor's representatives. Site personnel or other study team members (such as an Unmasked Medical Monitor) may request emergency unblinding as described above. If it is specifically necessary to provide a treatment assignment to the Sponsor, this information will be provided to an Unmasked Medical Monitor and/or Unmasked Clinical Operations Designee.

Any premature/unplanned unblinding requires a full written account by the site study physician of the event(s) that necessitated unblinding of the study medication for an individual participant. This account includes the reason(s) for unblinding, the name of the

sponsor's medical monitor who was notified of the unblinding, the names of the unblinded individual staff members and the date and time the unblinding occurred. The treatment assignment is confidential and should not be provided to blinded team members, as detailed in the ARC003 Treatment Masking Plan.

Planned Unblinding

End-of-study (planned) unblinding will be requested by the site for individual subjects. Subjects who discontinue the study prior to the Exit DBPCFC will remain blinded until study database lock. Otherwise, each subject will be unblinded after completing the DBPCFC and all known major data queries for the subject have been addressed.

ARC003 as a whole will remain blinded until after final database lock. However, most subjects in ARC003 are expected to be eligible to roll over to ARC004 prior to final database lock. ARC004 treatment assignments and schedules of events are based on treatment assignments and Exit DBPBFC results in ARC003. Therefore, the following additional measures will be implemented:

- Upon unblinding and rollover of subjects from ARC003 to ARC004, a new randomly generated subject identification number will be assigned by the IXRS to be used as the subject identification number for ARC004.

- Access to the link between ARC003 and ARC004 subject identification numbers will not be available to the Sponsor or Contract Research Organization (CRO) staff until after ARC003 study database lock.

- No member of the ARC003 project team (including data management and statistician) will have the ability to link subject identification data from ARC003 and ARC004. Subjects exiting ARC003 will be assigned a new subject ID upon entering ARC004. Only the ARC004 Statistician, clinical supply chain staff and relevant site staff, study monitors and IXRS vendor staff will have access to the key that links ARC003 and ARC004 subject identification numbers. This information will be kept strictly confidential until after the ARC003 study database has been locked and unblinded.

Refer to the ARC003 Treatment Masking Plan and ARC003 to ARC004 Rollover Procedures document for full details on treatment assignment and blinding procedures including those to be followed for emergency/unplanned and planned unblinding.

Database Lock and Study Unblinding

Once the database has been locked, randomized treatment assignments will be obtained from the IXRS vendor after obtaining proper authorization from Aimmune. These treatment assignments will then be incorporated into the analysis datasets, tables, listings, and figures.

2.6. ADMINISTRATION OF STUDY TREATMENT

The Initial Escalation period, Day 1, dosing schedule and Up-dosing period dosing schedule are shown in the following two tables:

Table 2.2 Initial Escalation Period, Day-1, Dosing Schedule

Initial Escalation Period, Day 1, Dosing Schedule		
Day 1 Dose #	**Study Product Dose (mg peanut protein or placebo)**	**Cumulative Study Product Dose (mg peanut protein or placebo)**
1	0.5	0.5
2	1	1.5
3	1.5	3
4	3	6
5	6	12
Doses will be delivered at 20 to 30 minute intervals. Subjects who are unable to tolerate a dose of 3 mg at the end of Day 1 will be considered escalation failures. All subjects who tolerate a dose of at least 3 mg on Day 1 will return on Day 2 to receive a single confirmatory 3 mg dose under direct observation. Subjects with either no symptoms or mild, non-dose-limiting symptoms after receiving 3 mg on Day 2 may start 2 weeks of daily dosing at 3 mg. Subjects who experience moderate or severe symptoms after receiving the 3 mg dose on Day 2 will be considered escalation failures. Future dose escalations will occur every 2 weeks with the initial dose increase administered in the clinical research center.		

Table 2.3 Up-dosing Period Dosing Schedule

Up-dosing Period Dosing Schedule			
Up-dosing Dose #	**Study Product Dose (mg peanut protein or placebo)**	**Interval (weeks)**	**% Increase**
1	3	2	
2	6	2	100%
3	12	2	100%
4	20	2	67%
5	40	2	100%
6	80	2	100%
7	120	2	50%
8	160	2	33%
9	200	2	25%
10	240	2	20%
11	300	Enter Maintenance Period	25%
Capsules and sachets (introduced during the Maintenance Period) are opened, contents sprinkled over an age-appropriate food, and mixed thoroughly. 300 mg capsules will be used for at least the first 2 weeks of dosing during the 24-Week Maintenance Period.			

All subjects who reach and tolerate 300 mg/d will continue at that dose level for the

duration of the Maintenance Period. The first Maintenance visit occurs 2 weeks after the last Up-Dosing visit, with visits every 4 weeks thereafter. For the Maintenance Period, 300 mg of peanut protein will be formulated in foil-laminate sachets. Matching placebo sachets will be used to maintain the double-blind. Any subject unable to achieve a dose of 300 mg/d of peanut protein by 40 weeks will be considered an escalation failure non-responder and will not undergo Exit DBPCFC.

2.7. DOUBLE BLIND PLACEBO CONTROLLED FOOD CHALLENGE

The DBPCFC at screening and exit will consist of administering gradually increasing challenge doses of peanut flour (containing ~50% peanut protein) or placebo (oat flour), mixed in a vehicle food, at 20 to 30 min intervals. The DBPCFC will be performed in accordance with Practical Issues in Allergology, Joint United States/European Union Initiative (PRACTALL) guidelines, but requiring progression in an unaltered sequence without repeating any dose. The DBPCFC is to be conducted as 2 challenges, each on a separate day, using a placebo for one challenge and peanut for the other, with similar artificial flavorings added to each as masking agents. The 2 challenge days should be scheduled as closely together as practicable and should not be scheduled more than 7 days apart. The oral food challenge is to be performed under double-blind conditions so that neither the subject, nor the subject's caregiver, nor any of the clinic staff (save for the unblinded preparer of the challenge foods) knows which challenge contains the peanut or the placebo. The clinic staff may not be unblinded as to the order of the two parts (peanut and placebo) of the DBPCFC until after completion of the observation period of the second part of the challenge.

For each subject, a "blinded" Evaluating Physician (Blinded Assessor) is to be designated to assess the tolerability of the challenge doses presented in the DBPCFC. The Blinded Evaluating Physician is not to be involved in the oversight of study product dosing (neither Initial Escalation, nor Up-dosing), nor the assessment or management of adverse events. To the extent practicable, the same Blinded Evaluating Physician who determines DLSs in the Screening DBPCFC should determine DLSs in the Exit DBPCFC.

The Screening and Exit DBPCFCs are based on a modified PRACTALL dosing regimen as illustrated in Table 2.4.

Table 2.4. Modified PRACTALL DBPCFC Doses Using Peanut Flour with 50% Peanut Protein Content for Screening and Exit DBPCFC

	Challenge Doses			
	Amount of Peanut Protein at Each Challenge Dose (mg)	**Amount of Peanut Flour with 50% Protein Content (mg)**	**Cumulative Amount of Peanut Protein (mg) at Screening**	**Cumulative Amount of Peanut Protein (mg) at Exit**
Screening only	1	2	1	0 (or 1)*
Screening and Exit	3	6	4	3 (or 4)
Screening and Exit	10	20	14	13 (or 14)
Screening and Exit	30	60	44	43 (or 44)
Screening and Exit	100	200	144	143 (or 144)
Exit only	300	600	-	443 (or 444)
Exit only	600	1200	-	1043 (or 1044)
Exit only	1000	2000	-	2043 (or 2044)

*For explanation of contingent/optional doses indicated in parentheses refer to protocol **Section 6.6.2**

3. ENDPOINTS

3.1. PRIMARY EFFICACY ENDPOINT

For North America, the primary efficacy endpoint is the proportion of subjects ages 4-17 who tolerate a single highest dose of at least 600 mg (1043 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC.

For Europe, the primary efficacy endpoint is the proportion of subjects ages 4-17 who tolerate a single highest dose of at least 1000 mg (2043 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC.

3.2. SECONDARY ENDPOINTS

3.2.1. Key Secondary Efficacy Endpoints

North America: The hierarchal order of key secondary efficacy endpoints is as follows:

1. The proportion of subjects ages 4-17 who tolerate a single highest dose of at least 300 mg (443 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC
2. The proportion of subjects ages 4-17 who tolerate a single highest dose of at least 1000 mg (2043 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC
3. The maximum severity of symptoms in subjects ages 4-17 occurring at any challenge dose of peanut protein during the Exit DBPCFC
4. The proportion of subjects aged 18 to 55 years who tolerate a single highest dose of at least 600 mg (1043 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC

Europe: The hierarchal order of key secondary efficacy endpoints is as follows:

1. The proportion of subjects ages 4-17 who tolerate a single highest dose of at least 600 mg (1043 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC
2. The proportion of subjects ages 4-17 who tolerate a single highest dose of at least 300 mg (443 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC
3. The maximum severity of symptoms in subjects ages 4-17 occurring at any challenge dose of peanut protein during the Exit DBPCFC

4. The proportion of subjects aged 18 to 55 years who tolerate a single highest dose of at least 1000 mg (2043 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC

3.2.2. Other Secondary Endpoints

The other secondary endpoints are as follows:

- Maximum dose achieved with no or mild symptoms at Exit DBPCFC in subjects ages 4-17
- Change from baseline in maximum tolerated dose (MTD) of peanut protein at DBPCFC in subjects ages 4-17
- Use of epinephrine as a rescue medication at Exit DBPCFC and comparison to its use at Screening DBPCFC in subjects ages 4-17
- Changes in peanut-specific (ps) serum IgE and IgG_4 levels in subjects ages 4-17
- Changes in peanut skin prick test (SPT) mean wheal diameters in subjects ages 4-17
- Quality of life assessment using the food allergy related quality of life questionnaire (FAQLQ), and the food allergy independent measure (FAIM) questionnaire (van der Velde et al., 2010) in subjects ages 4-17

3.3. SECONDARY SAFETY ENDPOINTS

Secondary safety endpoints are as follows:

- The safety of peanut OIT based on AEs including serious adverse events (SAEs) in the following 5 age groups: 4 to 17 years, 4 to 11 years, 12 to 17 years, 18 to 55 years, and 4 to 55 years, inclusive
- Use of epinephrine as a rescue medication during OIT (Initial Escalation, Up-dosing, and Maintenance Periods) in the following 5 age groups: 4 to 17 years, 4 to 11 years, 12 to 17 years, 18 to 55 years, and 4 to 55 years, inclusive
- Frequency of anaphylaxis during OIT (Initial Escalation, Up-dosing, and Maintenance Periods) in the following 5 age groups: 4 to 17 years, 4 to 11 years, 12 to 17 years, 18 to 55 years, and 4 to 55 years, inclusive
- Frequency of allergic reaction (hypersensitivity) AEs during OIT (Initial Escalation, Up-dosing, and Maintenance Periods)in the following 5 age groups: 4 to 17 years, 4 to 11 years, 12 to 17 years, 18 to 55 years, and 4 to 55 years, inclusive
- Frequency of accidental ingestions of peanut and other allergenic foods in the following 5 age groups: 4 to 17 years, 4 to 11 years, 12 to 17 years, 18 to 55 years, and 4 to 55 years, inclusive

- Severity of AEs associated with accidental ingestions of peanut and other allergenic foods in the following 5 age groups: 4 to 17 years, 4 to 11 years, 12 to 17 years, 18 to 55 years, and 4 to 55 years, inclusive
- Frequency of premature discontinuation of dosing due to AEs; and frequency of premature discontinuation of dosing due to chronic/recurrent gastrointestinal (GI) AEs in the following 5 age groups: 4 to 17 years, 4 to 11 years, 12 to 17 years, 18 to 55 years, and 4 to 55 years, inclusive
- Assessment of asthma control using the Asthma Control Test questionnaire in subjects with asthma in the following 5 age groups: 4 to 17 years, 4 to 11 years, 12 to 17 years, 18 to 55 years, and 4 to 55 years, inclusive

3.4. EXPLORATORY ENDPOINTS

- The primary endpoints identified above will be repeated in the following 3 age groups: 4 to 11 years, 12 to 17 years, and 4 to 55 years, inclusive
- The first 3 key secondary endpoints will be repeated in the following 4 age groups: 4 to 11 years, 12 to 17 years, 18 to 55 years, and 4 to 55 years, inclusive
- Treatment satisfaction assessment using the Treatment Satisfaction Questionnaire for Medication (TSQM-9) and an exit survey with palatability questions

4. ANALYSIS POPULATIONS

The following analysis populations will be defined for this study.

4.1. SAFETY POPULATION

The Safety population will consist of all subjects who receive at least one dose of randomized study treatment (i.e., not including food product administered during DBPCFC). The Safety population will be used for summaries of safety parameters. Subjects will be analyzed according to treatment received.

4.2. INTENT-TO-TREAT (ITT) POPULATION

The ITT population (i.e., the Full Analysis Set) will consist of all randomized subjects who received at least one dose of randomized study treatment. Subjects will be analyzed according to randomized treatment. The ITT population will be used as the primary analysis population for all analyses of efficacy endpoints. If no subjects received the incorrect treatment, the ITT population will be the same as the safety population.

Since only four subjects were randomized but not treated and due to the double blind nature of the trial, no bias is introduced by excluding these four subjects from the ITT population.

4.3. COMPLETER POPULATION

The Completer population includes all subjects in the ITT population who complete treatment and have an evaluable Exit DBPCFC, where an evaluable Exit DBPCFC is defined as completion of at least the peanut part of the food challenge. Sensitivity analyses and supportive analyses of the primary endpoint, and key secondary endpoints, and other secondary endpoints will be performed using the Completer population. These supportive analyses are considered important because they will provide the basis for informing patients and their families of their chances of achieving a clinically relevant level of desensitization if Up-dosing and maintenance therapy are achieved.

4.4. PER PROTOCOL (PP) POPULATION

The Per Protocol (PP) population will be a subset of the Completer population, limited to subjects who have no major protocol deviations that may influence the desensitization response. Additional criteria to exclude subjects from the PP population may be added. Any changes will be documented in a SAP amendment or other supporting document. Exclusions will be determined by blinded review before database lock and overall study unblinding. Subjects will be analyzed according to randomized treatment. The PP

population differs from the Completer population only in that it excludes subjects who may have undergone the Exit DBPCFC despite having major protocol deviations.

Analyses of the primary and all secondary efficacy endpoints will be performed on the PP population if the PP population differs from Completer population by >5% in either treatment arm. Sensitivity analyses of selected endpoints may, however, be performed if the PP population differs from the Completer population by ≤5% in both treatment arms.

The decision to conduct the optional analyses in the PP population will be made before database lock and overall study unblinding.

5. PROTOCOL DEVIATIONS

A Protocol Violation (i.e., major protocol deviation) is a deviation from the Institutional Review Board (IRB) approved protocol that may affect the subject's rights, safety, or well-being and/or the completeness, accuracy and reliability of the study data. In addition, protocol violations include willful or knowing breaches of human subject protection regulations, or policies, any action that is inconsistent with medical, and ethical principles, and a serious or continuing noncompliance with federal, state, local or institutional human subject protection regulations, policies, or procedures.

A non-major protocol deviation is any change, divergence, or departure from the study design or procedures of a research protocol that does not have a major impact on the subject's rights, safety or well-being, or the completeness, accuracy and reliability of the study data.

All protocol deviations will be reported in the data system on a specific case report form (CRF):

- Inclusion Criteria
- Exclusion Criteria
- Received incorrect study treatment
- Randomization issue/ randomized to wrong stratum
- ICF
- SAE Not Reported
- Visit Out of Window
- Missed Study Visit
- Procedure Not Per Protocol
- Prohibited Concominant Medication
- Lab Sample missed
- Study Drug Compliance
- Other (with free text field to record details)

These protocol deviations will be reviewed in a blinded fashion prior to database lock and their categorization as major or minor will be determined prior to database lock and will be used to help determine the PP population. All protocol deviations, both major and non-major, will be listed and included in the study report.

6. GENERAL ASPECTS FOR STATISTICAL ANALYSIS

6.1. STRATA AND COVARIATES

The primary efficacy endpoint of desensitization response rates in pediatric subjects will compare treatment groups based on Farrington-Manning confidence interval for the difference in binomial proportions without stratification.

The key secondary endpoint of maximum severity of symptoms in pediatric subjects at any challenge dose of peanut protein during the Exit DBPCFC will include a stratification factor for region (North America or Europe). Other secondary efficacy endpoints may include a stratification factor for region and/or a covariate for the corresponding baseline value, where appropriate.

Exploratory analyses of the primary efficacy endpoint and key secondary endpoints may be performed where specified baseline factors (see Section 9.6.3) are included as stratification factors in the analyses to explore their effects on study endpoints.

6.2. SUBGROUP ANALYSES

The primary analysis of efficacy and safety data will be performed in pediatric subjects ages 4-17. Additional analysis will be performed using the following subgroups of age, geographic region, or both age and geographic region.

Age Group:

- Ages 4-11 Years
- Ages 12-17 Years
- Ages 18-55 Years
- Ages 4-55 Years

Geographic Region:

- North America (NA)
- Europe (EU)

For countries with at least 3 subjects in each treatment arm, subgroup analyses may be performed by country.

In select analyses, age group and/or region will be included as factor(s) in the statistical analyses. When any region has fewer than 3 subjects with an evaluable response for the variable of interest in either treatment arm, region will be excluded from the model.

In addition to subgroup analyses described above, additional population subsets may be defined and utilized for select analyses. These subsets will include the following:

Baseline ps-IgE:

- < 100kU/L versus ≥100kU/L
- Cutoff predictive of screening MTD ≤30 mg (as determined from the Screened Population, see Section 9.6.6.)

Baseline SPT wheal diameter:

- < 8mm versus ≥ 8mm
- Cutoff predictive of screening MTD ≤30 mg (as determined from the Screened Population, see Section 9.6.6)

Additional subgroups for ps-IgE and SPT wheal diameter may be defined (e.g., those predictive of subjects who fail to reach the target dose of 300 mg, see Section 9.6.7) and evaluated in exploratory analyses. These will be described as such in the study report.

Other subgroups:

- Asthmatic subjects
- Subjects enrolled in protocol ARC004

6.3. MULTIPLE COMPARISONS AND MULTIPLICITY

The key secondary endpoints and other secondary efficacy endpoints will be tested in a hierarchical method, as described in Sections 9.3 and 9.5.

No other adjustments will be made for multiple comparisons.

6.4. SIGNIFICANCE LEVEL

Unless stated otherwise, all statistical tests will be two-sided, with a significance level of 0.05. Confidence intervals (CIs) will be calculated at the 95% level, reflecting a type I error rate of 0.05.

6.5. STATISTICAL NOTATION AND METHODOLOGY

Unless stated otherwise, the term “descriptive statistics” refers to the number of subjects (n), mean, median, standard deviation, minimum, and maximum for continuous variables and frequencies and percentages for categorical variables.

Unless specified otherwise, the denominator for percentages for categorical data will be based on the number of subjects or observations with non-missing data appropriate for

summary purposes. The denominator for percentages for incidence data (such as adverse events) will be based on the number of subjects in the analysis population "at risk".

Minimum and maximum values will be presented at the precision of the original value, means, medians will be rounded to 1 decimal place greater than the precision of the original value, standard deviations and standard errors will be rounded to 2 decimal places greater than the precision of the original value. Percentages will be rounded to 1 decimal place. Percentages that round down to 0 or up to 100% will be displayed as "<0.1%" and ">99.9%", respectively.

All summary tables will be presented by treatment group displayed as AR101 and Placebo. For disposition, demographic, and other summaries of baseline and history data, a Total column for both treatment groups combined will be included.

All relevant data collected in the database will be included in data listings and sorted by stratified age group, treatment group, subject number, test/measurement, and visit and time point as appropriate. The treatment group will be displayed in the same order as appeared in the summary tables.

Any change to the data analysis methods described in the protocol or SAP will require an amendment if it changes a principal feature of the protocol or SAP. Any other change to the data analysis methods described in the protocol or SAP, and the justification for making the change will be described in the clinical study report and where appropriate in a Note to File or amendment to the SAP before database lock. Additional exploratory analyses of the data will be conducted as deemed appropriate.

7. DATA HANDLING METHODS

7.1. VISIT WINDOWS

All information will be listed, summarized, and analyzed according to the nominal visit time point, study period, or dose. No visit windowing will be performed.

7.2. DATA PRESENTATION

Individual subject data listings will be provided to support summary tables and serve as a data source. Unless otherwise noted, all data collected during the study for all randomized subjects will be included in data listings. Listings will be sorted by age group (ages 4-17, ages 18-55), treatment group, subject identifier, and visit date or time point (where applicable). If a listing includes screen failures, they will be listed as treatment group 'NONE' and sorted after all randomized subjects.

Unscheduled visits will be listed but not included in by-visit summaries. Results from unscheduled visits may be used as baseline values, and for other derivations not tied to visit names (for example, unscheduled visits are included in the determination of worst post-baseline values for physical examination results).

7.3. MAXIMUM TOLERATED DOSE AT DBPCFC

The MTD for a DBPCFC is defined as the maximum single dose of peanut protein resulting in no more than mild symptoms and assessed by the investigator to have been tolerated (i.e., subject did not experience any dose-limiting symptoms). Any symptom requiring treatment is inherently dose-limiting; thus, a dose during a DBPCFC cannot be considered "tolerated" if treatment was deemed necessary by the investigator. The MTD at the Screening DBPCFC will be used as the baseline amount of peanut protein tolerated. If a subject is administered non-standard doses at a DBPCFC, the MTD will be considered as the highest standard dose (whether administered or not) less than the highest tolerated non-standard dose.

When describing the MTD at the Exit DBPCFC, in terms of the cumulative amount of peanut protein, the 1 mg dose will not be included. Thus subjects who tolerate all dose levels from 3 mg to 300 mg or 1 mg to 300 mg have a cumulative MTD of 443 mg.

7.4. CLASSIFICATION OF RESPONDER STATUS

7.4.1. Classification as a responder

To be a 600 mg responder, a subject must meet both of the following conditions:

1) Must have attained a maximum tolerated single dose (MTD) ≥600 mg of peanut protein on Exit DBPCFC
2) Must not have experienced more than mild symptoms through 600 mg of peanut protein on Exit DBPCFC.

Classification as a 300 mg responder and a 1000 mg responder are defined similarly.

7.4.2. Classification as a non-responder

If a subject cannot be classified as a responder, then that subject should be classified as a non-responder. If a subject meets any of the following conditions, then the subject is a non-responder:

1) Discontinued before the Exit DBPCFC or did not undergo an Exit DBPCFC for any other reason
2) Failed the Exit DBPCFC (e.g., a subject who did not tolerate a single dose of at least 600 mg of peanut protein with at most mild symptoms is a 600 mg non-responder).

Classification as a 300 mg non-responder and a 1000 mg non-responder are defined similarly.

7.4.3. Exit DBPCFC imputation rules for non-responders

As a general rule, subjects who do not undergo an Exit DBPCFC for any reason will be categorized as desensitization non-responders and their MTD at the Exit DBPCFC will be imputed using their MTD at the Screening DBPCFC.

7.4.4. Classification of Escalation and OIT Maintenance Failures

Non-responders who did not undergo the Exit DBPCFC are broadly categorized as either escalation failures or OIT maintenance failures. An escalation failure is a subject who is unable to achieve at least one at-home dose of 300 mg/d of peanut protein by 40 weeks (Protocol Section 3.2), discontinuing dosing prematurely either during the Initial Escalation period or during the Up-dosing period.

Subjects unable to tolerate a daily maintenance of 300 mg of study product for at least the last 2 weeks of dosing prior to Exit DBPCFC, free of any symptomatic therapy that was initiated during the course of OIT, will be considered OIT maintenance failures.

The 3 non-responder subcategories are summarized in Table 7.1.

Table 7.1 Non-Responder Subcategories for Subjects Who Do Not Undergo the Exit DBPCFC

Non-Responder Subcategory	Study Period Last Dose Was Taken	Definition of Period
Escalation Failure	Initial escalation	The last dose of study drug was taken during Initial Escalation Visit Day 1 or Day 2. No at-home doses were taken.
Escalation Failure	Up-dosing	The last dose of study drug was taken after Initial Escalation Visit Day 1 or Day 2 and before the first at-home dose of 300 mg.
OIT Maintenance Failure	Maintenance	At least one at-home dose of 300 mg was taken.

Classification of the primary reason a subject discontinued before the Exit DBPCFC provides a further sub-classification of non-responder subcategories, and will be determined during blinded data review prior to database lock and overall study unblinding.

The primary reason for discontinuation will be classified into one of the following categories:

- Discontinued due to reason related to treatment, including: escalation failure, OIT maintenance failure, treatment-related AE causing permanent discontinuation of study drug, or death
- Discontinued due to reason unrelated to treatment

7.5. DATA DERIVATIONS AND DEFINITIONS

The following definitions and derivations will be used throughout this study:

- Study Day is calculated as (assessment date – first dose date + 1) for assessments and visits performed on or after the first dose date, and (assessment date – first dose date) for assessments and visits prior to the first dose date.

- Baseline is defined as the last non-missing value prior to the first dose of randomized study treatment.

- Change from baseline is calculated as observed value after the first dose – baseline value.

- Screening period is defined as the time period beginning with the date and time of informed consent through the first dose of randomized study drug excluding the screening DBPCFC period. The last day of the screening period is the day before the first dose of randomized study treatment. Additional details on handling a DBPCFC given on non-consecutive days is provided below.

- Screening DBPCFC Period is the period of time starting with the first dose of DBPCFC product up through the first of:

    1. 24 hours after the last dose of DBPCFC product or

    2. first dose of randomized study drug

- Initial Escalation period is defined as the time period beginning with the date and time of the first dose of randomized study product in clinic and ending with the date of the last dose of randomized study product taken prior to Up-dosing.

- Up-dosing period is defined as the time period beginning with the date and time of the first home dose of study product at 3 mg, and ending with the date and time of first in-clinic dose at 300 mg. This period will be, ideally, 20 weeks in duration, but may be extended to a maximum of 40 weeks to accommodate dose reductions and re-escalations, if necessary.

- Maintenance period is defined as the time period beginning with the date and time of the first home dose of study product at 300 mg and ending with the date of the last dose of randomized study product taken prior to Exit DBPCFC. Ideally, this period will be 24 weeks in duration, but it may be extended to accommodate dose reductions and re-escalations by up to an additional 4 weeks (for a maximum Maintenance Period duration of 28 weeks), or to a total Treatment Phase duration of 68 weeks, whichever occurs first. For subjects who continue maintenance dosing after the Exit DBPCFC but prior to the rollover to ARC004, those data will be attributed to the maintenance period.

- Exit DBPCFC period is defined as the time period beginning with the date and time of the first DBPCFC dose after the Maintenance Period and through 24 hours after the last dose of DBPCFC product.

- The DBPCFC is given in 2 parts, either on the same day, on consecutive days, or occasionally on non-consecutive days. For the non-consecutive days, the period of time more than 24 hours after the first part and before the second part begins will be attributed as follows

    1. For the Screening DBPCFC this is attributed to the Screening Period

    2. For the Exit DBPCFC Period it is attributed to the maintenance period.

- The active treatment period is defined as the time period beginning with the date and time of the first dose of randomized study product and ending with the date and time of the last dose of randomized study product.

- Duration of active treatment period (days) for AR101 and placebo is calculated as the date of last dose minus the date of first dose plus 1, excluding the DBPCFC periods.

- The Exit DBPCFC is defined as indeterminate if the subject was not able to tolerate the placebo challenge up to and including a dose of 1000 mg (2043 mg cumulative).

7.6. MISSING DATA

All AEs with partial/missing dates and times will be considered Treatment Emergent Adverse Events (TEAEs) unless a partial date clearly indicates that it occurred prior to first dose of study treatment or more than 30 days after last dose of treatment. All therapies with partial or missing dates and times recorded on the Concomitant Medication or Non-Drug Therapy CRF pages will be considered concomitant unless a partial stop date and time clearly indicates it was stopped prior to the first dose of study treatment. Start and stop dates will be imputed when partial dates are present as needed to determine treatment emergent events and concomitant medications. No imputation will be done for a completely missing start/stop date or for subjects who did not receive study treatment.

Start dates with a missing day but which have month and year populated will be imputed such that:

- If the provided month and year match the month and year for that subject's first dose date, then the Day 1 date will be used
- In all other cases the 1st of the month will be used with the provided month and year

Start dates with a missing day and month but which have year populated will be imputed such that:

- If the provided year matches the year for that subject's first dose date, then the first dose date will be used
- In all other cases the 1st of January will be used with the provided year

Stop dates will be imputed as follows:

- Missing day with a provided year and month will use the last day of the month
- Missing day and month with provided year will use December 31

If the imputed stop date is greater than the last study date for the subject, then the imputed date will be replaced with the last known subject date.

The reported date of the most recent reaction to peanut on the peanut allergy history CRF page and date of diagnosis of peanut allergy will be imputed when the month or day is missing as follows:

- Missing day is set to 1 if the same year and month as the informed consent date. Otherwise it is set to 15
- Missing month and day are set to Jan 1 if the same year as the informed consent date. Otherwise it is set to July 1.

Where the severity score of a symptom is missing during screening or exit DBPCFC, the severity score will be imputed as severe.

For the primary and key secondary endpoints involving desensitization rates, if a subject discontinues prior to the exit DBPCFC, they will be considered non-responders. Other sensitivity analyses involving alternative methods for handling subjects with missing exit DBPCFC are described in section 9.2.

For the key secondary endpoint of maximum severity of symptoms, if a subject discontinues prior to the exit DBPCFC, the maximum severity of symptoms during the exit DBPCFC will be imputed using the maximum severity of symptoms during the screening DBPCFC.

No imputations will be made for other missing data, unless specified otherwise.

7.7. POOLING

Data pooling will be performed as necessary for exploratory analyses, as described in Section 9.6.2. No other data pooling is planned.

8. DEMOGRAPHIC, OTHER BASELINE CHARACTERISTICS

8.1. SUBJECT DISPOSITION AND WITHDRAWALS

The number and percentage of subjects screened, randomized, completed and discontinued, entered each study period, and in each analysis population will be summarized by age group (4-17 years, 18-55 years, 4-55 years) and treatment group (Placebo, AR101, and Total).

Reasons for discontinuation from the study will be summarized as well. In addition, subjects who discontinued study drug will be reviewed to determine if discontinuation was treatment-related and whether discontinuation was due to chronic/recurrent GI AEs. Classification of subjects will be determined during blinded data review prior to study unblinding.

Subject completion status, date of study completion/discontinuation, study treatment discontinuation, and reason for discontinuation will be listed based on the information collected in the case report form (CRF). Other information collected at screening and information regarding emergency unblinding will be included in the data listings as well.

Inclusion and exclusion eligibility will be listed separately.

8.2. PROTOCOL DEVIATIONS

All protocol deviations as defined in Section 5 will be listed by subject. Major protocol deviations (identified by blind data review before database lock) will be summarized by age group (4-17 years and 18-55 years) and treatment group (Placebo, AR101, and Total).

8.3. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Summary statistics for demographic and baseline characteristics will be provided for the Safety, ITT, Completer, and PP populations for subjects ages 4-17 and for the Safety, ITT, and Completer populations for subjects ages 4-11 years, 12-17 years, 18-55 years, and 4-55 years by treatment group (Placebo, AR101, and Total). A separate summary of demographics and baseline characteristics will be presented for screen failure subjects. Demographic data will include age, race, ethnicity, sex, body weight, height and BMI. Baseline characteristics include total IgE, ps-IgE, ps-IgG_4, ps-IgE/IgG_4 ratio, SPT mean wheal diameter at the 15 minute time point, MTD of peanut protein at Screening DBPCFC, and Childbearing Potential. History of asthma will also be included.

Age will be calculated relative to date of informed consent, as follows:

- If the month and day portion of the informed consent date is prior to the month and day portion of the birthdate, age will be calculated as the year of informed consent minus the year of birth, minus one;

- If the month and day portion of the informed consent date is on or after the month and day portion of the birthdate, age will be calculated as the year of informed consent minus the year of birth.

Demographic and baseline characteristic informationwill be listed for all randomized subjects.

8.4. ALLERGY HISTORY

Allergy history will be listed by subject.

The duration of peanut allergy (months since peanut allergy diagnosis), number of anaphylactic reactions to peanut in lifetime, number of anaphylactic reactions to peanut per year during lifetime, months since most recent anaphylactic reaction to peanut, months since most recent reaction to peanut that was treated with a therapy, the type of therapy administered for the most recent anaphylactic reaction to peanut, and the symptoms experienced during the most recent peanut exposure will be summarized by age group (4-17 years, 18-55 years, and 4-55 years) and treatment group (Placebo, AR101, and Total).

The reported date of the most recent reaction to peanut and date of diagnosis of peanut allergy will be imputed based on the logic in section 7.6.

8.5. NON-PEANUT ALLERGY HISTORY

All non-peanut allergy history will be listed by subject. The presence of non-peanut allergy history and causative allergens will also be summarized by age group (4-17 years, 18-55 years, and 4-55 years) and treatment group (Placebo, AR101, and Total).

8.6. OTHER MEDICAL HISTORY

Medical history will be listed by subject and body system. Subjects experience abnormal medical history events will be summarized by age group (4-17 years, 18-55 years, and 4-55 years), treatment group (Placebo, AR101, and Total) and by MedDRA system organ class and preferred term.

9. EFFICACY

9.1. PRIMARY ENDPOINT

The primary endpoint for North America is a desensitization response as determined by tolerating a single highest dose of at least 600 mg of peanut protein at the Exit DBPCFC with no more than mild symptoms. The primary endpoint for Europe is a desensitization response as determined by tolerating a single highest dose of at least 1000 mg of peanut protein at the Exit DBPCFC with no more than mild symptoms. Subjects who do not have an Exit DBPCFC are non-responders.

The number and percent of subjects with a desensitization response will be reported by treatment group. The desensitization response rate and its 95% confidence interval will be calculated for each treatment group with Wilson (score) confidence limits for the binomial proportion. The 95% confidence interval for the treatment difference (AR101 desensitization rate minus placebo desensitization rate) will be based on Farrington-Manning CI for the difference in binomial proportions. The number and percent of subjects at each dose level for highest tolerated dose at the Exit DBPCFC will also be summarized by treatment group.

The primary efficacy analysis of the primary endpoint will be evaluated in subjects ages 4-17 and will test for the difference in response rates (AR101 minus Placebo) in the ITT Population. For North America, the Farrington-Manning test will be used to test the null hypothesis that the difference is equal to 0.15 at the 0.05 significance level. AR101 is considered to have met primary efficacy endpoint if the lower bound of the corresponding test-based 95% confidence interval is greater than the pre-specified margin of 0.15.

This corresponds to the following hypotheses:

- H_0: Proportion of active subjects with a desensitization response – proportion of placebo subjects with a desensitization response = 0.15

- H_A: Proportion of active subjects with a desensitization response – proportion of placebo subjects with a desensitization response > 0.15 or < 0.15

For Europe, the Farrington-Manning test will be used to test the null hypothesis that the difference is equal to 0 at the 0.05 significance level. AR101 is considered to have met primary efficacy endpoint if the lower bound of the corresponding test-based 95% confidence interval is greater than 0.

9.2. ADDITIONAL ANALYSES OF THE PRIMARY EFFICACY ENDPOINT

As a sensitivity analysis in order to determine the impact of missing data on the robustness of the study results, the primary efficacy endpoint will be analyzed using a worst case approach to missing data imputation. Placebo subjects who have missing data (i.e., do not have an Exit DBPCFC) for the primary efficacy endpoint for any reason will be considered as responders while AR101 subjects will be considered as non-responders if they have missing data for the endpoint. Analytical methods will follow those described above in Section 9.1.

If the primary analysis of the primary endpoint shows a statistically significant treatment effect, a tipping point analysis will be conducted to identify the point at which the number placebo subjects with missing data are imputed as responders will make the treatment effect non-significant. The first step will be to test the treatment effect when all AR101 subjects with missing data are imputed as non-responders and only one placebo subject with missing data is imputed as a responder while the remaining placebo subjects with missing data are imputed as non-responders. If that test shows a statistically significant treatment effect the number of placebo subjects who have missing data that are imputed as responders will be increased by one until the treatment effect becomes non-significant.

The primary efficacy analysis will be repeated in the Completer population as a sensitivity analysis and also in the PP population (if sufficiently different from the Completer population; section 4.7). Additionally, subjects with an indeterminate Exit DBPCFC as defined in section 7.5 will be excluded for a sensitivity analysis of the primary efficacy analysis in the ITT population.

The primary efficacy endpoint will be analyzed in the following subgroups in the ITT and Completer populations as supportive analyses to the primary efficacy analysis using the methods described in Section 9.3.1:

- By Region (NA and EU) in pediatric subjects
- By Pediatric Age group (Children 4-11 and Adolescents 12-17)
- By Region and Pediatric Age group (NA Children, NA Adolescents, EU Children, and EU Adolescents)

If each region stratum has at least 3 placebo subjects, then a sensitivity analysis will be conducted on the common risk difference of the response rates (AR101 minus Placebo) in the ITT Population, stratified by Region. The risk difference will be estimated using the Mantel-Haenszel estimate of the common risk difference. A 95% confidence interval will be calculated using stratified Newcombe confidence limits. The common risk difference assumption will be assessed by visual comparison of the 95% Farrington-Manning CIs across the 2 strata and by a 1 degree of freedom Wald statistic calculated to test for unequal risk differences across the 2 strata.

Similarly, if each pediatric age group stratum has at least 3 placebo subjects, then a sensitivity analysis will be conducted on the common risk difference of the response rates (AR101 minus Placebo) in the ITT Population, stratified by Pediatric Age group.

An additional similar sensitivity analysis will be conducted on the common risk difference of the response rates (AR101 minus Placebo) in the ITT Population, stratified by pediatric age group and region, if each region and pediatric age group combination has at least 3 placebo subjects.

9.3. KEY SECONDARY EFFICACY ENDPOINTS

If the primary efficacy analysis is significant at the 0.05 level, then analysis of the treatment effect for key secondary efficacy endpoints will be tested in the following hierarchical order:

North America:

1. Response at a single dose of 300 mg (443 mg cumulative) of peanut protein: The proportion of subjects ages 4-17 who tolerate a single dose of 300 mg of peanut protein with no more than mild symptoms at the Exit DBPCFC
2. Response at a single dose of 1000 mg (2043 mg cumulative) of peanut protein: The proportion of subjects ages 4-17 who tolerate a single dose of 1000 mg of peanut protein with no more than mild symptoms at the Exit DBPCFC
3. The maximum severity of symptoms in subjects ages 4-17 occurring at any challenge dose of peanut protein during the Exit DBPCFC
4. Response at a single dose of 600 mg (1043 mg cumulative) of peanut protein: The proportion of subjects ages 18-55 who tolerate a single dose of 300 mg of peanut protein with no more than mild symptoms at the Exit DBPCFC

Europe:

1. Response at a single dose of 600 mg (1043 mg cumulative) of peanut protein: The proportion of subjects ages 4-17 who tolerate a single dose of 600 mg of peanut protein with no more than mild symptoms at the Exit DBPCFC
2. Response at a single dose of 300 mg (443 mg cumulative) of peanut protein: The proportion of subjects ages 4-17 who tolerate a single dose of 300 mg of peanut protein with no more than mild symptoms at the Exit DBPCFC
3. The maximum severity of symptoms in subjects ages 4-17 occurring at any challenge dose of peanut protein during the Exit DBPCFC
4. Response at a single dose of 1000 mg (2043 mg cumulative) of peanut protein: The proportion of subjects ages 18-55 who tolerate a single dose of 1000 mg of peanut protein with no more than mild symptoms at the Exit DBPCFC

Each comparison will be evaluated for statistical significance (two-sided $p<0.05$) only if all of the preceding tests in the hierarchy and the primary analysis of the primary endpoint are statistically significant in favor of AR101. Similar to the primary efficacy analysis of the primary endpoint, the ITT Population will be used as the primary population in pediatric subjects for analysis of key secondary endpoints as part of the hierarchical testing procedure. This closed testing procedure maintains the overall Type I error rate at 0.05 (EMEA CPMP, 2002; Cook et al., 2008). If any of the preceding tests are not significant, the p-value will be displayed for informational purposes only.

9.3.1. Desensitization Response Rates

The number and percent of subjects with a desensitization response will be reported by treatment group within pediatric subjects. The desensitization response rate and its 95% confidence interval will be calculated for each treatment group with Wilson (score) confidence limits for the binomial proportion. The 95% confidence interval for the treatment difference (AR101 desensitization rate minus placebo desensitization rate) will be based on Farrington-Manning CI for the difference in binomial proportions.

The analysis will test for the difference in response rates (AR101 minus Placebo) in the ITT Population. The Farrington-Manning test will be used to test for a non-zero treatment effect at the two-sided 0.05 significance level. AR101 is considered to have met this key secondary efficacy endpoint if the lower bound of the corresponding test-based 95% confidence interval is greater than 0.

This corresponds to the following hypotheses:

- H_0: Proportion of active subjects with a desensitization response ≤ proportion of placebo subjects with a desensitization response

- H_A: Proportion of active subjects with a desensitization response > proportion of placebo subjects with a desensitization response

9.3.2. Maximum Severity of Symptoms

The objective of analyzing this key secondary efficacy endpoint is to determine if pediatric subjects from the AR101 group will have less chance of developing more severe levels of symptom severity compared to pediatric subjects from the placebo group. Symptom severity will be determined at 4 levels: 0-None, 1-Mild, 2-Moderate, 3-Severe or higher (severe, life threatening, fatal). Subjects who experience no symptoms will be assigned a severity of 0-None. Symptom severity data will be collected at each challenge dose of peanut protein during the Exit DBPCFC (3 mg, 10 mg, 30 mg, 100 mg, 300 mg, 600 mg, and 1000 mg). The analysis of this key secondary endpoint will be conducted in the pediatric subject subset of the ITT population.

The maximum severity of symptoms observed in the DBPCFC at any dose (1000 mg or lower) will be used for each subject in the primary analysis of this key secondary endpoint.

The number and percent of subjects by maximum severity at the Exit DBPCFC will be tabulated by treatment arm. The Cochran-Mantel-Haenszel statistic (with equally spaced scores) will be used to test for a treatment difference. The test will be stratified by Region. The mean of the numerical values assigned to maximum severity by treatment group will also be reported.

The primary analysis of this key secondary endpoint will test the following hypotheses:

- H_0: The maximum severity of symptoms at the Exit DBPCFC at any dose is the same for the AR101 and placebo treatment arms in all strata.

- H_A: The maximum severity of symptoms at the Exit DBPCFC at any dose is different between the AR101 and placebo treatment arms in at least one stratum.

For the primary analysis of this key secondary endpoint, subjects without an Exit DBPCFC will be assigned their maximum severity during the Screening DBPCFC, which equates to no change from screening.

9.4. ADDITIONAL ANALYSIS OF KEY SECONDARY ENDPOINTS

Analysis of the key secondary endpoints will be repeated in the pediatric subset of the Completer population as a sensitivity analysis and also in the PP population (if sufficiently different from the Completer population; section 4.7). Additionally, subjects with an indeterminate Exit DBPCFC will be excluded as a sensitivity analysis of the key secondary efficacy endpoints in the ITT subset.

Analysis of the key secondary efficacy endpoints will be repeated in the following subgroups in the ITT and Completer populations as supportive analyses:

- By Region (NA and EU)
- By Pediatric Age group (Children 4-11 and Adolescents 12-17)
- By Region and Pediatric Age group (NA Children, NA Adolescents, EU Children, and EU Adolescents)

For countries with at least 3 subjects in each treatment arm, subgroup analyses may be performed by country.
A sensitivity analysis will be conducted on the common risk difference of the desensitization response rates similar to the sensitivity analysis of the primary efficacy endpoint (Section 9.2). This sensitivity analysis will be performed in the ITT and Completer populations.

As supportive analyses, the analysis of maximum severity of symptoms during the Exit DBPCFC will be repeated for the maximum severity observed in the DBPCFC at any dose up to a maximum challenge of 600 mg. The analysis will be repeated up through a maximum challenge of 300 mg, 100 mg, 30 mg, 10 mg, and 3 mg. Together with the primary analysis of this key secondary endpoint in the ITT population of pediatric subjects, this family of tests will be conducted hierarchically to maintain a familywise Type I error rate of 5%. Each test within this family will be conducted at the 5% level if the preceding test was significant at the 5% level; starting with the primary analysis of this key secondary endpoint (maximum severity at any dose up through and including 1000 mg) and followed by each lower level (600 mg, 300 mg, 100 mg, 30 mg, 10 mg, and 3 mg), until a test is not significant at the 5% level. Analytical methods described in Section 9.3.2 will be used for these analyses.

A sensitivity analysis for the maximum severity of symptoms in the ITT population of pediatric subjects using multiple imputation (MI) methodology may be explored to assess the impact of missing symptom severity data. We assume subjects with a missing exit DBPCFC tend to have more severe symptoms than Completers and thus are missing not at random (MNAR). The data to be used for the MI sampling will be chosen from subjects in the placebo group in sequential order of severity starting with the most severe symptoms. The percentage of placebo subjects with non-missing data to be used in the MI sample will be based on the percentage of AR101 subjects with missing data at the Exit DBPCFC. However, no fewer than 15 and no more than 30 placebo subjects with non-missing data will be used for the MI sample. The number of imputations will be set to the percentage of subjects across both treatment groups with missing data at the Exit DBPCFC but no fewer than 20 imputations will be used. For example, if 25% of AR101 subjects have missing Exit DBPCFC, then the worst 25% of observed Exit DBPCFC responses from the Placebo group will be used as the MI sample for missing AR101 and Placebo responses and a total of 25 imputations would be used.

These additional analyses will not be considered as part of the hierarchical testing procedure.

Subgroup analyses based on region will employ similar analytical methods as the non-subgroup analyses with the exception of not including region as a factor in the model or as a stratification factor.

9.5. ANALYSIS OF OTHER SECONDARY ENDPOINTS

If the primary efficacy analysis of the primary endpoint and all of the hierarchical testing of key efficacy secondary endpoints in pediatric subjects as described in Section 9.3 are found to be statistically significant, then statistical testing of the other secondary efficacy endpoints will continue in the order that they are listed in Section 3.2.2 according to the same hierarchical closed testing procedure used for the key efficacy secondary endpoints.

The ITT population of pediatric subjects will similarly be used for the hierarchical testing of the other secondary endpoints.

The other secondary endpoints are included in the overall hierarchical testing procedure. However, these endpoints are considered supportive in nature to the primary and key secondary endpoints. As such, no labeling claims will be made based on these endpoints if statistical significance is achieved, regardless of the success of the variables above them in the hierarchy.

Analysis of other secondary endpoints, as described below, will be performed on the pediatric subject subset (Ages 4-17) of the ITT and Completer populations. Analysis in the Completer population will not be considered for the hierarchical testing procedure.

9.5.1. Maximum Tolerated Dose at Screening and Exit DBPCFCs

The Screening and Exit DBPCFCs are based on a modified PRACTALL dosing regimen as described in section 2.7 (see Table 2.4). With the exception of the 600 mg dose, the modified PRACTALL doses are approximately on a logarithmic scale.

Estimates for the probability of tolerating each challenge dose or higher in the Exit DBPCFC will be based on the discrete hazards model with terms for treatment group effect, region (NA, EU), and the MTD at the Screening DBPCFC (baseline) in the $\log_{10}$ scale (Chinchilli et al., 2005). Following Chinchilli, the extreme value hazard function will be used for the discrete-time hazard function and the model will be fit with logistic regression with the complementary log-log link function. Subjects with no dose eliciting response at the 1000 mg single dose will be censored at that dose.

The probability estimates for each dose level will be tabulated by treatment group based on LS Mean estimates from the above model (i.e., adjusted for the MTD at baseline and region). The adjusted probability estimates will also be plotted. An unadjusted probability estimate will also be calculated by removing the MTD at baseline and region terms from the model. All subjects in the analysis population are considered eligible for the 1 mg dose. If the optional 1 mg dose was not taken, the subject is considered to have passed it.

The treatment group effect will be assessed using the model with terms to adjust for the MTD at baseline and region. The hazard ratio for the treatment group effect with its 95% confidence interval and the p-value will be based on the Wald statistic. If a subject did not participate in the Exit DBPCFC, the screening results will be used.

The hazard ratio is an estimate of the ratio of the conditional probability of not tolerating a single dose of the DBPCFC given the subject tolerated the lower doses for AR101 subjects relative to Placebo subjects. The comparison of treatment groups using the discrete hazards model corresponds to the following hypothesis:

- H_0: The hazard ratio for subjects within active subjects relative to placebo subjects = 1 when controlling for baseline MTD and region

- H_A: The hazard ratio for subjects within active subjects relative to placebo subjects ≠ when controlling for baseline MTD and region

The proportional hazard model assumption for treatment effect will be checked graphically with a log-log plot of Kaplan-Meier estimates for each treatment arm by dose level. Overall validity of model assumptions will be checked by visual comparison of Kaplan-Meier estimates with estimates from the unadjusted and adjusted models.

The peanut MTD at Screening and Exit DBPCFCs will be listed. Imputed peanut MTD at the Exit DBPCFC will be flagged.

Analyses will be performed in the ITT and Completer populations.

9.5.2. Change from Baseline in MTD at DBPCFC

Frequencies and percent of subjects will be presented for each MTD level at the Screening and Exit DBPCFC and for all possible ratios (X-fold increase) of the MTD at the Exit DBPCFC relative to the Screening DBPCFC. In order to calculate fold, if subjects do not tolerate any dose level they will be assigned an MTD of 0.3 mg.

Analyses of change from baseline MTD will be performed using change calculated on the $\log_{10}$ scale. Summary statistics of the change in MTD in the $\log_{10}$ dose scale will be presented. As above, if subjects do not tolerate any dose level they will be assigned an MTD of 0.3 mg prior to converting to the $\log_{10}$ scale. An analysis of covariance (ANCOVA) model of change from screening MTD at Exit DBPCFC ($\log_{10}$ mg) will be fit with terms for treatment group, region, and MTD at screening ($\log_{10}$ mg). The values for MTD for subjects who do not undergo the Exit DBPCFC will be imputed using the MTD from their Screening DBPCFC.

The comparison of treatment groups using the ANCOVA model corresponds to the following hypotheses:

- H_0: Mean change from baseline within active subjects = mean change from baseline within placebo subjects when controlling for baseline and region

- H_A: Mean change from baseline within active subjects ≠ mean change from baseline within placebo subjects when controlling for baseline and region

The p-value is based on the F-test for treatment group effect adjusted for the MTD at screening ($\log_{10}$ mg) and region. The ANCOVA model will also be assessed for unequal slopes by adding a term for the treatment group by baseline interaction and using an F-test

to test for unequal slopes. If there is a signal of unequal slopes at the 0.10 significance level, then the least squares means with 95% confidence interval, and p-value for the F-test of the treatment group difference will be calculated and reported at screening MTDs of 1 mg, 3 mg, 10 mg, and 30 mg. P-values and confidence intervals are based on the normality assumption.

Model residuals will be calculated and assessed for non-normality using the Shapiro-Wilk test and graphically to check the model assumptions of homoscedasticity and normality. If the model assumptions of homoscedasticity and normality are not met, then the Wilcoxon rank-sum test, adjusted for region will be used to test for a treatment group difference to examine the robustness of the ANCOVA F-test. Further exploration may be warranted if unequal slopes are noted.

Least squares mean statistics (point estimate and confidence interval) from the ANCOVA analysis of change from baseline analysis in the $\log_{10}$ scale will be transformed back to the original scale to obtain geometric least squares mean statistics of the ratio of AR101 to Placebo. Similarly, geometric least squares mean statistics of the ratio of the MTD at the Exit DBPCFC to Screening DBPCFC will be obtained and presented by treatment group.

Analyses will be performed in the ITT and Completer populations.

9.5.3. Use of Epinephrine as a Rescue Medication at Exit DBPCFC

The number and percent of pediatric subjects using Epinephrine as a rescue medication at the Exit DBPCFC will be summarized by type of challenge (peanut or placebo) and treatment group. Fisher's exact test will be performed to test for a treatment difference for each type of challenge at the Exit DBPCFC.

This corresponds to the following hypotheses and the analysis at the Exit DBPCFC in pediatric subset of the Completer population will constitute the primary analysis of this secondary endpoint:

- H_0: Percentage of active subjects with epinephrine use as a rescue medication at the Exit DBPCFC = percentage of placebo subjects with epinephrine use as a rescue medication at the Exit DBPCFC

- H_A: Percentage of active subjects with epinephrine use as a rescue medication at the Exit DBPCFC ≠ percentage of placebo subjects with epinephrine use as a rescue medication at the Exit DBPCFC

The number and percent of pediatric subjects using Epinephrine as a rescue medication at the Exit peanut DBPCFC will be summarized by Epinephrine use at the Screening peanut DBPCFC and treatment group. A logistic regression model will be fit to assess the relationship between epinephrine use at the screening DBPCFC with epinephrine use at the

Exit DBPCFC. Epinephrine use at the Exit DBPCFC will be modelled with terms for treatment, epinephrine use at the Screening DBPCFC, and the interaction between treatment and epinephrine use at the Screening DBPCFC. The odds ratio for epinephrine use at the Exit DBPCFC with 95% CIs of the treatment effect (AR101 versus Placebo) by epinephrine use at Screening will be calculated. Similarly, the odds ratio for epinephrine use at the Exit DBPCFC with 95% CIs of epinephrine use (Yes versus No) at the Screening DBPCFC by treatment will be calculated. The score statistic will be used to calculate a p-value for each of the model terms.

Summary of epinephrine use as a rescue medication at the Exit peanut DBPCFC (yes or no) will also be summarized by each dose level (up through 1000 mg, up through 600 mg, up through 300 mg, etc.). Treatment groups will be compared using a logistic regression model with terms for treatment, epinephrine use at the Screening DBPCFC, and the interaction between treatment and epinephrine use at the Screening DBPCFC, when the dose level under examination has at least 5% of subjects with epinephrine use.

The number of epinephrine doses as a rescue medication during the Exit DBPCFC will be summarized by dose level (up through 1000 mg, up through 600 mg, up through 300 mg, etc.). Doses will be categorized as 0 doses, 1 dose, 2 doses, and 3 or more doses. Treatment groups will be compared using Cochran-Mantel-Haenszel tests (with equally spaced scores).

Listings of all Epinephrine use will be provided, sorted by time Screening DBPCFC and DBPCFC, then by treatment group, and subject ID.

Analyses will be performed in the Completer population.

9.5.4. Peanut-specific IgE and IgG_4

Blood samples to measure ps-IgE, ps-IgG_4 levels, and total IgE levels will be collected prior to the Screening DBPCFC and prior to the Exit DBPCFC. Ps-IgE/IgG_4 ratio will be calculated, listed by subject, and summarized by visit and treatment group. Results outside the limits of quantification will be displayed as less than the lower limit of quantification (LLOQ), or greater than the upper limit of quantification (ULOQ), as appropriate. These values will be summarized as either the LLOQ or the ULOQ. If the ps-IgE or ps-IgG_4 is outside of the limits of quantification, the ps-IgE/IgG_4 ratio will be calculated using the LLOQ or ULOQ as appropriate.

Summary statistics, including geometric means and geometric standard deviations, will be presented by time point and treatment group.

Analyses of change from baseline ps-IgE, ps-IgG_4, and ps-IgE/IgG_4 ratio will be performed using change calculated on the $\log_{10}$ scale. An ANCOVA model of change from Screening DBPCFC to Exit DBPCFC will be fit with terms for treatment group, region,

and immunoglobulin value at Screening DBPCFC on the $\log_{10}$ scale. Similar analysis methods as described for the change from baseline in MTD at DBPCFC will be used. Geometric least squares mean statistics for the ratio of AR101 to Placebo and geometric least squares mean statistics for the ratio of the Exit DBPCFC to Screening DBPCFC by treatment group will be presented.

The comparison of treatment groups using the ANCOVA model corresponds to the following hypotheses:

- H_0: Mean change from baseline within active subjects = mean change from baseline within placebo subjects when controlling for baseline and region

- H_A: Mean change from baseline within active subjects ≠ mean change from baseline within placebo subjects when controlling for baseline and region

Analyses will be performed in the ITT and Completer populations.

9.5.5. Peanut Skin Prick Test

Results from the SPT will be listed, including test date, timing of test reading (e.g., 10 minutes), and measurements of the mean wheal diameter (in mm) of the following: peanut wheal (long axis), peanut erythema/flare (short axis), saline wheal (long axis), saline-glycerin erythema/flare (short axis), histamine wheal (long axis), and histamine erythema/flare (short axis).

A derived mean wheal diameter score will be calculated as the average of the long and short axis from the peanut wheal minus the average of the long and short axis from the saline wheal. Summary statistics for the derived SPT mean wheal diameter will be presented at each visit by treatment group for the 15-minute time interval and change from baseline to Exit visit will be presented.

An ANCOVA model of change from baseline at Exit DBPCFC will be fit with terms for treatment group, region, and wheal diameter at baseline. Similar analysis methods as described for the change from baseline in MTD at DBPCFC will be used. Least squares mean change estimates along with 95% confidence intervals will be presented by treatment group and the least squares mean difference (AR101 minus placebo) with corresponding 95% confidence interval will be presented.

The comparison of treatment groups using the ANCOVA model corresponds to the following hypotheses:

- H_0: Mean change from baseline within active subjects = mean change from baseline within placebo subjects when controlling for baseline and region

- H_A: Mean change from baseline within active subjects ≠ mean change from baseline within placebo subjects when controlling for baseline and region

Analyses will be performed in the ITT and Completer populations.

9.5.6. Quality of Life Assessments

Quality of life assessment using the food allergy related quality of life questionnaire (FAQLQ), and the food allergy independent measure (FAIM) questionnaires will be performed at the Baseline and Exit visits.

Separate FAQLQ and FAIM instruments are administered based on the subject's age group. Caregiver/parent versions are also administered for subjects who are 17 years old or younger. Due to differences between the various instruments, separate summaries will be provided by age group and person who completed the questionnaire (i.e., subject or caregiver).

FAQLQ:

The FAQLQ is a self-report instrument that is intended to assess the effect of food allergy on the subject's quality of life. Evaluations are done by the subject (for subjects 13 years and older) and parent (for subjects 17 years and younger).

The number of items and domains varies by instrument administered. Each question is scored on a seven-point scale (van der Velde et al., 2010). The total score is the arithmetic average of all non-missing items. Domain scores are calculated similarly. The total score ranges from 1 (limited severity perception) to 7 (greatest severity perception). If less than 80% of the items are complete, then the total score will not be calculated. Similarly, if less than 80% of the items within a domain are complete, then the domain score will not be calculated.

Descriptive statistics and scores of total and domain scores along with their changes from baseline will be provided. Data will be summarized separately by age group and responder (subject or caregiver).

Analyses of change from baseline in the total and domain scores will be performed using ANCOVA models of change from baseline at the Exit visit. The model will be fit with terms for treatment group, region, and screening value. Region will be excluded from the model if any region has fewer than 3 subjects with evaluable scores within either treatment group.

Listing of the raw scores as recorded in the CRF will be provided, sorted by treatment group and subject ID.

FAIM:

The FAIM is a self-report instrument that is intended to reflect the perception of food allergy severity and related risk as evaluated by the subject (for subjects 13 years and older) and parent (for subjects 17 years and younger). The instrument consists of 6 questions (4 expectation of outcome questions and 2 disease severity questions). The parent versions only include the 4 questions related to expectation of outcome. The total FAIM score is calculated as the arithmetic average of all non-missing items and ranges from 1 (limited severity perception) to 7 (greatest severity perception).

Descriptive statistics for each item and the total scores along with their changes from baseline will be tabulated. Data will be summarized separately by age group and responder (subject or caregiver). Analyses of change from baseline in the total score will be performed using ANCOVA models of change from baseline at the Exit visit. The model will be fit with terms for treatment group, region, and screening score. Region will be excluded from the model if any region has fewer than 3 subjects with evaluable scores within either treatment group.

Listing of the raw scores as recorded in the CRF will be provided, sorted by treatment group and subject ID. The comparison of treatment groups using the ANCOVA model for both FAQLQ and FAIM corresponds to the following hypotheses:

- H_0: Mean change from baseline within active subjects = mean change from baseline within placebo subjects

- H_A: Mean change from baseline within active subjects ≠ mean change from baseline within placebo subjects

Additional exploratory analyses of the FAQLQ and FAIM may be performed. A description of these exploratory analyses will be documented in an addendum SAP for exploratory Quality of Life assessments.

9.6. EXPLORATORY ANALYSES

9.6.1. Treatment Satisfaction

Assessment of treatment satisfaction will be performed using the Treatment Satisfaction Questionnaire for Medication (TSQM-9), an exit survey including palatability questions.

The TSQM-9 is a widely used instrument to assess treatment satisfaction with medication in studies where patient reported side effects have a potential to interfere with the objectives of the study. The instrument consists of 9 questions that comprise 3 scales.

Responses to the 9 individual items will be presented using descriptive statistics. The scale scores (effectiveness, convenience, and global satisfaction) will be calculated and summarized using descriptive statistics.

The Effectiveness scale includes items 1-3, the Convenience scale includes items 4-6, and the Global Satisfaction scale includes items 7-9. Each scale will be scored as: 100*[(sum of non-missing responses) minus the number of non-missing responses] divided by the maximum possible score of the sum of non-missing responses. If more than one item within the scale has a missing result then the scale score will not be calculated.

Analysis of the scale scores will be performed using ANOVA models with terms for treatment group and region.

In addition to the TSQM-9, an exit survey will be performed at study exit. The survey includes questions on study drug palatability, frequency of taking study drug as instructed, impact on attending clinic visits, interest in continuing to take study drug, if the subject would recommend the study drug to others, and burden of treatment. Responses to each item will be summarized with descriptive statistics and type of instrument administered (parent or subject ages 12 and older). No statistical testing is planned for this questionnaire.

9.6.2. Secondary Efficacy Endpoints Analyzed in Adults and Overall

Analyses described above for key secondary efficacy and other secondary efficacy endpoints will be repeated in adults (Ages 18-55) as exploratory analyses. Selected analyses may be repeated overall (pediatric and adult subjects, Ages 4-55). Where appropriate, age group (pediatric or adult) will be included as a factor in the relevant statistical models and as a stratification factor in the statistical tests. Treatment by age group interaction terms will also be included in the models. This will only be applicable for the analyses based on all subjects, if performed.

9.6.3. Baseline Factors that may Influence Desensitization

The influence of age, region and other potentially prognostic baseline factors on treatment group differences in the desensitization response rate and MTD at Exit DBPCFC may be explored using the methods described in sections 9.6.4 and 9.6.5. If done, the analyses will be performed in the pediatric subject subset.

The factors to be explored will be:

- Pediatric age group (age <12,12-17)
- Age as a continuous variable
- Region
- Sex
- Sex by Age group
- MTD of peanut protein at Screening DBPCFC, in $\log_{10}$ scale

- Presence of asthma, as recorded as part of non-peanut allergy history
- Presence of allergic rhinitis, as recorded as part of non-peanut allergy history
- History of allergy to a food other than peanut, as recorded as part of the non-peanut allergy history
- Severity of the peanut allergy, defined as the number of anaphylactic reactions to peanut experienced, divided by the duration of the peanut allergy, and categorized into: subjects with 0 anaphylactic reactions, those with at least one reaction but less than 1 every 5 years, and those with more than 1 reaction every 5 years
- Duration of peanut allergy in months, as recorded as part of allergy history
- ps-IgE level at baseline, in $\log_{10}$ scale
- ps-IgE group (based on a cutoff predictive of screening MTD ≤30 mg as determined from the Screened Population, see Section 9.6.7)
- Mean SPT peanut wheal diameter at baseline
- BMI at baseline
- Use of epinephrine at the Screening DBPCFC
- GI symptoms as part of the Screening DBPCFC reaction
- Presence of eczema / atopic dermatitis, as recorded as part of non-peanut allergy history

The examination of age group and region are considered supportive and not exploratory.

If necessary, categorical variables will be pooled such that each pooled level will have at least 3 observations for each treatment group.

For nominal categorical variables, pooling will start with the smallest level which will be pooled with the next larger level until each pooled level has at least 3 observations for each treatment group. For ordinal categorical variables, pooling will proceed in the same manner but will start with the minimum ordinal value and pool in ascending order. Ties will be broken by sorting (in ascending order) the desensitization rate for AR101, then by the desensitization rate for placebo, then by the geometric mean of the MTD for AR101, and then by the geometric mean of the MTD for placebo. If the tie is still not broken, then tied levels (that are subject to pooling) are pooled together as the initial step.

If model parameters for a categorical variable are not estimable, then the minimum number of observations requiring at least 3 observations per pooled level for each treatment group will be increased by 1 in the above pooling algorithm until the model parameters are estimable. If the model parameters for a binary categorical variable are not estimable, then that factor will not be explored for its effect.

9.6.4. Effect on Desensitization Response Rate

The influence of factors described in Section 9.6.3 on treatment group differences in the desensitization response rate at 600 mg, 300 mg, and 1000 mg will be explored using

logistic regression. These analyses will be performed on the ITT and Completer populations. For the analyses in the ITT population, missing data will be based on the rules defined for the primary efficacy analysis. For the Completer population, missing data will be excluded.

The logistic regression model will include terms for treatment group and the baseline factor to be explored, and an interaction term between the baseline factor with treatment group. The odds ratio for the treatment group effect will be calculated at each level for categorical variables. The odds ratio for the treatment group effect will be calculated at 3 values for continuous variables (the 25% percentile, median, and 75% percentile). Main effects for treatment group and the baseline factor will also be calculated after dropping the interaction term from the model. A 95% confidence interval and p-value will be presented for each odds ratio using the Wald statistic. The p-value for the interaction term will also be presented and based on the Wald statistic. For categorical variables, the number and percent of responders for each level of the categorical variable by treatment arm will be tabulated.

9.6.5. Effect on Maximum Tolerated Dose at Exit DBPCFC

Each baseline factor described in Section 9.6.3, will independently be evaluated using the discrete hazards model similar to the model described in Section 9.5.1. These analyses will be performed on the ITT population and Completer population and PP population, if applicable.

The discrete hazards model will include terms for treatment, the baseline factor to be explored, and an interaction term between the baseline factor with treatment will be included. The hazard ratio for the treatment effect will be calculated at each level for categorical variables. The hazard ratio for the treatment effect will be calculated at 3 values for continuous variables (the 25% percentile, median, and 75% percentile). Main effects for treatment and the baseline factor will also be calculated after dropping the interaction term from the model. A 95% confidence interval and p-value will be presented for each hazard ratio using the Wald statistic. The p-value for the interaction term will also be presented and based on the Wald statistic.

9.6.6. Imputation of MTD based on Maximum AR101 Dose Achieved

Subjects who do not undergo the Exit DBPCFC have the MTD for the Exit DBPCFC imputed as the MTD at the Screening DBPCFC, as described in Section 7.4.3. As an exploratory analysis, subjects randomized to AR101 who do not undergo the Exit DBPCFC will have their MTD for the Exit DBPCFC imputed as the higher of the Screening DBPCFC MTD or the highest Exit DBPCFC dose level (1, 3, 10, 30, 100, 300 mg) less than or equal to the highest tolerated dose of randomized study treatment tolerated during Up-dosing. The highest tolerated dose is defined as the highest dose level taken for

at least two weeks, which is not immediately followed by a dose reduction or withdrawal from the study.

The number and percent of subjects by their MTD at the Exit DBPCFC will be summarized for the AR101 treatment group with this imputation scheme, with missing Exit DBPCFC imputed using their MTD at screening, and without any imputation.

9.6.7. Cutoff Predictive of MTD ≤ 30

For both ps-IgE and SPT wheal diameter, analyses may be performed in a post-hoc manner to identify cutpoints, with good performance characteristics, that discriminate between subjects eligible for randomization (MTD≤30 mg) versus those not eligible (MTD>30 mg). Performance characteristics to be considered, include positive predictive value (PPV), sensitivity, and specificity.

9.6.8. Relationship of ps-IgE and SPT on Premature Discontinuations

The influence of ps-IgE and SPT wheal diameter on failure to achieve the target dose of 300 mg/d (and similarly of subjects who develop gastrointestinal symptoms and discontinue prematurely) may be explored in a post-hoc manner using logistic regression using the methods described in section 9.6.3. Additionally, the methods described in Section 9.6.7 will be used to identify cutoffs for ps-IgE and SPT wheal diameter predictive of failure to achieve the target dose of 300 mg/d (and similarly of subjects who develop gastrointestinal symptoms and discontinue prematurely).

These analyses will be performed on the ITT population.

9.7. IMMUNE TOLERANCE NETWORK (ITN) SAMPLES

Optional blood samples are collected for exploratory analysis by the Immune Tolerance Network. These analyses will be conducted separately and are outside the scope of this analysis plan. Sample collection dates and times will be listed by subject and time point. Selected analyses including demographics and baseline characteristics, medical history, and select efficacy endpoints may be performed in the subgroup of subjects participating in this substudy.

9.8. INTERIM ANALYSIS

There is no interim analysis of efficacy planned for this study.

9.9. DATA SAFETY MONITORING COMMITTEE

A DSMC will monitor the study for safety. The DSMC will meet periodically to review accruing safety data. The DSMC will not be prospectively provided with any efficacy data, and the trial will not be stopped for any reasons related to efficacy. The DSMC is an

independent group consisting of three clinicians and one biostatistician who have pertinent experience in the management of Adult and Pediatric patients with Peanut Allergies, as well as in the conduct and monitoring of randomized clinical trials. These individuals will be entirely independent of the conduct of the study. Further details, including treatment masking, will be provided in the DSMC Charter.

10. SAFETY

Safety will be assessed on the basis of extent of exposure, concomitant medications, physical examinations, and all the safety endpoints defined in Section 3.3.

Unless otherwise noted, safety data will be summarized descriptively and the Safety population will be used for all summaries of safety parameters. In general, safety data will be summarized separately by age group (Pediatric 4-17 and Adults 18-55) and overall (ages 4-55). Further breakdown by age group (Children 4-11 and Adolescents 12-17) will be provided where indicated. Safety listings will include all randomized subjects, sorted by age group and treatment group (AR101 then placebo).

10.1. STUDY TREATMENT EXPOSURE

Study treatment exposure will be summarized by age group (4-17 years and 18-55 years), treatment (AR101 and placebo), and study period (initial escalation, up-dosing, maintenance, and overall). Due to data collection issues with the electronic diary, the main calculation of exposure will be based on in clinic dosing data and the dose level of the dispensed study medication.

First and last dose dates for each study period will be identified as follows:

Study Period	First Dose Date	Last Dose Date
Initial Escalation	Date of first in-clinic dose at Initial Escalation Day 1	Date of last in-clinic dose at Initial Escalation Day 1 or Day 2
Up-Dosing	The day following the date study product was dispensed at Initial Escalation Day 2	Date of first in-clinic dose of 300 mg at the End Up-Dosing 300 mg Visit. For subjects who do not reach the 300 mg dose: the latest of the last in-clinic dose during the Up-Dosing period or the date of last study drug administration as entered on the Early Termination CRF page.
Maintenance	The day following the date study product was dispensed at End Up-Dosing 300 mg Visit	Latest of the last in-clinic dose during the Maintenance period or the date of last study drug administration as entered on the Early Termination CRF page

The total amount of study product consumed will be calculated as the sum of in-clinic doses plus the estimated amout of doses taken at home. At-home doses will be estimated by calculating the number of days between in clinic visits multiplied by the dose amout dispensed at the previous in clinic visit. This definition assumes that the subject took all doses between visits. If possible, the data will be reviewed to determine if subjects were temporarily taken off study drug for a period of time and those days will be removed from the number of at-home doses.

The following calculations of study drug exposure will be made and summarized:

- Duration of Exposure (days): calculated as the date of last dose of study drug minus the date of the first dose of study drug plus one during the study period, except for the Initial Escalation period, the duration of exposure will only be 1 or 2 days depending on whether drug was taken on Initial Escalation Day 1 and Initial Escalation Day 2. Duration of exposure will be summarized using descriptive statistics for continuous endpoints as well as categorically by 28 day increments for the overall treatment period: ≤ 28 days, 29 – 56 days, … 337 – 364 days, and >364 days.

- Total dose consumed (mg): calculated as the cumulative sum of all doses taken during the study period.

- Average dose per day (mg): calculated as the total dose consumed divided by the number of days during the study period.

- Number of unsuccessful dose increases: where an unsuccessful dose increase is defined as a single in clinic dose at a higher dose level followed by an immediate return to the previous dose level or a lower dose level.

- Number of dose reductions: where a dose reduction is defined as any decrease in dose level that does not qualify as an unsuccessful dose increase.

- Maximum dose achieved (mg/day): summarized using descriptive statistics for continuous endpoints as well as categorically using all possible dose levels: 0.5, 1, 1.5, 3, 6, 12, 20, 40, 80, 120, 160, 200, 240, or 300 mg/day.

- Time to 300 mg dosing and time to 80 mg dosing for the overall treatment period using Kaplan-Meier methodology. Time will be calculated as date of the first 300 mg (or 80 mg) dose minus the first dose date +1. Subjects who do not reach the specified dose will be censored at the date of their last study drug dose.

The non-missing valid diary entries will be used to estimate at-home dosing compliance. The following measures of compliance with at-home dosing will be calculated:

- Total number of planned at-home dosing days: calculated as the number of days where a valid diary entry was made, but excluding entries where a dose was missed because of doctor's orders.

- Percentage of planned dosing days where a full or partial dose was consumed.

- Percentage of planned dosing days where a full dose was consumed.

- Percentage of planned dosing days where a partial dose was consumed.
- Percentage of planned dosing days where a dose was missed.

At-home dosing data will be listed. Daily diary records, including date and time, whether a full or partial dose was consumed (or the dose was missed), reason for partial or missed dose will be listed. If any dose-related symptoms are present, the symptoms, onset time, and resolution time will also be listed.

10.2. PRIOR, CONCOMITANT, AND RESCUE MEDICATIONS AND THERAPIES

All medications recorded on the Concomitant Medications CRF page and on the On-Study Rescue Medication CRF page will be coded using the World Health Organization drug dictionary (WHODRUG), June 2015 version. Medications will be listed and summarized by Anatomical Therapeutic Chemical (ATC) Level 1 and Preferred Name.

Prior medications are defined as those which are only taken prior to the date of the first dose of study drug on Day 1 (i.e, medication end date is prior to the date of first dose of study drug).

Concomitant medications are medications taken at any time during the active treatment period. Any medications recorded for which dosing began after the last dose of randomized study treatment will also be classified as concomitant medications. As needed (PRN) medications, which may or may not be taken for long periods of time, but which are prescribed to the subject for a period that overlaps with the active treatment period, will be considered concomitant medications. If it cannot be determined whether a medication was received prior to the start of study drug dosing due to partial or missing medication start and/or end dates, it will be considered a concomitant medication.

Rescue medications are any medication used to treat symptoms of an acute allergic reaction and are recorded on a separate on-study rescue medication CRF. Unless administered during a DBPCFC, each use of a rescue medication during OIT should be associated with a corresponding adverse event (AE).

Prior medications and concomitant medications excluding rescue medications will be summarized by ATC Class, Preferred Name and treatment group. Subjects will be counted no more than one time per Preferred Name and no more than one time per ATC Level 4 in the summary.

Rescue medications will be summarized by ATC Class, Preferred Name, and treatment group for the following study periods:

- Screening (excluding rescue medications taken as a result of the Screening DBPCFC)

- Screening DBPCFC, where the medication start date is during the screening period and the CRF page indicates the medication was taken as a result of DBPCFC symptoms

- Initial Escalation period or the Up-dosing period

- Maintenance period (excluding rescue medications taken as a result of the Exit DBPCFC)

- Exit DBPCFC, where the medication start date is after the start of the maintenance period and the CRF page indicates the medication was taken as a result of DBPCFC symptoms

- Overall, including all rescue medications reported on the CRF

Concomitant non-drug therapies will be listed by subject.

10.3. ADVERSE EVENTS

Symptoms that are collected during in-clinic dosing visits (including DBPCFC) are not captured on the adverse event CRF page, unless the symptom is determined to be an SAE or the subject experienced an anaphylaxis episode. Thus, each of the summaries of adverse events described below will be presented in 2 different ways:

1. Including only events collected on the adverse event CRF page, excluding in-clinic dosing symptoms other than SAEs and anaphylaxis episodes.

2. Combining adverse events collected on the adverse event CRF page and dosing symptoms collected at the in-clinic dosing visits.

All reported adverse events (AEs) will be classified into System Organ Class (SOC) and Preferred Term (PT) using Medical Dictionary for Regulatory Activities (MedDRA) version 18.1.

Treatment-emergent adverse events (TEAEs) are defined as those AEs with onset after the first dose of study drug. AEs with onset prior to first dose of study drug will be included in subject listings, but not summarized. TEAEs will be summarized for the safety population by age group (4-17 years, 4-11 years, 12-17 years, 18-55 years, and 4-55 years) and by study period as follows:

- Initial Escalation: All events beginning after the first dose of study drug on Day 1 and prior to the first at-home Up-dosing administration.

- Up-dosing: All events beginning after the first dose at-home Up-dosing administration and prior to the first at-home dose of 300 mg (which starts the Maintenance Period).

- Maintenance: All events beginning after the first dose at-home dose of 300 mg, including events after the Exit DBPCFC but prior to the rollover to ARC004.

- Overall: Across initial escalation, up-dosing, and maintenance periods.

Summaries that are displayed by system organ class and preferred terms will be ordered by descending incidence of system organ class and preferred term within each system organ class. Summaries displayed by preferred term only will be ordered by descending incidence of preferred term. Summaries of the following types will be presented:

- Overall summary of number of unique TEAEs and treatment-emergent serious adverse events (TESAEs), subject incidence of TEAEs and TESAEs meeting various criteria, and exposure adjusted incidence rates of TEAEs and TESAEs

meeting various criteria, where exposure incidence rates are defined as the total number of events divided by the total number of subject-years at risk during the study period;

- Subject incidence of TEAEs by MedDRA system organ class and preferred term;
- Exposure adjusted event rates for the most frequently-occurring TEAEs (i.e., TEAEs occurring in ≥ 5% of the Safety Population) by MedDRA preferred term;
- Subject incidence of TEAEs by severity grade, MedDRA system organ class, and preferred term;
- Subject incidence of TEAEs related to study drug by MedDRA system organ class, and preferred term;
- Exposure adjusted event rates for the most frequently-occurring TEAEs related to study drug (i.e., related TEAEs occurring in ≥ 5% of the Safety Population) by MedDRA preferred term;
- Subject incidence of ≥ grade 3 severity TEAEs related to study drug by MedDRA system organ class and preferred term;
- Subject incidence of SAEs by MedDRA system organ class and preferred term;
- Subject incidence of TEAEs leading to discontinuation of study drug by MedDRA system organ class and preferred term;
- Subject incidence of hypersensitivity TEAEs by MedDRA system organ class and preferred term, where hypersensitivity TEAEs are those TEAEs flagged as allergic reactions;
- Subject incidence of hypersensitivity TEAEs related to study drug by MedDRA system organ class and preferred term;
- Subject incidence of TEAEs associated with an accidental food allergen exposure by MedDRA system organ class and preferred term; and
- Subject incidence of TEAEs with onset < 90 minutes after study drug dosing by MedDRA system organ class and preferred term.

At each level of summarization (e.g., any AE, system organ class, and preferred term), subjects experiencing more than one TEAE will be counted only once within each study period. In the summary of TEAEs by severity grade, subjects will be counted once at the highest severity reported at each level of summarization.

Adverse event data will be presented in data listings by age group, treatment group, subject, study period, and event. Serious AEs; severe, life-threatening, or fatal adverse events; and AEs leading to discontinuation, reduction, or interruption of the study drug will be presented in separate data listings.

10.4. FOOD ALLERGY EPISODES

The occurrence of a safety event associated with accidental food ingestion will be reported as a food allergy episode, as per Section 7.3 of the protocol. Any such event that meets the definition of an SAE will also be reported as an adverse event. All reported food allergy episodes will be listed by age group, treatment group, and subject. Episodes of allergic reaction associated with foods other than peanut will be flagged.

Food allergy episode will be summarized by study period (overall, Initial Escalation, Up-dosing and Maintenance). For each period, the number of subjects experiencing any food allergy episode, the number of subjects experiencing a food allergy episode in response to peanut (or non-peanut), the number of episodes of each (peanut-related and non-peanut related) experienced per subject, and the total number of food allergy episodes (peanut and non-peanut related) will be summarized. The number of episodes with unscheduled clinic visits, treatment administered, or considered SAEs will also be summarized.

10.5. SYMPTOMS DURING DBPCFC

During each food challenge, the severity of pre-specified symptoms is rated as mild, moderate, severe, life-threatening, or death at each dose level of each food product. In addition, the presence of any dose-limiting symptoms is identified.

The number of subjects experiencing any dose related symptoms and the maximum severity of any symptoms will be summarized by individual dose level and overall during the Screening peanut challenge, Screening placebo challenge, Exit peanut challenge, and Exit placebo challenge for the Safety population. Subjects will be counted at most once per type of challenge and dose level for symptom severity at the most severe level recorded for that subject.

Symptoms at the Screening and Exit DBPCFCs will be listed by subject.

10.6. SYMPTOMS DURING IN-CLINIC STUDY DRUG DOSING

During each on-treatment, in-clinic study drug dosing, the severity of pre-specified symptoms is similarly rated as mild, moderate, severe, life-threatening, or death. In addition, the presence of dose-related symptoms, and/or whether the dose was tolerated is recorded for each dose.

The number of subjects experiencing any dose related symptoms during any on-treatment, in-clinic study drug dosing visit and the maximum severity of any symptoms will be

summarized at each dose level. If a subject is administered the same dose at more than one in-clinic visit (either the subject remained at the same dose as the previous visit, or a subject had a prior dose increase and subsequent dose reduction), the most severe symptoms will be summarized for that dose level.

The time from dose administration to the time of onset of the first symptom will be summarized using descriptive statistics and presented by dose level. Time from onset of the first symptom to resolution of the last symptom will be summarized similarly. If a subject receives more than one in-clinic dose at the same dose level that subject will be counted only once using the shortest time from dose administration to onset of first symptom and the longest time from onset of first symptom to resolution of last symptom.

10.7. ADVERSE EVENTS DURING AT-HOME DOSING

The subject diaries will serve as source documentation for the collection of AEs occurring at home and these will be reviewed by the investigator, entered into the eCRF, and reported within the safety database. As for all AEs, these will be reported with a brief description of the at-home events, and their onset and resolution dates and times. The investigator will assess the severity of the AEs, their etiology (principally determining whether allergic or non-allergic), and their relatedness to study-product.

10.8. PREGNANCY TEST RESULTS

Pregnancy test results will be listed by age group, treatment group, subject, and visit.

10.9. SPIROMETRY AND PEFR

Spirometry and/or Peak Expiratory Flow Rate (PEFR) assessments are performed prior to any DBPCFC. PEFR is also performed at each Up-dosing and Maintenance visit. Spirometry may be performed at any time during the study where a subject's pulmonary status is in question. Three attempts of FEV_1 are performed, and the best (highest) value flagged in data listings. Similarly, three attempts of PEFR are performed, and the best (highest) value flagged in data listings. Only the best FEV_1 value and the best PEFR value will be summarized. For analysis of FEV_1/FVC ratio, the result corresponding to the best FEV_1 value will be included in the applicable summaries.

Observed values for PEFR, FEV_1, FEV_1 percent predicted, FEV_1/FVC ratio, and FEV_1/FVC percent predicted as well as changes from baseline will be summarized at each applicable visit by treatment group for the Safety population. Results will be listed by age group, treatment group, subject, and visit.

10.10. VITAL SIGNS

BMI will be calculated as (weight in kilograms)/ (height in meters)^2.

Vital signs (pulse rate, systolic/diastolic blood pressure, body temperature, height, BMI, and weight) will be listed by subject and visit. Observed values and change from baseline will be summarized by treatment group at each scheduled visit and time point. At Screening and Exit DBPCFC vital signs are scheduled to be taken prior to each dose given. At Initial Escalation, Up-dosing, and Maintenance visits, vital signs are to be taken pre-dose and within 15-30 minutes after each dose given.

Additional vital signs measurements taken due to extension of the observation period will not be included in the summaries. If a subject is administered the same dose at more than one in-clinic visit (either the subject remained at the same dose as the previous visit, or a subject had a prior dose increase and subsequent dose reduction), an additional summary will be presented for that dose level (e.g., Up-dosing 3 mg Visit 2).

10.11. PHYSICAL EXAMINATION

Physical examination results will be listed by age group, treatment group, subject, and visit. Results will be summarized by body system and treatment in a shift table of worst change from baseline that occurred within Up-dosing and Maintenance study periods, as well as overall.

Missing data will not be imputed.

The worst post-baseline result will be defined, within body system, as:

- Abnormal, clinically significant, if any clinically significant abnormal result is present after the first dose of randomized study treatment
- Abnormal, not clinically significant, if any abnormal, not clinically significant result is present after the first dose of randomized study treatment, but no clinically significant results
- Normal, if all non-missing results after the first dose of treatment are normal
- Not Done, if no results are present after the first dose of treatment

Population totals will be used as the denominator for all percentages.

10.12. ASSESSMENT OF ASTHMA CONTROL

Assessment of asthma control in asthmatic subjects using the Asthma Control Test (ACT) questionnaire will be performed at Baseline, Up-dosing Interim visit, End of Up-dosing Period visit, and Exit visit.

For subjects 12 years old or older, the ACT has 5 questions each recorded on a scale of 1 (worst control) to 5 (complete control). The total ACT is the sum of the 5 scores and ranges from 5 (worst control) to 25 (total control). A total score of 19 or less indicates asthma is not adequately controlled. Missing data will not be imputed. If any of the 5 questions have a missing response, the total ACT score will not be calculated.

For subjects under 12, there are 4 questions for the subject and 3 questions for the parents to complete. Subject responses range from 0 (worst control) to 3 (complete control). Parent responses range from 0 (every day) to 5 (no days). The sum of all 7 questions will make up the total score. The total ACT score for subjects under 12 will range from 0 (worst control) to 27 (worst control). Missing data will not be imputed. If any of the questions have a missing response, the total ACT score will not be calculated for that subject.

All analyses of the ACT will be performed separately by subject age group. Summary statistics of the score for question, total score and change from baseline will be tabulated by visit and treatment. A shift table of asthma control (adequate, not adequate, missing) will be summarized by treatment at each visit. The number of subjects with completed ACT questionnaires will be used as the denominator for all percentages.

Listing of the results from the questionnaire, including the total score, will be provided, sorted by age group, treatment group, subject, and visit.

10.13. ASSESSMENT OF GI SYMPTOMS BY PEESS

Subjects under the age of 18 who discontinue treatment due wholly or in part to GI AEs will be instructed to complete the Pediatric Eosinophilic Esophagitis Symptom Scores (PEESSTM v2.0) questionnaire (Franciosi, et al., 2011) monthly for 6 months.

The PEESS questionnaire is composed of 20 items investigating 2 domains. The total score consists of all 20 items. The frequency domain consists of items 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, and 20, where each item is scored as: 0=Never, 1=Almost never, 2=Sometimes, 3=Often, 4=Almost always. The severity domain consists of items 2, 4, 6, 8, 10, 12, 14, 16, and 18, where each item is scored as: 0=Not bad at all, 1=A little bad, 2=Kind of bad, 3=Bad, 4=Very bad. Each item score is transformed to 0-100 as follows: 0=0, 1=25, 2=50, 3=75, 4=100.

The total and domain scores are computed as the sum of the items divided by the number of items answered. If more than 50% of the items for the calculation of the total score or either of the domain scores are missing, the score will not be calculated.

Summary statistics for the toal score and each PEESS domain will be tabulated by time point and treatment. PEESS results including the domain scores, will be listed.

10.14. EPINEPHRINE USE AS RESCUE MEDICATION

Epinephrine use is defined as any rescue medication with a preferred name of ‘EPINEPHRINE’ when coded as described in Section 10.2, and with documentation of administration to a subject on a specific date.

All subjects, per protocol, are required to have epinephrine autoinjectors for use in case of a suspected anaphylactic reaction occurring outside of the clinic. There are, however, differences in how, and even if, physicians record the prescription of epinephrine autoinjectors for as-needed (PRN) use. As a consequence of this, the presence or absence

of a PRN prescription for epinephrine cannot be taken to indicate epinephrine usage, regardless of whether the prescription was written prior to, or after, enrollment in the study. What is important is to be able to quantitate the number of subjects receiving doses of epinephrine and the number of doses. As epinephrine should only be administered to treat a discrete allergic reaction, each dose of epinephrine should be closely temporally associated with a specific safety event and its use recorded on the Rescue Medication eCRF form. In cases where a PRN epinephrine prescription is issued after the start of study-product dosing, the sites will be queried as to if, and when, epinephrine was actually administered and to treat what specific event.

Epinephrine use will be categorized by the following period or events, based on epinephrine use on the same day(s) as any of the following events:

- Any use on the same date as an accidental exposure (defined as a recorded food allergy episode) occurring on or after the first dosing date of randomized treatment
- Screening DBPCFC, further categorized into
    - Peanut Challenge
    - Placebo Challenge
- Initial Escalation
- Up-dosing
- Maintenance
- Exit DBPCFC, further categorized into
    - Peanut Challenge
    - Placebo Challenge

If both peanut and placebo DBPCFCs were on the same date (either at Screening or Exit), any epinephrine use on that date will be categorized as used for the peanut challenge.

Frequency of use will be summarized overall (for any frequency) and by frequency administered (Once, twice, etc.). The number of subjects at risk at each time point will be used as the denominators; for instance, only subjects undergoing Exit DBPCFC will be used as the denominator for epinephrine use during Exit DBPCFC. The summary will be conducted in the Safety population. Data will be listed for all randomized subjects.

10.15. ANAPHYLAXIS EPISODES

All reported anaphylactic reaction episodes will be listed by age group, treatment group, and subject.

Each anaphylaxis reaction will be identified by the following triggers:

- DBPCFC
- Study drug

- Food allergen other than study drug while on treatment
- Non-food allergen while on treatment
- Other

Anaphylaxis reactions will be summarized in the Safety population separately for the following study periods:

- Screening DBPCFC, Peanut Challenge
- Screening DBPCFC, Placebo Challenge
- Overall Treatment Period (including Initial Escalation, Up-dosing, Maintenance)
- Initial Escalation
- Up-dosing
- Maintenance
- Exit DBPCFC, Peanut Challenge
- Exit DBPCFC, Placebo Challenge

For study periods of overall treatment period, Initial Escalation, Up-dosing, and Maintenance, additional summaries will be included by trigger (any trigger, study drug, food allergen other than study drug while on treatment, and non-food allergen while on treatment)

The number of anaphylactic reactions, the number and percent of subjects experiencing an anaphylactic reaction, the number and percent of subjects experiencing an anaphylactic reaction by maximum severity using the Murano Grading Scale (Muraro, et. Al), the number of subjects experiencing an anaphylactic reaction that was an SAE, the number of subjects experiencing an anaphylactic reaction that required use of epinephrine, and the number of subjects experiencing an anaphylactic reaction that involved individual symptoms will be summarized for each study period and trigger combination.

10.16. ALGORITHMIC SEARCH OF SYSTEMIC HYPERSENSITIVITY EVENTS

Systemic hypersensitivity events involving two or more body systems that occur together within a 2-hour window and not already determined as anaphylaxis events by the study investigators will be identified using Standard MedDRA Queries (SMQ) for Anaphylactic reaction AEs, hypotonic-hyporesponsive episode AEs, and selected sponsor-defined gastrointestinal AEs possibly associated with an allergic reaction to oral immunotherapy.

Systemic hypersensitivity cases must include the following categories or combination of categories (see below for categories' PTs):

1. Any term from category A or,
2. Any term from category D1 or,

3. Any term from category B and any term from categories C or D2 or E or,
4. Any term from category C and any term from categories D2 or E or,
5. Any term from categories D2 and any term from category E

Table of AE terms preferred terms:

Category	SMQ or Category Description	Preferred Term
A	Anaphylactic reaction (SMQ)	Anaphylactic reaction
		Anaphylactic shock
		Anaphylactic transfusion reaction
		Anaphylactoid reaction
		Anaphylactoid shock
		Circulatory collapse
		Dialysis membrane reaction
		Kounis syndrome
		Shock
		Shock symptom
		Type I hypersensitivity
B	Upper airway/respiratory AEs	Acute respiratory failure
		Asthma
		Bronchial oedema
		Bronchospasm
		Cardio-respiratory distress
		Chest discomfort
		Choking
		Choking sensation
		Circumoral oedema
		Cough
		Cyanosis
		Dyspnoea
		Hyperventilation
		Dyspnoea
		Hyperventilation
		Irregular breathing
		Laryngeal dyspnoea
		Laryngeal oedema
		Laryngospasm
		Laryngotracheal oedema
		Mouth swelling
		Nasal obstruction
		Oedema mouth
		Oropharyngeal spasm
		Oropharyngeal swelling
		Respiratory arrest
		Respiratory distress
		Respiratory failure
		Reversible airways obstruction
		Sensation of foreign body
		Sneezing
		Stridor
		Swollen tongue
		Tachypnoea
		Throat tightness
		Tongue oedema

		Tracheal obstruction
		Tracheal oedema
		Upper airway obstruction
		Wheezing
C	Angioedema/urticaria/pruritus/flush AEs	Allergic oedema
		Angioedema
		Erythema
		Eye oedema
		Eye pruritus
		Eye swelling
		Eyelid oedema
		Face oedema
		Flushing
		Generalised erythema
		Injection site urticaria
		Lip oedema
		Lip swelling
		Nodular rash
		Ocular hyperaemia
		Oedema
		Periorbital oedema
		Pruritus
		Pruritus allergic
		Pruritus generalised
		Rash
		Rash erythematous
		Rash generalised
		Rash pruritic
		Skin swelling
		Swelling
		Swelling face
		Urticaria
		Urticaria papular
D1	Cardiovascular/hypotension AEs	Blood pressure decreased
		Blood pressure diastolic decreased
		Blood pressure systolic decreased
		Cardiac arrest
		Cardio-respiratory arrest
		Cardiovascular insufficiency
		Diastolic hypotension
		Hypotension
D2	Hypotonic-hyporesponsive episode (SMQ)	Hypotonic-hyporesponsive episode
		Hypotonia
		Hypotonia neonatal
		Altered state of consciousness
		Depressed level of consciousness
		Hypokinesia
		Hypokinesia neonatal
		Hyporesponsive to stimuli
		Loss of consciousness
		Neurogenic shock
		Presyncope
		Shock
		Shock symptom
		Syncope

		Unresponsive to stimuli
		Cyanosis
		Cyanosis central
		Cyanosis neonatal
		Pallor
		Skin discolouration
E	GI Related Symptoms	Abdominal discomfort
		Abdominal pain
		Abdominal pain upper
		Diarrhoea
		Enlarged uvula
		Gastrointestinal disorder
		Lip oedema
		Lip pain
		Lip pruritus
		Lip swelling
		Mouth swelling
		Nausea
		Odynophagia
		Oral discomfort
		Oral mucosal erythema
		Oral pruritus
		Paraesthesia oral
		Retching
		Saliva altered
		Salivary hypersecretion
		Swollen tongue
		Tongue pruritus
		Vomiting

The following incidences will be summarized by treatment groups:

- Subjects with systemic hypersensitivity events that were:
    - Related to study drug
    - Related to study drug and led to study drug discontinuation
    - Serious
    - Severe

11. PRE-DATABASE LOCK BLINDED DATA REVIEWS

As stated previously and in the ARC003 Treatment Masking Plan and ARC003 to ARC004 Rollover Procedures document, major data queries that have or may have bearing on the efficacy and/or safety aspects of the study will be resolved prior to unblinding individual subjects. These include:

- Desensitization status from the DBPCFC (Screening and Exit)
- Major protocol deviations that might affect assignment of subjects into analysis populations,
- Fulfillment of inclusion/exclusion criteria,
- Accidental peanut exposures that resulted in SAEs or AEs
- Severity and relationship to investigational product of SAEs and AEs

The purpose of the pre-database lock blinded data review is for:

- Identification of major and minor protocol deviations (Section 5).
- Assignment of subjects to their appropriate analysis populations (Section 4).
- Classification of the treatment related and recurrent/chronic GI discontinuations (Section 8.1).
- Classification of their desensitization status (Responder or Non-responder; Section 7.4) and non-responder subcategories (Section 7.4.4).
- Classification of accidental food allergen exposures (Section 10.4) as peanut-related or non-peanut related.

Only ARC003 project team members who are blinded to study treatment assignments (i.e., Medical Monitor, Statistician, Data Manager, Sponsor clinical project staff) will be involved in the data reviews. No member of the ARC003 project team (including data management and statistician) will have the ability to link subject identification data from ARC003 and ARC004.

12. CHANGE FROM ANALYSIS PLANNED IN PROTOCOL

Any deviations from the plans detailed in this SAP will be described and justified in the final clinical study report. A separate document to this SAP will provide a table of contents and mockups for the expected layout and titles of the tables, listings, and figures. Any changes to format, layout, titles, numbering, or any other minor deviation will not necessitate a revision to the SAP nor will it be considered a deviation from planned analyses. Only true differences in the analysis methods or data handling will necessitate such documentation.

13. REFERENCE LIST

Anagnostou K, Islam S, King Y, Foley L, Pasea L, Bond S, Palmer C, Deighton J, Ewan P, Clark A. Assessing the efficacy of oral immunotherapy for the desensitisation of peanut allergy in children (STOP II): a phase 2 randomised controlled trial. Lancet. 2014 Apr 12;383(9925):1297-304.

Chinchilli VM, Fisher L, Craig TJ. Statistical issues in clinical trials that involve the double-blind, placebo-controlled food challenge. J Allergy Clin Immunol 2005;115:592-7.

Cook, Thomas D, and David L DeMets. Introduction to Statistical Methods for Clinical Trials. Boca Raton, FL: Chapman & Hall/CRC, 2008, p374

EMEA. Points to Consider on Multiplicity Issues in Clinical Trials, London (UK), EMEA CPMP 2002

Farrington, C. P. and Manning, G. 1990. 'Test Statistics and Sample Size Formulae for Comparative Binomial Trials with Null Hypothesis of Non-Zero Risk Difference or Non-Unity Relative Risk.' Statistics in Medicine, Vol. 9, pages 1447-1454.

Fleischer DM, Burks AW, Vickery BP, Scurlock AM, Wood RA, Jones SM, Sicherer, SH, Liu AH, Stablein D, Henning AK, Mayer L, Lindblad R, Plaut M, Sampson HA; Consortium of Food Allergy Research (CoFAR). Sublingual immunotherapy for peanut allergy: a randomized, double-blind, placebo-controlled multicenter trial. J Allergy Clin Immunol. 2013 Jan;131(1):119-27.e1-7.

Glaumann S, Nopp A, Johansson SG, Borres MP, Nilsson C. Oral peanut challenge identifies an allergy but the peanut allergen threshold sensitivity is not reproducible. PLoS One. 2013;8(1):e53465.

Kirby S, Burke J, Chuang-Stein C, Sin C. Discounting phase 2 results when planning phase 3 clinical trials. Pharm Stat. 2012 Sep-Oct;11(5):373-85.Le UH, Burks, AW. Oral and Sublingual Immunotherapy for Food Allergy. World Allergy Organization Journal 2014;7:35.

Sampson HA, Leung DY, Burks AW, Lack G, Bahna SL, Jones SM, Wong DA. A phase II, randomized, double‐blind, parallel‐group, placebo‐controlled oral food challenge trial of Xolair (omalizumab) in peanut allergy. J Allergy Clin Immunol. 2011 May;127(5):1309-10.e1.

Sampson H, et al. Epicutaneous Immunotherapy (EPIT) Is Effective and Safe to Treat Peanut Allergy: A Multi-National Double-Blind Placebo-Controlled Randomized Phase IIb Trial. AAAAI Annual Meeting 2015; Abstract L28 (American Academy of Allergy, Asthma & Immunology Meeting DBV Investor Update - February 2015)

van der Velde JL, Flokstra-de Blok BM, Vlieg-Boerstra BJ, Oude Elberink JN, DunnGalvin A, Hourihane JO, Duiverman EJ, Dubois AE. Development, validity and reliability of the food allergy independent measure (FAIM). Allergy. 2010 May;65(5):630-5

Varshney P, Jones SM, Scurlock AM, Perry TT, et al. A randomized controlled study of peanut oral immunotherapy: clinical desensitization and modulation of the allergic response. J Allergy Clin Immunol, 2011. 127(3): p. 654-60.

Wang SJ, Hung HM, O'Neill RT. Adapting the sample size planning of a phase III trial based on phase II data. Pharm Stat. 2006 Apr-Jun;5(2):85-97.

14. PROGRAMMING CONSIDERATIONS

All tables, data listings, figures (TLFs) will be generated using SAS® for Windows, Release 9.4 or higher (SAS® Institute Inc., Cary, NC, USA). Computer-generated table, listing and figure output will adhere to the following specifications.

14.1. GENERAL CONSIDERATIONS

Report summaries will be generated using validated Base SAS® software, version 9.4 or higher, on a PC or server-based platform. Additional validated software may be used to generate analyses, as needed.

Data will be analyzed by Agility Clinical biostatistics personnel. Statistical analyses will be reported with tables, figures, and listings, presented in rich text format, and using recommended ICH numbering. Output specifications for all tables, figures, and listings will be in conformance with guidelines specified by the ICH in Appendix 7 of the *Electronic Common Technical Document Specification* (Apr 2003).

14.2. REPORTING CONVENTIONS

Tables and figures will be summarized by treatment group. Tables summarizing demographics and other baseline characteristics will also include a column for all subjects combined. In general, all data collected and any derived data will be presented in subject data listings, for all randomized subjects. Listings will be ordered by age group, treatment group, subject number, and assessment or event date. The treatment group presented in listings will be based on the planned assignment, unless otherwise noted.

In general, continuous variables will be summarized to indicate the population sample size (N), number of subjects with available data (n), mean, SD, median, minimum, and maximum values. Categorical variables will be summarized by the population size (N), number of subjects with available data (n), number of subjects in each category, and the percentage of subjects in each category. Unless otherwise noted, the denominator to determine the percentage of subjects in each category will be based on the number of subjects with available data. Select ordinal data may be summarized using both descriptive statistics and counts and percentages of subjects in each category, as appropriate.

Non-zero percentages will be rounded to one decimal place. Rounding conventions for presentation of summary statistics will be based on the precision of the variable of summarization, as it is collected in its rawest form (ie, on the electronic case report form [eCRF] or as provided within an external file) and are outlined as follows:

- The mean and median will be rounded to one more decimal place than the precision of the variable of summarization;

- Measures of variability (eg, SD, SE) will be rounded to two more decimal places than the precision of the variable of summarization; and

- Minimum and maximum values will be presented using the same precision as the variable of summarization.

Other statistics (eg, CIs) will be presented using the same general rules outlined above, or assessed for the most appropriate presentation based on the underlying data.

Statistical significance testing will be two-sided and performed using α=0.05. P-values will be reported for all statistical tests, rounded to four decimal places. P-values less than 0.0001 will be displayed as "<0.0001"; p-values greater than 0.9999 will be displayed as ">0.9999". Tests of interaction terms, if applicable, will be two-sided and performed using α=0.10.

Unless otherwise noted, values reported as greater than or less than some quantifiable limit (eg, "< 1.0") will be summarized with the sign suppressed in summary tables and figures, using the numeric value reported. Data will display on subject listings to include the sign.

15. QUALITY CONTROL

All SAS programs that create outputs or supporting analysis datasets will be validated by a second statistical programmer or biostatistician. At a minimum, validation of programs will consist of a review of the program log, review of output or dataset format and structure, and independent confirmatory programming to verify output results or dataset content. Additionally, all outputs will undergo a review by a senior level team member before finalization.

The content of the source data will be reviewed on an ongoing basis by project statistical programmers and statisticians. Data will be checked for missing values, invalid records, and extreme outliers through defensive programming applications, analysis-based edit checks, and other programmatic testing procedures. All findings will be forwarded to the project data manager for appropriate action and resolution.

16. STUDY SCHEDULE

Refer to the protocol for the full study schedule of events (Appendix 1)

17. INDEX OF TABLES, LISTINGS AND FIGURES

An index of the planned statistical outputs will be provided in the shell TLF document.