STATISTICAL ANALYSIS PLAN

ARC010
AR101 Trial in Europe Measuring Oral Immunotherapy Success in Peanut Allergic Children (ARTEMIS)

Version 1.0 – 04 Dec 2018
Reference Numbers: NCT03201003, EudraCT 2016-005004-26

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

Sponsor Name:	**Aimmune Therapeutics, Inc.**
Protocol Number and Title:	**ARC010** **AR101 Trial in Europe Measuring Oral Immunotherapy Success in Peanut Allergic Children (ARTEMIS)** Protocol Version 4.0 (Amendment 3) 28 August 2018
Authors:	David Norval Director, Biometrics EU Aimmune Therapeutics, Ltd. Brian Beus / Laura Holland Precision for Medicine, Oncology and Rare Disease 6005 Hidden Valley Road, Suite 170 Carlsbad, CA 92011
SAP Version:	Version 1.0
SAP Version Date:	04Dec2018

Version: Version 1.0	**Version Date: 04DEC2018**
Document Effective Date:	04 December 2018
Document File Name:	Aimmune_ARC010_SAP_v1.0_04DEC2018.pdf

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

Precision for Medicine, Oncology and Rare Disease

Laura Holland, MS
Biostatistician I — Date: 10 JAN 2019

Brian Beus, MS
Principal Biostatistician — Date: 10JAN2019

Aimmune Therapeutics, Inc

Stuart McIntosh, BM BS, MRCS, MD, FFPM
Vice President, Clinical Development, Europe — Date: 06 DEC 18

Rezi Zawadzki, DrPH
Vice President, Biometrics — Date: 17 Dec 18

David Norval
Director, Biometrics EU — Date: 06 DEC 18

TABLE OF CONTENTS

GLOSSARY OF ABBREVIATIONS

Abbreviation	Description
ACT	Asthma Control Test
AE	Adverse Event
ANCOVA	Analysis of Covariance
AR101	Characterized Peanut Allergen
ATC	Anatomical Therapeutic Class
CI	Confidence Interval
CoFAR	Consortium of Food Allergy Research
CRF	Case Report Form
CRO	Contract Research Organization
CTCAE	Common Terminology Criteria for Adverse Events
DBPCFC	Double-Blind Placebo-Controlled Food Challenge
DLS	Dose-Limiting Symptom
DSMC	Data Safety Monitoring Committee
EMEA	European Medicines Agency
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 Harmonisation
IgE	Immunoglobulin E
IgG	Immunoglobulin G
ITN	Immune Tolerance Network
ITT	Intent-to-Treat
IUS	Intrauterine System
IXRS	Interactive Voice or Web-Based Response System

Abbreviation	Description
LLOQ	Lower Limit of Quantification
MedDRA	Medical Dictionary for Regulatory Activities
MTD	Maximum Tolerated Dose
OIT	Oral Immunotherapy
PEFR	Peak Expiratory Flow Rate
PP	Per Protocol
PRACTALL	Practical Allergy, a joint initiative between the European Academy of Allergy and Immunology and The American Academy of Asthma, Allergy and Immunology
PRN	As needed (pro re nata)
ps	Peanut-specific
PT	Preferred Term
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SOC	System Organ Class
SOP	Standard Operating Procedure
SPT	Skin Prick Test
TD	Treatment difference
TEAE	Treatment-Emergent Adverse Event
TLF	Table, Listing and Figure
TSQM-9	Treatment Satisfaction Questionnaire for Medication
ULOQ	Upper Limit of Quantification
WHO-DDE	World Health Organization Drug Dictionary

1. PURPOSE

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

1.1. DOCUMENT HISTORY

This is the first draft.

1.2. RESPONSIBILITIES

The SAP text was drafted by David Norval (Aimmune Therapeutics Ltd.) and then further revised by Precision for Medicine, Oncology and Rare Disease ("Precision") to create table, listing and figures shells. This collaboration with Precision will include further development and finalization of the SAP.

Precision will perform the final statistical analyses and be responsible for the production and quality control of all tables, listings and figures. In addition, Precision will also 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 Precision 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 to 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 and adolescents (ages 4 to 17 years, inclusive).

2.2. SECONDARY OBJECTIVES

The secondary objectives of the study are:

- To demonstrate the safety of AR101 as measured by the frequency of adverse events (AEs), including serious adverse events (SAEs).
- To evaluate the immunological effects of peanut oral immunotherapy (OIT).

2.3. BRIEF DESCRIPTION OF STUDY DESIGN

This is a European, 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 Protocol Section 6.6 for additional details of the DBPCFC.

Peanut-allergic children (ages 4-17 years inclusive) will have an initial screening DBPCFC of up to 300 mg (444 mg cumulative) peanut protein or placebo. Those experiencing dose-limiting symptoms (DLS) at or before the 300 mg dose of peanut protein (measured as 600 mg of peanut flour) will be enrolled. Those who successfully consume and tolerate a 300 mg (444 mg cumulative) dose of peanut protein during the Screening DBPCFC (i.e., without manifesting DLS), will be considered screen failures and will not be randomized. Further, any subject who is assessed to have had DLS to the placebo part, or both parts, of the Screening DBPCFC (i.e., to oat flour as well as peanut flour) will be considered a screen failure and will not be randomized.

Approximately 160 subjects who pass Screening will be randomized 3:1 to either AR101 (active treatment) or placebo using a proprietary interactive response system. See Protocol Section 4 for a complete list of inclusion and exclusion criteria.

All eligible subjects will receive escalating doses of either AR101 or placebo. The Treatment Phase comprises 3 periods:

- **Initial Escalation** (2 consecutive days) – Eligible subjects will be randomized and initiate OIT starting at a dose of 0.5 mg of IP and then increase the dose incrementally at 20 to 30 minute intervals over the course of a single day to a maximum dose of 6 mg. Subjects who fail to tolerate at least a 3 mg dose will be considered escalation failures and will be discontinued from the study. Subjects who tolerate both the 3 mg and 6 mg doses of IP, or who tolerate the 3 mg but not the 6 mg dose, will undergo confirmatory testing of the tolerability of a 3 mg dose the following consecutive day (refer to Initial Escalation Schedule at end of synopsis). Subjects who tolerate this confirmatory dose will enter the Up-dosing Period. Subjects who do not tolerate this confirmatory dose will be discontinued. Therapy details are found in Section 3 and Section 6 of the protocol.
- **Up-dosing Period** – After the initial escalation period, subjects will report to the Clinical Research Center (CRC) every 2 weeks to escalate their OIT dose to an expected daily dose of 300 mg of peanut protein. This constitutes the Up-dosing Period. Subjects will receive daily oral dosing of peanut or placebo OIT for approximately 20 weeks, if up-dosing proceeds without holding at, or reducing, a dose level; up to 40 weeks, maximum. All escalation doses will be administered in a CRC. All up-dosing activities will be performed under direct observation. Therapy details can be found in Section 3 and Section 6 of the protocol.
- **Maintenance Period** – Those subjects who reach the target maintenance dose of 300 mg/d of IP will enter an approximately 12-week Maintenance Period of continued dosing at 300 mg/d, which may be extended by up to an additional 4 weeks (for a maximum Maintenance Period duration of 16 weeks), or to a total Treatment Phase duration of 56 weeks, whichever occurs first.

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 300 mg Up-dosing Visit, with visits every 4 weeks thereafter. 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.

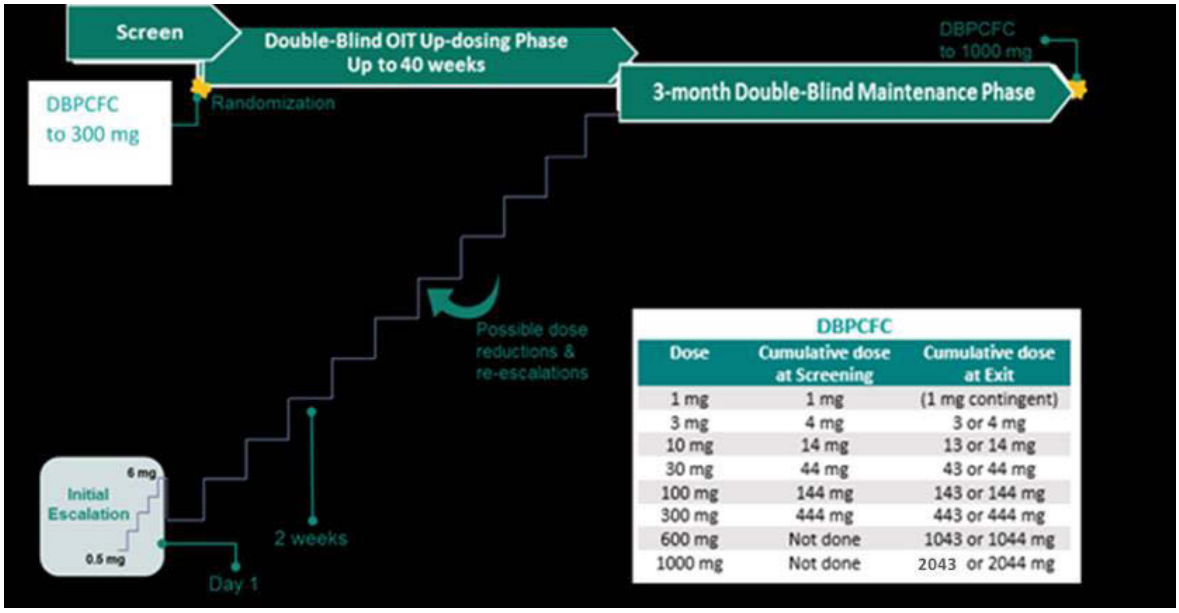
All subjects who reach the targeted daily dose of 300 mg and maintain that dose through the Maintenance Period will undergo an Exit DBPCFC. The Exit DBPCFC will include a 600 mg challenge dose, and the top challenge dose will be capped at 1000 mg. For details of the Exit DBPCFC refer to Section 3.3 of the protocol.

Each subject will be unblinded after he/she completes the DBPCFC and all major data queries to date (i.e., queries that could influence allocation to one or another analysis population) for the subject have been resolved.

Subjects who do not reach the target dose of 300 mg/d by Week 40 are not eligible for the Exit DBPCFC and will be considered escalation failure non-responders. These subjects will be unblinded once the database has been locked and unblinded.

The study 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 the Figure below.

Figure 2.1 Illustration of study design



2.4. DETERMINATION OF SAMPLE SIZE

As stated in the protocol the sample size for the study is approximately 160 subjects randomized 3:1 (AR101:placebo):

"The sample size for the study, approximately 160 subjects randomized 3:1 (AR101:placebo), has sufficient power to detect a treatment effect for the primary efficacy analysis. In Study ARC001, the placebo group response rate (95% CI) in the ITT population was 19% (7%, 39%) and 0% (0%, 13%) at 443 mg and 1043 mg cumulative peanut, respectively. Subjects were enrolled in Study ARC001 if at the Screening DBPCFC they expressed DLS at ≤ 144 mg cumulative peanut protein whereas in Study ARC010 they will be enrolled if they express DLS ≤ 443 mg cumulative protein. Extrapolating to Study ARC010, where the primary efficacy endpoint is response rate in the ITT population at 2043 mg cumulative peanut protein, the placebo response rate is

assumed to be at most 15%. With these considerations, a sample size of 120 AR101 and 40 placebo subjects and a 15% placebo response rate, there is at least 80% power to detect a response rate of 39% or higher in the AR101 group using Fishers exact test with a 0.05 two-sided Type I error rate. Under the same assumptions, there is at least 90% power to detect a response rate of at least 43% in the AR101 group."

The sample size has been independently checked using nQuery 8. The power for the 39% and 43% response rates using the assumptions as above was calculated as 80.4% and 90.8% respectively.

2.5. TREATMENT ASSIGNMENT AND BLINDING CONSIDERATIONS

Refer to the protocol and to the ARC010 Treatment Masking Plan and ARC010 to ARC008 Rollover Procedures document for full details on treatment assignment and blinding procedures including those to be followed for emergency/unplanned and planned 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.

3. ENDPOINTS

3.1. PRIMARY EFFICACY ENDPOINT

The primary efficacy endpoint is the proportion of subjects who tolerate at least 1000 mg as a single dose (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

The key secondary efficacy endpoints are as follows:

- The proportion of subjects who tolerate at least 600 mg as a single dose (1043 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC
- The proportion of subjects who tolerate at least 300 mg as a single dose (443 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC
- The maximum severity of symptoms occurring following ingestion of peanut protein during the Exit DBPCFC

3.2.2. Other Secondary Efficacy Endpoints

The other secondary efficacy endpoints are as follows:

- The MTD with no more than mild symptoms at Exit DBPCFC
- Change from baseline in MTD of peanut protein at Exit DBPCFC
- Use of epinephrine as a rescue medication at Exit DBPCFC and comparison to its use at Screening DBPCFC
- Changes in serum peanut- and peanut component-specific IgE, total IgE, and peanut-specific IgG4 levels
- Changes in peanut SPT mean wheal diameter
- Quality of life assessments using the food allergy related quality of life questionnaire (FAQLQ) (DunnGalvin, 2008; Flokstra-de Blok, 2008; Flokstra-de Blok, 2009) and the food allergy independent measure (FAIM; van der Velde et al, 2010) questionnaire

3.3. EXPLORATORY ENDPOINTS

The exploratory endpoints are as follows:

- Treatment satisfaction assessment using the Treatment Satisfaction Questionnaire for Medication (the TSQM-9) and an exit questionnaire
- Assessment of palatability (taste and after-taste)

3.4. SAFETY ENDPOINTS

The safety endpoints are as follows:

- The safety of peanut OIT based on AEs and SAEs
- Use of epinephrine as a rescue medication during OIT (Initial Escalation, Up-dosing, and Maintenance Periods)
- Frequency and severity of anaphylaxis during OIT (Initial Escalation, Up-dosing, and Maintenance Periods), attributable to IP or accidental food exposure
- Frequency and severity of allergic reaction (hypersensitivity) AEs occurring during the Up-dosing versus the Maintenance Period, normalized for duration of treatment
- Frequency of accidental ingestion of peanut and other allergenic foods; frequency and severity of reactions resulting from accidental ingestion of peanut and other allergenic foods
- Frequency and severity of premature discontinuation of dosing due to AEs; and frequency of premature discontinuation of dosing due to chronic/recurrent GI AEs
- Assessment of asthma control using the 2007 National Heart, Lung and Blood Institute (NHLBI) classification and the Asthma Control Test questionnaire in subjects with asthma

4. ANALYSIS POPULATIONS

The following analysis populations will be defined for this study.

4.1. INTENT-TO-TREAT (ITT) POPULATION

The ITT population (i.e., the Full Analysis Set) will consist of all randomized subjects who receive at least one dose of randomized study treatment (AR101 or placebo). 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.

4.2. COMPLETER POPULATION

The Completer population will include 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 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.

4.3. 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. Exclusions will be determined by blinded data 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 key secondary efficacy endpoints will be performed on the PP population if the PP population differs from the 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.

4.4. SAFETY POPULATION

The Safety population will consist of all subjects who receive at least one dose of randomized study treatment. The Safety population will be used for summaries of safety parameters. Subjects will be analyzed according to treatment received.

5. PROTOCOL DEVIATIONS

All protocol deviations will be reported on a specific case report form (CRF) as follows:

- 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 Concomitant Medication
- Lab Sample missed
- Study Drug Compliance
- Other (with free text field to record detail)

Additional criteria to exclude subjects from the PP population may be added and will be documented in a SAP amendment or in a supporting document.

Protocol deviations will be reviewed in a blinded fashion and categorized into major or minor prior to database lock. Major protocol deviations that may influence the desensitization response will exclude the subject from the PP population. All protocol deviations, both major and minor, will be listed and included in the study report.

6. GENERAL ASPECTS FOR STATISTICAL ANALYSIS

6.1. 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". Select ordinal data may be summarized using both descriptive statistics and counts and percentages of subjects in each category, as appropriate.

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

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 (e.g., "< 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.

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 and any derived data will be included in data listings and unless otherwise noted sorted by 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.

All statistical methods will be based on the International Conference on Harmonisation (ICH)-E9 Guidance for Industry "Statistical Principles for Clinical Trials".

A review of the database will be conducted before database lock. Any decision to amend the planned statistical analysis will be documented in an amendment to the statistical analysis plan prior to unblinding of the study and details will be included in the CSR.

If the assumptions underlying the formal statistical methods proposed are not met during the analysis of the final data, an alternative, more appropriate, statistical method will be used and any changes documented in the statistical methods section of the CSR, including the rationale for use.

Additional exploratory analyses of the data will be conducted as deemed appropriate. These analyses will be fully documented and clearly identified as post-hoc and exploratory.

6.2. STRATA AND COVARIATES

The primary efficacy endpoint of the proportion of subjects who achieve desensitization will be compared between treatment groups using Fisher's exact test. In addition, and as a supportive analysis, the Cochran–Mantel–Haenszel test will also be performed with country as a factor.

The key secondary endpoints of 1) the proportion of subjects who tolerate at least 600 mg as a single dose (1043 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC and 2) the proportion of subjects who tolerate at least 300 mg as a single dose (443 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC will be analyzed using the same methods as for the primary efficacy endpoint.

Further, the analysis of the key secondary endpoint of maximum severity of symptoms occurring following ingestion of peanut during the Exit DBPCFC will include a factor for country.

Post-hoc exploratory analyses (logistic regression) of the primary efficacy endpoint and key secondary endpoints may be performed if numbers allow where specified baseline factors are fitted in the analyses to explore their effects on study endpoints.

6.3. SUBGROUP ANALYSES

Analysis of the primary and key secondary endpoints will be repeated for the following subgroup of age on the ITT population and the completer population:

Age:

- Children 4-11 years inclusive
- Adolescents 12-17 years inclusive

6.4. 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.

6.5. 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.

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.

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., the 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) that is 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 1000 mg responder, a subject must meet both of the following conditions:

1) Must have attained a maximum tolerated single dose (MTD) ≥ 1000 mg of peanut protein on Exit DBPCFC
2) Must not have experienced more than mild symptoms through 1000 mg of peanut protein on Exit DBPCFC.

Classification as a 300 mg responder and a 600 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 1000 mg of peanut protein with at most mild symptoms is a 1000 mg non-responder).

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.

7.5. 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 12 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 16 weeks), or to a total Treatment Phase duration of 56 weeks, whichever occurs first. For subjects who continue maintenance dosing after the Exit DBPCFC but prior to the rollover to ARC008, those data will be attributed to the maintenance period and identified as post-DBPCFC in listings.

- 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.

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 the 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.

For any anaphylactic reactions that are missing severity, severe, the highest severity on the Muraro scale, will be imputed.

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

7.7. POOLING

Data pooling will not be performed unless it becomes necessary to group countries for analyses where country is included in the analysis model. Any grouping will be carried out prior to unblinding.

8. DEMOGRAPHIC, OTHER BASELINE CHARACTERISTICS

8.1. SUBJECT DISPOSITION AND WITHDRAWALS

The number of subjects screened and randomized as well as the number and percentage of subjects who entered each study period, completed all dosing as defined in the protocol, and were included in each analysis population will be summarized by age overall (ages 4 to 17) and by age group (Children 4-11 and Adolescents 12-17) and treatment group (AR101, Placebo and Total).

Primary reason for early discontinuation from the study will be summarized as well as whether subjects required follow-up due to chronic/recurrent GI symptoms.

The date of last study visit and reason for discontinuation will be listed as well as follow-up status for subjects who required follow-up due to recurrent GI symptoms.

Inclusion and exclusion eligibility and screen failures 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 for the Safety population by age overall (ages 4 to 17) and by age group (Children 4-11 and Adolescents 12-17) and treatment group (AR101, Placebo and Total).

8.3. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Summary statistics for demographic and baseline characteristics will be provided for the ITT, Completer, PP and Safety populations.

Demographic data will include age, race, country, ethnicity, sex, body weight, height and body mass index (BMI). Baseline characteristics include total IgE, ps-IgE, ps-IgG_4, ps-IgE/IgG_4 ratio, SPT mean wheal diameter at screening, MTD of peanut protein at Screening DBPCFC and Childbearing Potential. Whether the subject had asthma (intermittent, mild persistent, moderate persistent or severe persistent) will also be summarized.

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 information will be listed for all randomized subjects. Available demographic data for screening failure subjects will also be listed.

8.4. PEANUT ALLERGY HISTORY

The duration of peanut allergy (months since peanut allergy diagnosis), number of anaphylactic reactions to peanut in lifetime, months since most recent allergic reaction, and the symptoms experienced during the most recent peanut exposure will be summarized for the Safety population by age overall (ages 4 to 17) and by age group (Children 4-11 and Adolescents 12-17) and treatment group (AR101, Placebo and Total).

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

Peanut allergy history will be listed by subject.

8.5. NONPEANUT ALLERGY HISTORY

The presence of nonpeanut allergy history and causative allergens will also be summarized for the Safety population by age overall (ages 4 to 17) and by age group (Children 4-11 and Adolescents 12-17) and treatment group (AR101, Placebo and Total).

All nonpeanut allergy history will be listed by subject.

8.6. OTHER MEDICAL HISTORY

Medical history will be listed by subject. Subjects who experience medical history events will be summarized for the Safety population by age overall (ages 4 to 17) and by age group (Children 4-11 and Adolescents 12-17) and treatment group (AR101, Placebo and Total), and by MedDRA system organ class and preferred term.

9. EFFICACY

9.1. PRIMARY ENDPOINT

The primary efficacy endpoint is the proportion of subjects who achieve desensitization as determined by tolerating 1000 mg as a single dose (2043 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC. These subjects are termed responders.

All subjects failing to achieve the success definition described above are termed non-responders. This includes subjects who drop out of the study or discontinue OIT prior to undergoing the Exit DBPCFC (treatment failure non-responders) as well as subjects who fail to achieve and maintain a 300 mg daily dose of IP (escalation failure non-responders).

The primary efficacy endpoint of desensitization response rates will be compared between treatment groups using Fisher's exact test on the ITT population.

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

- H_A: Proportion of active subjects with a desensitization response – proportion of placebo subjects with a desensitization response ≠ 0

Further, the desensitization response rates and their associated 95% confidence intervals will be presented for each treatment group using exact Clopper-Pearson confidence intervals. The 95% confidence interval for the treatment difference (AR101 desensitization rate minus placebo desensitization rate) will be based on exact unconditional confidence limits using the score statistic (RISKDIFF(METHOD=SCORE)).

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.

9.2. ADDITIONAL ANALYSES OF THE PRIMARY EFFICACY ENDPOINT

As a supportive analysis, the Cochran–Mantel–Haenszel test, stratifying for country, will also be performed to compare the desensitization response rates between treatment groups.

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.

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). The primary efficacy endpoint will also be analyzed by country in the ITT population as supportive analyses to the primary efficacy analysis using the methods described in Section 9.1. Further, efficacy results will be presented overall and by country using a forest plot of the treatment difference and corresponding 95% confidence interval.

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 on the ITT population will be tested in the following hierarchical order:

1. The proportion of subjects who tolerate at least 600 mg as a single dose (1043 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC

2. The proportion of subjects who tolerate at least 300 mg as a single dose (443 mg cumulative) of peanut protein with no more than mild symptoms at the Exit DBPCFC

3. The maximum severity of symptoms occurring following ingestion of peanut protein during the Exit DBPCFC

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 statistically significant, all subsequent p-values will be displayed for informational purposes only.

The first two key secondary endpoints (desensitization response rates) will be analyzed using the same methods as the primary analysis described in section 9.1.

The third key secondary endpoint (maximum severity of symptoms at the Exit DBPCFC) will be analyzed using the Cochran-Mantel-Haenszel test. The objective of this analysis is to determine if subjects from the AR101 group will have lower chance of developing more severe levels of symptom severity compared to subjects from the placebo group. Symptom severity is determined according to the CoFAR scale at 5 levels: 1-Mild, 2-Moderate, 3-Severe, 4-Life Threatening, and 5-Fatal. Analysis is done on 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 is 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 maximum severity of symptoms observed in the

DBPCFC at any dose (1000 mg or lower) will be used for each subject. 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 number and percentage of subjects by maximum severity at the Exit DBPCFC will be tabulated by treatment arm. The Cochran-Mantel-Haenszel statistics with equally spaced scores (Statistic: Row Mean Score Differences) stratified by country will be used to test for a treatment difference.

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.

Each comparison will be evaluated for statistical significance (two-sided $p < 0.05$) only if the preceding test in the hierarchy is statistically significant in favor of AR101.

Analysis of the key secondary endpoints will be repeated in the Completer population as a sensitivity analysis and also in the PP population (if sufficiently different from the Completer population).

9.4. ANALYSIS OF OTHER SECONDARY ENDPOINTS

If the primary efficacy analysis of the primary endpoint and all the hierarchical testing of key efficacy secondary endpoints as described in Section 9.3 are found to be statistically significant, then statistical testing on the ITT population 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 other secondary endpoints are included in the overall hierarchical testing procedure although these endpoints are considered supportive in nature.

9.4.1. Maximum Tolerated Dose with no more than mild symptoms at Exit DBPCFC

The Screening and Exit DBPCFCs are based on a modified PRACTALL dosing regimen as described in the protocol (Section 3). With the exception of the 600mg 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, country, 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 fitted 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 country). The adjusted probability estimates will also be plotted. An unadjusted probability estimate will also be calculated by removing the MTD at baseline and country 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 country. 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
- H_A: The hazard ratio for subjects within active subjects relative to placebo subjects ≠ 1

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.

The analysis will be performed using the ITT and Completer populations.

9.4.2. Change from Baseline in MTD of peanut protein at Exit 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. 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 fitted with terms for treatment group, country, 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

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

The p-value is based on the F-test for treatment group effect adjusted for the MTD at screening ($\log_{10}$ mg) and country.

Least squares mean statistics (point estimate and 95% 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.

If the assumptions underlying the ANCOVA are not met then alternative methods of analysis will be explored, e.g., non-parametric methods as appropriate. Further, the ANCOVA model will also be assessed for unequal slopes by adding a term for the treatment by baseline interaction. If there is evidence of unequal slopes then the least squares means and their associated 95% confidence intervals and p-value for the treatment group difference will be calculated and reported at screening MTDs of 1 mg, 3 mg, 10 mg, 30 mg and 100 mg.

The analyses will be performed using the ITT and Completer populations.

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

The number and percentage of 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 the Completer population:

- 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 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 fitted 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 (Statistic: Row Mean Score Differences).

Listings of all Epinephrine use will be provided.

The analyses will be performed using the Completer population.

9.4.4. Changes in Peanut-specific IgE, total IgE and peanut specific IgG_4

Blood samples to measure ps-IgE, ps-IgG_4 levels, and total IgE levels will be collected prior to the Screening DBPCFC, end of up-dosing visit, 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 fitted with terms for treatment group, country 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

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

Analyses will be performed using the ITT and Completer populations.

9.4.5. Peanut Skin Prick Test

Results from the SPT will be listed, including test date and time, 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 and change from baseline to Exit visit will be presented.

An ANCOVA model of change from baseline at Exit DBPCFC will be fitted with terms for treatment group, country 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 their associated 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

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

Analyses will be performed using the ITT and Completer populations.

9.4.6. 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 prior to the screening DBPCFC and after the Exit DBPCFC (and subject's treatment assignment has been unblinded).

Separate FAQLQ and FAIM instruments are administered based on the subject's age group. Parent versions are also administered for all subjects. 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 parent).

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 using a different form by age group (8-12 years and 13-17 years) (Flokstra-de Blok, 2008; Flokstra-de Blok, 2009). Evaluations are also done by the parent/caregiver using a different form by age group (4-6 years, 7-12 years, and 13-17 years) (DunnGalvin, 2008).

Each question is presented on a scale from 0 (not at all) to 6 (extremely) and scaled from 1 to 7. The number of items and domains varies by instrument administered. For reporting, the domains for each included form are as follows:

Teen form –

a) Allergen Avoidance and Dietary Restrictions
b) Risk of Accidental Exposure
c) Emotional Impact

Child form –

a) Allergen Avoidance and Dietary Restrictions (Allergen Avoidance + Dietary Restrictions)
b) Risk of Accidental Exposure
c) Emotional Impact

Parent form teen –

a) Social and Dietary Limitations (Dietary Restrictions + Social Restrictions)
b) Food Anxiety
c) Emotional Impact

Parent form 7-12 –

a) Social and Dietary Limitations
b) Food Anxiety
c) Emotional Impact

Parent form 4-6 –

a) Social and Dietary Limitations
b) Food Anxiety
c) Emotional Impact

For each domain, the domain average score is the arithmetic average of the non-missing items comprising the domain. For all forms, the total score will be calculated as the average of the domain averages: (average a + average b + average c) / 3. For FAQLQ-PF 4-6, if items are completed that were not included for this age group, these items will not be included towards the scoring.

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 fitted with terms for treatment group, country and screening value.

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 using a different form by age group (8-12 years and 13-17 years) (van der Velde, 2010). Evaluations are also done by the parent/caregiver using a different form by age group (4-12 years and 13-17 years). The instrument consists of 6 questions (4 expectation of outcome questions and 2 disease severity questions). The parent/caregiver versions include questions related to perception of disease severity and expectation of allergen exposure outcome. The parent/caregiver form includes 8 questions for subjects aged 4-12 years and 4 questions for subjects aged 13 to 17 years. The FAIM is scored on a 7-point scale from 1 (limited severity perception) to 7 (greatest severity perception), and the total FAIM score is calculated as the arithmetic average of all nonmissing items.

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 parent). 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 fitted with terms for treatment group, country and screening score.

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

9.5. EXPLORATORY ANALYSES

9.5.1. Treatment Satisfaction

Assessment of treatment satisfaction will be performed using the Treatment Satisfaction Questionnaire for Medication (TSQM-9) and an exit survey including palatability questions. Both are to be completed after the Exit DBPCFC and the subject's treatment assignment has been unblinded.

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 country.

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 for each type of instrument administered (parent or subject ages 12 and older, and subjects ages 4 to 11 for the drug palatability as appropriate). No statistical testing is planned for this questionnaire.

9.6. INTERIM ANALYSIS

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

9.7. 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 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, are provided in the DSMC Charter.

10. SAFETY

Safety will be assessed based on extent of exposure, concomitant medications, physical examinations, and all the safety endpoints defined in Section 3.4.

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 (Children 4-11 and Adolescents 12-17) and overall (ages 4 to 17). 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 overall (ages 4 to 17) and by age group (Children 4-11 and Adolescents 12-17), treatment (AR101 and placebo) and study period (initial escalation, up-dosing, maintenance and overall). The 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 dose at home.
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 dose at home prior to Exit DBPCFC, or in the case of delayed roll-over subjects, the date of last dose following Exit DBPCFC.
Overall	Date of first in-clinic dose at Initial Escalation Day 1	Last dose of study product

The total amount of study product consumed will be calculated as the sum of in-clinic doses plus the estimated number 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 amount 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 (in days and in months): 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 data as well as categorically by 28 day increments for the overall treatment period: ≤ 28 days, 29 – 56 days, … 253 – 280 days, and >280 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.

- Maximum dose achieved (mg/day): summarized using descriptive statistics for continuous data 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.

10.2. PRIOR, CONCOMITANT, AND RESCUE MEDICATIONS AND THERAPIES

All medications recorded on the Concomitant Medications CRF page will be coded using the World Health Organization Drug Dictionary Enhanced (WHO-DDE), September 2016 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 indicated as such on the concomitant medication page. 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
- 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

Treatment-emergent adverse events (TEAEs) will be summarized excluding symptoms recorded during the food challenges. The AE summaries will report AEs and Allergic AEs unless stated otherwise.

If symptoms are recorded as part of an anaphylaxis reaction as reported on the Allergic AE or in-clinic dosing form, only the single anaphylaxis reaction event will be summarized and not the individual symptoms.

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 19.1.

Treatment-emergent adverse events are defined as those AEs with onset after the first dose of study drug and no more than 30 days after the last dose of study drug. Non treatment-emergent AEs will be included in subject listings, but not summarized. TEAEs will be summarized for the safety population by age group (4-11 years, 12-17 years and 4-17 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 ARC008.

- 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;
- Subject incidence of TEAEs by MedDRA preferred term;
- Exposure-adjusted event rates for the most frequent 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 treatment-related AEs by MedDRA system organ class, and preferred term;
- Subject incidence of treatment-related AEs by MedDRA preferred term;
- Exposure-adjusted event rates for the most frequent treatment-related AEs (i.e., treatment-related AEs occurring in ≥ 5% of the Safety Population) by MedDRA preferred term;
- Subject incidence of grade ≥ 3 severity TEAEs by MedDRA system organ class and preferred term;
- Subject incidence of grade ≥ 3 severity TEAEs by MedDRA preferred term;
- Subject incidence of grade ≥ 3 severity treatment-related AEs 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 product by MedDRA system organ class and preferred term;
- Subject incidence of TEAEs with onset < 90 minutes after study product dosing by MedDRA system organ class and preferred term.
- Subject incidence of TEAEs associated with Epinephrine use 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 all TEAEs from the in-clinic dosing case report forms and all TEAEs from the allergic adverse event case report form;

- Subject incidence of hypersensitivity treatment-related AEs by MedDRA system organ class and preferred term;

- Subject incidence of hypersensitivity TEAEs by MedDRA preferred term;

- Exposure-adjusted event rates for the most frequent hypersensitivity TEAEs (i.e., hypersensitivity TEAEs occurring in ≥ 5% of the Safety Population) by MedDRA preferred term;

- Subject incidence of TEAEs associated with nonstudy product food allergen exposure by MedDRA system organ class and preferred term; and

- Subject incidence of TEAEs leading to early withdrawal 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 (4-11 years, 12-17 years and 4-17 years), treatment group, subject, study period, and event. Serious AEs; severe, life-threatening, or fatal adverse events; anaphylactic reactions; GI AEs in subjects who discontinue due to chronic/recurrent GI AEs; 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 7.3.3.3 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 nonpeanut), the number of episodes of each (peanut-related and nonpeanut related) experienced per subject, and the total number of food allergy episodes (peanut and nonpeanut related) will be summarized. The number of episodes considered SAEs, those that required treatment, and those that required epinephrine use will also be summarized.

10.5. SYMPTOMS DURING DBPCFC

During each food challenge, the severity of pre-specified symptoms is rated on the CoFAR scale as mild, moderate, severe, life-threatening, or death at each dose level of each food product. In addition, the presence of any dose related 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. PREGNANCY TEST RESULTS

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

10.7. SPIROMETRY AND PEFR

Spirometry and/or Peak Expiratory Flow Rate (PEFR) assessments are performed at screening, initial escalation, and 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.

*Only for subjects 6 years of age and older who are able to adequately perform spirometry.

10.8. 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 treatment, subject and visit. Observed values and change from baseline will be summarized by treatment group at each scheduled visit and time

point for pulse rate, systolic/diastolic blood pressure, and body temperature. At Screening and Exit DBPCFC vital signs are scheduled to be taken just prior to each dose given or at 15-20 minutes post-dose, if the between challenge dosing interval is prolonged. Screening DBPCFC vital signs will not be summarized. 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).

10.9. PHYSICAL EXAMINATION

Physical examination results will be listed by age group, treatment group, subject and visit.

Missing data will not be imputed.

10.10. 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 or Early Discontinuation visit.

For subjects 12 years old or older, the ACT has 5 questions each recorded on a scale of 1 (least control) to 5 (greatest control). The total ACT is the sum of the 5 scores and ranges from 5 (least control) to 25 (greatest 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 (least control) to 3 (greatest 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 (least control) to 27 (greatest control). A total score of 19 or less indicates asthma is not adequately controlled. 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.

10.11. ASSESSMENT OF GI SYMPTOMS BY PEESS

Subjects 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 (Martin, 2015; Franciosi, et al., 2011) monthly for 6 months.

The PEESS questionnaire is composed of 20 items investigating the frequency and severity of eosinophilic esophagitis (EoE) symptoms in the last month. The total score consists of all 20 items. The frequency of symptoms is assessed by 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 of symptoms is assessed by 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, frequency total, and severity total 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 a score are missing, the score will not be calculated.

Summary statistics for frequency of symptoms, severity of symptoms, and the total score will be tabulated by time point and treatment. PEESS results including the frequency total, severity total, and total scores, will be listed.

10.12. 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.

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 CRF 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.

All on-study epinephrine use will be listed for the Safety population.

In addition, epinephrine use excluding food challenges will be summarized by study period. The number and percent of subjects with at least 1 episode, where episode refers to 1 or more doses of epinephrine within a 2-hour window, will be summarized as well as the number of episodes experienced by each subject (1, 2, 3, and > 3). For these subjects, demographics including age, age category, baseline PS IgE, and baseline PS IgE category will be included. The total number of episodes will also be presented along with the number and percent of episodes by number of doses per episode, by associated adverse event severity, by seriousness of associated adverse events, by relatedness of associated adverse events, and by location of episode (home/study site). For subject counts, the number of subjects at risk within each study period will be used as the denominators. For unique episode counts, the total number of episodes within each study period will be used as the denominators. Data will be summarized for the safety population by age group (4-11 years, 12-17 years, 4-17 years), study period (initial escalation, up-dosing, maintenance, overall), and treatment group (AR101, placebo).

10.13. 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 product
- Peanut or peanut-containing food
- Other food allergen
- Medication
- Insect sting
- Environmental allergen(s)
- 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

By each study period, the summary will include the number of anaphylactic reactions; the number of anaphylactic reactions by trigger (study product, peanut or peanut-containing food, other food allergen, medication, insect sting, environmental allergen, and other); and subject incidence of subjects experiencing an anaphylactic reaction by number of episodes, subjects experiencing an anaphylactic reaction by maximum severity using the Muraro grading scale (Muraro, 2007; Muraro, 2014), subjects experiencing an anaphylactic reaction that was a serious adverse event, subjects experiencing an anaphylactic reaction that required epinephrine use, the location of epinephrine episodes (home or study site), and the individual symptoms involved.

If any symptom associated with an anaphylactic reaction is serious, the reaction will be classified as serious. This is in addition to, not instead of, the Muraro grade. If any symptom associated with an anaphylactic reaction is related to study product, then the reaction will be classified as related to study product.

11. PRE-DATABASE LOCK BLINDED DATA REVIEWS

The ARC010 Treatment Masking Plan and ARC010 to ARC008 Rollover Checklist Guidance document detail the major data queries that have or may have bearing on the efficacy and/or safety aspects of the study and which will be resolved prior to unblinding individual subjects.

A pre-database lock blinded data review will take place prior to database lock and breaking the blind at the end of the study.

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

- Identification of major and minor protocol deviations
- Assignment of subjects to their appropriate analysis populations

Only ARC010 project team members who are blinded to study treatment assignments (i.e., Medical Monitor, Statistician, Data Manager and Sponsor clinical project staff) will be involved in the data reviews. No member of the ARC010 project team (including the data manager(s) and statistician) will have the ability to link subject identification data from ARC010 and ARC008.

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

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

DunnGalvin A, de BlokFlokstra BM, Burks AW, Dubois AE, Hourihane JO. Food allergy QoL questionnaire for children aged 0-12 years: content, construct, and cross-cultural validity. Clin Exp Allergy. 2008;38(6):977-86.

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

Flokstra-de Blok BM, DunnGalvin A, Vlieg-Boerstra BJ, Oude Elberink JN, Duiverman EJ, Hourihane JO, et al. Development and validation of the self administered Food Allergy Quality of Life Questionnaire for adolescents. J Allergy Clin Immunol. 2008;122(1):139 44.

Flokstra-de Blok BM, DunnGalvin A, Vlieg-Boerstra BJ, Oude Elberink JN, Duiverman EJ, Hourihane JO, et al. Development and validation of a self administered Food Allergy Quality of Life Questionnaire for children. Clin Exp Allergy. 2009;39(1):127 37.

Franciosi JP, Hommel KA, DeBrosse CW, Greenberg AB, Greenler AJ, Abonia JP, et al. Development of a validated patient-reported symptom metric for pediatric eosinophilic esophagitis: qualitative methods. BMC Gastroenterol. 2011;11:126.

Martin LJ, Franciosi JP, Collins MH, Abonia JP, Lee JJ, Hommel KA, et al. Pediatric Eosinophilic Esophagitis Symptom Scores (PEESS v2.0) identify histologic and molecular correlates of the key clinical features of disease. J Allergy Clin Immunol. 2015;135(6):1519-28.

Muraro A, Roberts G, Clark A, Eigenmann PA, Halken S, Lack G, et al. The management of anaphylaxis in childhood: position paper of the European academy of allergology and clinical immunology. Allergy. 2007;62(8):857-71.

Muraro A, Roberts G, Worm M, Bilò MB, Brockow K, Fernández Rivas M, et al. Anaphylaxis: guidelines from the European Academy of Allergy and Clinical Immunology. Allergy. 2014;69(8):1026-45.

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

This page intentionally left blank.

14. PROGRAMMING CONSIDERATIONS

All tables, data listings and figures (TLFs) will be generated using SAS® for Windows, Release 9.4 or higher (SAS® Institute Inc., Cary, NC, USA).

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 Precision 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).

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.