NCT01603628

Study ID: 191622-111

Title: BOTOX® Treatment in Pediatric Lower Limb Spasticity: Double-blind Study

Statistical Analysis Plan Date: August 16, 2017

Allergan

Biostatistics

Analysis Plan

Study ID: 191622-111

Study Title: BOTOX® Treatment in Pediatric Lower limb Spasticity: Double-blind Study

Table of Contents

1 Introduction

This analysis plan details the analyses that will be performed for BOTOX® (Botulinum Toxin Type A) Study 191622-111, a phase 3 study designed to evaluate the safety and efficacy of 2 doses (4 U/kg and 8 U/kg) of BOTOX with standardized physical therapy in the treatment of pediatric lower limb spasticity involving the ankle plantar flexors in monoplegic and hemiplegic children.

1.1 Primary Study Objectives and Design

The primary objective of the study is to evaluate the safety and efficacy of a single treatment of 2 doses (4U/kg and 8U/kg) of BOTOX with standardized physical therapy (PT) in pediatric patients with lower limb spasticity.

This is a multicenter, randomized, double-blind, placebo-controlled, parallel-group, 12-week study evaluating 2 doses of BOTOX, 4 and 8 U/kg, and placebo with standardized PT for the treatment of spasticity involving the ankle plantar flexors in monoplegic and hemiplegic children. Patients who successfully complete this study without major protocol deviations (eg, noncompliance to protocol-required procedures) will be given the option of enrolling in the open-label extension Study 191622-112 if they meet eligibility criteria.

A centralized randomization scheme stratified by age group, and baseline modified Ashworth Scale – Bohannon (MAS-B) will be used. Patients will be randomized in a 1:1:1 ratio ratio to one of the following treatment groups: 4 U/kg (not to exceed 150 U) + PT, 8 U/kg (not to exceed 300 U) + PT, or placebo + PT within each combination of strata. The stratification categories are as follows:

1. Age (≤ 6 years and > 6 years)
2. Baseline (day 1) MAS-B ankle score with knee extended (MAS-B = 2 and MAS-B > 2)

There will be 1 treatment cycle in the study. Only 1 affected lower limb (referred to as the study limb) will be treated. The maximum study duration per patient will be approximately 16 weeks (including up to 4 weeks in the screening period). Patients will complete 8 study visits as follows: screening (week -4 to week -2), initiation of standardized PT (week -2), randomization and treatment visit (day 1), and post-injection follow-up visits at weeks 2, 4, 6, 8, and 12. All patients will receive weekly PT sessions from week -2 through week 11.

Approximately 412 patients will be enrolled to ensure that at least 351 patients (117 per treatment group) will complete the study. Details on sample size calculation and power consideration are presented in Section 7.5 of the study protocol.

1.2 Secondary and Other Objectives

There are no secondary or other objectives for this study.

2 Analysis Populations and Data Conventions

2.1 Analysis Populations

Two populations will be used in the statistical analysis: safety population and modified intent-to-treat (mITT) population. If any patient is mis-stratified by the site at randomization on day 1, this patient's data will be included in the stratum to which the patient should have been stratified instead of the interactive voice response system (IVRS) stratum for analysis.

The safety population will include all treated patients based on the treatment received. All safety analyses will be based on the actual treatment received by each patient in this population.

The mITT population will include all randomized patients with a valid MAS-B baseline ankle score with knee extended and at least one post-baseline measurement at weeks 2, 4, or 6 for the MAS-B ankle score with knee extended and the Clinical Global Impression of Overall Change (CGI) by Physician. Here weeks 2, 4, and 6 refer to the analysis visits defined in Section 2.2. The mITT population will be analyzed according to randomization assignment, regardless of treatment actually received. Efficacy data, demographics, and background information will be analyzed on the basis of the mITT population.

2.2 Treatment Visits

All days are referenced to day 1. Day 1 is the randomization/treatment day. In the cases where randomization and treatment happened on different days, the day of treatment will be used as day 1. Visits according to the windows below will be determined for each variable analyzed unless noted otherwise in the following sections. If there are values from multiple visits in a given window, the value collected from the visit closest to the target day will be used to represent the window. If 2 visits with values fall within the same window and are equidistant from the target day, the value from the first of the 2 visits chronologically will be used to represent the window. For clinical laboratory variables, the last non-missing observation for that visit window will be used. This rule will be applied separately for each variable for non-missing data only.

(The exception is because re-runs of laboratory variables may only involve one or a few variables and would thus include missing data for many variables).

Visit	Target Day of the Visit	Visit Window (Day*)
Screening	-28	≤ Day -1
Day 1 (Randomization/Treatment)	1	Day 1
Week 2	15	Day 2 to Day 22
Week 4	29	Day 23 to Day 36
Week 6	43	Day 37 to Day 50
Week 8	57	Day 51 to Day 71
Week 12	85	≥ Day 72

* All days are referenced to day 1. For post–treatment visits, study day = visit date – day 1 date+1 while for pre-treatment visits, study day = visit date - day 1 date.

The visit windows will be used for by-visit analyses in both modified intent-to-treat (mITT) and safety population analyses.

2.3 Data Conventions

The data conventions and definitions listed in this section are to be applied to all analyses unless stated otherwise in a subsequent section of this Analysis Plan.

- Data will be analyzed according to the 3 treatment groups, ie, (1) BOTOX 8 U/kg, (2) BOTOX 4 U/kg, and (3) Placebo.
- Concomitant medication and prior medication will be calculated using the following convention: 1) If the stop date of taking medication is prior to injection date, then it will be counted as prior medication; 2) If the start date of taking medication is after or on injection date, then it will be counted as concomitant medication; 3) If the start date of taking medication is prior to injection date and stop date is on or after injection date, then it will be counted as both prior medication and concomitant medication.
- Concurrent procedures are defined as those taken on or after the first injection date.
- All data will be listed with the exceptions of commonly expected results such as physical examination, pregnancy tests results for females when not applicable, etc.
- The level of significance used for all statistical tests will be 0.05. All hypotheses will be tested for statistical significance with 2-sided alternatives.

- Summary statistics of continuous variables will include the sample size (N), mean, standard deviation, median, minimum (Min), and maximum (Max).

- Summary statistics of categorical variables will include the sample size (N) and/or frequency counts and percentages.

- Medical Dictionary for Regulatory Activities (MedDRA) nomenclature will be used to code all adverse events and medical history.

- World Health Organization Drug Dictionary Enhanced (WHO DDE) preferred name and MedDRA will be used to code all medications.

- Metric systems will be used for all applicable measures including (kg), height (cm), and clinical laboratory data (SI: Standard International units).

- Unless otherwise defined, baseline will be the last non-missing value of an analysis variable before treatment. Therefore, in general values collected prior to the treatment on day 1 will be used as baseline except for clinical laboratory tests which will be performed at screening instead of day 1. In the event a variable value is missing at the day 1 before treatment time point, the last non-missing value before day 1 will be used as baseline. Change from baseline is calculated as follow-up minus baseline value.

- Missing values will not be imputed. The multiple imputation (MI) method is only intended for sensitivity analyses of MAS-B ankle scores with knee extended and CGI by Physician.

- Study days from randomization/day 1 will be calculated as:

  Study days = visit date – randomization/day 1 visit date + 1.

- The study day of an event that has occurred before randomization/day 1 will be calculated as:

  Study day = date of event – randomization/day 1 visit date.

- Partial dates occurring when calculation of study day is germane to the study outcome will be handled as follows:

  If the year is missing then the date is treated as missing.

If either the month or the day is missing, then the closest date to baseline possible will be used. That is, an event occurring the year before baseline would use December and/or the last day of the month. An event occurring the year after baseline would use January and/or the first day of the month.

All partial dates will be listed "as is" in the data listings.

- All adverse events with missing onset dates will be identified and the missing dates will be imputed as follows for analysis. These imputations are only for assigning adverse events to a particular study phase; calculations such as days to onset will not use these imputations. All partial dates will be listed "as is" in the data listings.

  (a) For adverse event onset date: If day and month are missing but year is available (../../yy) then the imputed day and month will be 01/01/yy or the injection date if they have the same year, whichever is later. If the day is missing but the month and year are available (mm/../yy), then the imputed day will be mm/01/yy or the injection date if they have the same month and year, whichever is later.

  (b) For adverse event stop dates: If day and month are missing but year is available (../../yy) then the imputed day and month will be 12/31/yy or the study exit date if they have the same year, whichever is earlier. If day is missing but the month and year are available (mm/../yy), then the imputed day will be the last day of the month mm in the year yy or the study exit date if they have the same month and year, whichever is earlier.

  (c) If the adverse event onset or stop month, day, and year are missing (../../..), then the study entry date will be used as the adverse event onset date and the exit date will be used as the adverse event stop date.

- Data from unscheduled visits or re-test visits will be treated the same way as scheduled visits, ie, the dates of such visits will be used in the categorization of visit window. The algorithm handling multiple visits within a visit window will also be applied to unscheduled visits and re-test visits.

- For the primary and secondary efficacy analyses where study center is a factor in the analysis model, centers that recruited only a few patients will be pooled together. The following algorithm for pooling study centers will be used:

A. Pooling of study centers will by country.

B. Within each country, the following will be conducted:

1. Study centers will be ranked in descending order of the total number of patients enrolled.

2. Any center with at least 12 patients will stand on its own.

3. The first center with fewer than 12 patients will be combined with the next center until the total number in the pooled center reaches or exceeds 12 patients.

4. The algorithm of combining centers will continue down the list as described above.

5. If the last pooled center still has less than 12 patients, the last pooled center will be combined with the previous pooled center.

C. If there are less than 12 subjects enrolled in a country, the country will be pooled to the smallest center in the same region (alphabetical order if there is a tie). The regions include Asia-Pacific, Europe, and North America. For example, if Canada enrolled less than 12 patients, Canada will be pooled with a smallest center in the United States.

- The mean change from baseline for a given treatment group will be calculated by taking the sum of all individual changes from baseline divided by the number of patients with observations in that treatment group. As a consequence, patients without the particular follow-up data will not be taken into account, except if there is a missing value replacement method used.

[redacted]

3 Disposition and Exit Status

3.1 Screening Log Data

The number of subjects who are screen failures versus those who are qualified will be tabulated. The characteristics (sex, race, and age) of the failed and qualified subjects will be summarized. In addition, reasons for failure will be tabulated by “inclusion” and “exclusion”. The screen failures will be listed by site.

3.2 Disposition and Exit Status

A patient disposition and exit status table will be done for the all randomized subjects population. A cumulative frequency table showing patient disposition (enrolled, completed, and discontinued) in the study with sub-categorization by reason for premature discontinuation for each treatment group will be generated.

4 Demographics and Other Baseline Characteristics

4.1 Demographics

The demographics will be summarized for the mITT population.

Age, race, sex will be summarized using frequency tabulations and/or summary statistics, both by treatment group and overall patients. Race will be classified as Caucasian and non-Caucasian (black, Asian, Hispanic, and other). In addition, race will be displayed as collected in the case report form (Caucasian, black, Asian, Hispanic, and other).

4.2 Disease Characteristics

For the mITT population, the MAS-B ankle score with knee extended and knee flexed at baseline, the history of disease including the type (hemiplegia and monoplegia), onset, duration, and previous botulinum toxin exposure status will be summarized by treatment group for the mITT population.

4.3 Prior Medications

Using the mITT population, the number and percent of patients who had taken prior medication (as defined in Section 2.3) will be presented by treatment group for the base preferred (drug) name according to the World Health Organization Drug Dictionary Enhanced (WHO DDE) with the same active ingredients by the primary system organ class (SOC) and preferred term of the MedDRA code.

4.4 Concomitant Medications and Concurrent Procedures

Concomitant medications (as defined in Section 2.3) will be presented by WHO DDE preferred drug name for each treatment group by the primary SOC and preferred term of the MedDRA code using the mITT population. Concurrent procedures will be listed for the mITT population.

4.5 Medical History

Data for medical history with onset dates prior to the initial informed consent form signed date will be tabulated and presented by the SOC and preferred term of MedDRA code for each treatment group using the mITT population.

4.6 Pre-treatment Adverse Events

Data for pre-treatment adverse events, adverse events with onset date in-between informed consent date and randomization/day 1 visit or medical history that worsened duration this period will be listed.

4.7 Pre-treatment Serious Adverse Events

All pre-treatment serious adverse events will be flagged in the pre-treatment adverse events listing.

5 Efficacy Analyses

The primary, secondary, and other efficacy endpoints are listed below. The efficacy analysis for US FDA is described in Section 10.

Primary efficacy endpoints:

- The average change from baseline in MAS-B ankle score with knee extended at weeks 4 and 6
- The average CGI by Physician at weeks 4 and 6 (co-primary efficacy endpoint for US FDA only)

Secondary efficacy endpoints:

- The average CGI by Physician at weeks 4 and 6 (for non-US FDA analyses)
- Goal Attainment Scale (GAS) by Physician
- Modified Tardieu Scale (MTS) of the ankle with knee extended and knee flexed

[redacted]

[redacted]

5.1 Collection and Derivation of Efficacy Assessments

The MAS-B will be used to evaluate spasticity of each of the injected muscle (gastrocnemius, soleus, tibialis posterior) based on rating the resistance encountered in a specific muscle by means of passively moving a limb through its range of motion at a specified velocity.

At day 1 and weeks 2, 4, 6, 8 and 12, the same investigator (when possible) will assess the patient's ankle muscle tone with knee extended and knee flexed . The investigator will extend the patient's ankle from a position of maximum possible plantar flexion to a maximum possible dorsiflexion. The resistance encountered to the passive stretch will be rated on the MAS-B as follows:

0 =	No increase in muscle tone
1 =	Slight increase in muscle tone, manifested by a catch and release, or by minimal resistance at the end of the range of motion when the affected part(s) is moved in flexion or extension
1+ =	Slight increase in muscle tone, manifested by a catch, followed by minimal resistance throughout the remainder (less than half) of the range of motion
2 =	More marked increase in muscle tone through most of the range of motion, but affected part(s) easily moved. There can be a catch, but movement should be stiff through most of range.
3 =	Considerable increase in muscle tone, passive movement difficult
4 =	Affected part(s) rigid in flexion or extension

The MAS-B scores of 0, 1, 1+, 2, 3, or 4 will be coded as a grade of 0, 1, 2, 3, 4, or 5, respectively. The MAS-B ankle score change from baseline to each post-randomization visit and the primary endpoint, average change from baseline at weeks 4 and 6, will be derived for each patient. Additionally, a responder status will be determined according to whether or not a patient has at least a 1-grade reduction from baseline in MAS-B ankle score.

The CGI by Physician will be measured at weeks 2, 4, 6, 8, and 12. There is no CGI assessment at baseline. The CGI is a 9-point scale from -4 (very marked worsening) to +4 (very marked improvement). Each assessment is evaluated relative to the patient's baseline status obtained at day 1. The secondary endpoint for CGI will be the average of weeks 4 and 6 scores. Additionally,

a responder status will be determined according to whether or not a patient has at least +1 score of CGI by Physician.

The secondary efficacy measures of average CGI by Physician at weeks 4 and 6, GAS, and MTS will also be collected at scheduled office visits according to the Schedule of Visits and Procedures in the protocol.

The GAS is used to measure functional goal attainment relative to the effect of the study intervention. The achievement of 2 functional goals, one active and one passive, preset at week -2 by the patient and family in consultation with the investigator and/or treating physical therapist relative to the lower limb impairment due to spasticity and confirmed at day 1, will be assessed at weeks 8 and 12 using a scale of -3 (worse than start) to +2 (improvements clearly exceed the defined therapeutic goal).

For MTS, the difference between slow and fast range of motion (R2-R1) and respective change from baseline to each post-treatment office visit will be derived. The MTS of the ankle will be used to determine the passive range of movement at different movement velocities, V1 (as slow as possible) and V3 (as fast as possible) with the relative difference between a slow and a fast velocity passive stretch determining the dynamic component of the muscle contracture for the joint. At each visit, the same investigator, when possible will measure two joint angles by goniometer: the R1 angle which is the "angle of catch" after a fast velocity (V3) stretch and the R2 angle defined as the passive joint range of movement following a slow velocity (V1) stretch. The R2 – R1 value indicates the level of dynamic contracture in the joint. The difference between slow (R2) and fast (R1) range of motion and respective change from baseline to each post-treatment office visit on the MTS will be derived.

[redacted]

[redacted]

[redacted]

[redacted]

5.2 Primary Efficacy Analyses

5.2.1 Primary Analyses of Primary Efficacy Variable

The primary efficacy analyses will be based on the mITT population. The primary analysis will be conducted for the average of change from baseline in MAS-B of the ankle plantar flexors with

knee extended at weeks 4 and 6. For the assessment of treatment groups in the primary endpoint, the statistical null hypothesis is that both BOTOX higher dose (8 U/kg) and lower dose (4 U/kg) have the same mean as placebo. The alternative hypothesis is that at least one study dose of BOTOX has a different mean than placebo. The hypothesis will be tested using Mixed Model Repeated Measures (MMRM) at significance level of 0.05. The primary variable will be summarized by treatment group. The differences of each BOTOX dose group and the placebo group will be used to test the null hypothesis. The analysis will include weeks 2, 4, 6, 8, 12 data. The model will include baseline MAS-B ankle score with knee extended as a covariate and factors of age group, treatment group, visit, treatment-by-visit interaction, study center and previous botulinum toxin exposure. Age group is represented by stratification categories (≤ 6 years and > 6 years). Baseline MAS-B ankle score with knee extended is a continuous variable included as a covariate. Pooled study center will be produced to ensure each contains at least 12 patients. Pooling method is described in Section 2.3.

A gate-keeping approach will be used to control the type I error rate for the primary endpoint. Specifically, 8 U/kg versus placebo will be tested first followed by the 4 U/kg versus placebo comparison, each at 2-sided 0.05 level. The 4 U/kg versus placebo comparison will only be performed if the test for 8 U/kg versus placebo is statistically significant.

The co-primary analyses for US FDA only is described in Section 10.

5.2.2 Other Analyses of Primary Efficacy Variable

For sensitivity analysis purposes to supplement the primary analyses, an analysis of covariance (ANCOVA) model using MI for missing values will be conducted. A summary of the percentage

of missing MAS-B ankle scores with knee extended by visit will also be provided. The ANCOVA will include baseline MAS-B ankle score with knee extended as a covariate and factors of age group, treatment group, study center and previous botulinum toxin exposure.

The ANCOVA model will be used to analyze each imputation dataset obtained from MI as described in Section 5.1. [redacted] will then be used to pool the analyses results across the imputation datasets and produce final parameter estimates. Additional sensitivity analyses for the MAS-B ankle score with knee extended changes from baseline at each assessment time point will also be performed using the same ANCOVA as specified above with observed data, and linear contrasts will be used to estimate the treatment differences between each study dose and placebo.

In addition, the frequency distribution of the grade changes from baseline on the MAS-B ankle score with knee extended, as well as the percentage of responders (with at least a 1-grade reduction in MAS-B), will be summarized by treatment group at each post-treatment office visit. A logistic regression analysis will also be performed on the responder status at each post-treatment visit with a covariate of baseline MAS-B ankle score with knee extended and factors of age group, treatment group, study center, and previous botulinum toxin exposure, if applicable. The additional analyses will be conducted using observed cases unless stated otherwise.

5.3 Secondary Efficacy Analyses

For secondary efficacy analyses, the observed cases will be used unless stated otherwise.

5.3.1 CGI by Physician

CGI by Physician will be rated at study weeks 2, 4, 6, 8, and 12. Descriptive statistics will be presented by treatment group at each assessment timepoint and for the average of weeks 4 and 6. CGI by Physician will be summarized and analyzed by MMRM, as well as ANCOVA at each assessment timepoint using observed cases. MMRM model will include baseline MAS-B ankle score with knee extended as a covariate and factors of age group, treatment group, visit, treatment-by-visit interaction, study center and previous botulinum toxin exposure. The frequency distribution of the CGI scores, as well as the percentage of responders (CGI score ≥ 1), will be summarized by treatment group at each post-treatment visit. A logistic regression analysis will also be performed on the responder status at each post-treatment visit with factors of age group, treatment group, study center, and previous botulinum toxin exposure, if applicable.

5.3.2 Global Attainment Scale by Physician

Each of the 2 goals for each patient at each visit will be analyzed by an ANCOVA model with age group, treatment group, study center and previous botulinum toxin exposure as factors and baseline MAS-B ankle score with knee extended as a covariate to assess treatment differences. The frequency distribution of raw scores for each goal will be summarized by treatment group. Additionally, the percentages of responders who scored 0 or higher (meet or exceed expectation) and the responders who scored -1 or high (slight improvement but not meet expectation) on each functional goal will also be summarized and responder status will be analyzed by logistic regression with factors of age group, treatment group, study center and previous botulinum toxin exposure, if applicable.

5.3.3 Modified Tardieu Scale (MTS) of Ankle

The change from baseline of the angle difference, R2-R1, and the individual R1and R2 respective changes from baseline will be derived at each post-treatment visit and summarized by treatment group, excluding observations where R1>R2. An ANCOVA model with age group, treatment group, study center and previous botulinum toxin exposure as factors and baseline MTS of ankle with knee extended as a covariate will be performed to assess treatment differences. The raw scores of R2-R1, R2 and R1 will also be summarized and analyzed using similar methods. MTS of ankle with knee flexed will be analyzed using the same method as MTS of ankle with knee extended.

[redacted]

[redacted]

[redacted] [redacted]

[redacted]

[redacted] [redacted]

[redacted]

[redacted] [redacted]

[redacted]

[redacted] [redacted]

[redacted]

5.5 Subgroup Analyses for Efficacy Variables

Subgroup analyses by the stratification factors of age (≤ 6 years and > 6 years), baseline MAS-B score (2 and > 2) and previous exposure to botulinum toxin will be performed for primary efficacy variable(s). The "None" subgroup for previous exposure to botulinum toxin is defined as the patients who never received botulinum toxin in the past.

6 Safety Analyses

Safety variables including the incidence of adverse events, the incidence of suicide-related events using the C-SSRS (for patients ≥ 6 years of age at day 1), lab tests values and their changes from baseline, as well as vital signs and their change from baseline, immunogenicity tests will be summarized by treatment group using the safety population.

6.1 Exposure to Study Treatments

Patient exposure to the study medication will be characterized by total dose in U, and dose in U/kg body weight.

Duration of follow up is defined as the number of days from the first injection to exit; if the date of exit is missing, the date of the last visit will be used. Data will be summarized with descriptive statistics for each treatment group.

In addition, the total dose in U and dose in U/kg body weight will be summarized by individual muscle, as well as for all muscles combined.

6.2 Treatment Emergent Adverse Events

Adverse events are collected both for the screening/baseline period prior to treatment (which are referred to as pre-treatment adverse events as cited in Section 4.6) and for the follow-up period after treatment is initiated.

A treatment-emergent adverse event (TEAE) is a post-baseline adverse event with onset after the initiation of study treatment or an adverse event with onset prior to study treatment that worsened in severity or became serious after the initiation of study treatment. The number and percentage of patients reporting each event will be summarized and presented for all events and tabulated by treatment groups (8 U/kg, 4 U/kg, and placebo) and overall as follows:

i. By primary SOC and preferred term. At each level of summarization (overall, primary SOC, and preferred term) a patient will be counted once if he/she reports 1 or more experiences at that level.

ii. By primary SOC, preferred term and severity. For a given adverse event of a patient, if more than 1 severity grade was reported, the worst severity grade will be included in the tabulation.

All treatment-related TEAEs (study drug-related and injection procedure-related TEAE) will be presented by primary SOC and preferred term. A listing of adverse events leading to discontinuation will also be provided.

The number and percentage of patients with any of the preferred terms will be summarized for each primary SOC and preferred term. All patients will be counted only once for each preferred term when multiple occurrences of the same preferred term are reported.

In addition, individual patient data listings (including adverse event listings that include the investigator verbatim terms and MedDRA lower level terms) will be provided.

Furthermore, the assessment of possible distant spread of toxin will be performed using the list of 39 MedDRA preferred terms (version 19.0 or its updated version as appropriate during the time of data analysis).

MedDRA Preferred Terms Evaluated for Possible Distant Spread of Toxin

Cardiac Disorders
- Bradycardia

Eye Disorders
- Accommodation disorder
- Diplopia
- Extraocular muscles paresis
- Eyelid function disorder
- Eyelid ptosis
- Pupillary reflex impaired
- Vision blurred

Gastrointestinal Disorders
- Constipation
- Dry mouth
- Dysphagia
- Ileus paralytic

Infections and Infestations
- Botulism

Musculoskeletal and Connective Tissue Disorders
- Muscular weakness

Nervous System Disorders
- Bulbar palsy
- Cranial nerve palsies multiple
- Cranial nerve paralysis
- Dysarthria
- Facial paralysis
- Facial paresis
- Hyporeflexia
- Hypotonia
- Paralysis
- Paresis cranial nerve
- Peripheral nerve palsy
- Peripheral paralysis
- Speech disorder
- Vocal cord paralysis
- Vocal cord paresis

Renal and Urinary Disorders
- Urinary retention

Respiratory, Thoracic and Mediastinal Disorders
- Aspiration
- Diaphragmatic paralysis
- Dysphonia
- Dyspnoea
- Pneumonia aspiration
- Respiratory arrest
- Respiratory depression
- Respiratory failure

Reproductive System and Breast Disorders
- Pelvic floor muscle weakness

6.3 Serious Adverse Events

The number and percent of patients with serious TEAEs will be tabulated by treatment groups (8 U/kg, 4 U/kg, and placebo) and overall for all serious adverse events regardless of relation to treatment by primary SOC and preferred term.

6.4 Clinical Laboratory Evaluations

Hematology and non-fasting clinical chemistry assays will be obtained at screening and exit as specified below.

Hematology	hemoglobin	hematocrit	red blood cells
	white blood cells	neutrophils	bands
	lymphocytes	monocytes	basophils
	eosinophils	platelets	
Electrolytes	sodium	potassium	bicarbonate
	chloride	phosphorus	calcium
Chemistry	glucose	creatinine	urea nitrogen
	total bilirubin	alkaline phosphatase (total plus bone and liver fractions)	aspartate aminotransferase (serum glutamic oxalacetic transaminase)
	alanine aminotransferase (serum glutamic pyruvic transaminase)	albumin	
Other	hemoglobin A1c	25 hydroxyl vitamin D	Parathyroid hormone if alkaline phosphatase bone fraction is abnormal

These values at each visit and the change from baseline will be summarized by the treatment group by visits using descriptive statistics. All laboratory values will be listed. Laboratory values at each visit will also be categorized as low, normal, and high according to the respective reference normal ranges.

6.5 Vital Signs

Vital signs are assessed at screening, day 1, and exit. Blood pressure (diastolic and systolic measured in mmHg), pulse rate (beats/minute), respiratory rate (breaths/minute), and body temperature (°C), weight (kg) will be summarized by treatment group for each visit and change from baseline to exit visit will also be summarized by treatment group.

6.6 Columbia-Suicide Severity Rating Scale

Suicidality for age over 6 years will be assessed using the Columbia-Suicide Severity Rating Scale (C-SSRS), conducted at every assessment visit. Suicide-related events corresponding to the C-SSRS questions will be classified into the following 11 categories:

- Self-injurious behavior, no suicidal intent
- Suicidal ideation
    1. Wish to be dead
    2. Active: Nonspecific (no method, intent, or plan)
    3. Active: Method, but no intent or plan
    4. Active: Method and intent but no plan
    5. Active: Method, intent and plan
- Suicidal behavior
    1. Preparatory acts or behaviors
    2. Aborted attempt
    3. Interrupted attempt
    4. Suicide attempt
    5. Completed suicide

The number and percentage of patients with the above suicide-related events during treatment will be presented by treatment group. If a patient has more than one suicide-related event during an interval, the most severe event in increasing order listed above will be used.

6.7 Immunogenicity Testing

Blood samples (approximately 7 mL per visit) will be collected for immunogenicity testing at the screening visit and study exit from patients who weigh at least 15 kg at the screening visit. A 2-stage assay approach will be used for the detection of binding antibodies against BoNT-A and neutralizing antibodies against BOTOX in patients' serum. In stage 1, serum samples will be screened for reactivity using a validated enzyme-linked immunosorbent assay (ELISA). The reactive serum samples will be subsequently titered to assess the extent of antibodies present and immunodepleted to confirm that the reactivities are specific to BoNT-A. In stage 2, only samples that are confirmed positive in the ELISA will be tested for neutralizing antibodies to BOTOX using a validated assay.

Immunogenicity results of binding antibodies (BABs) against BoNT-A will be reported as positive or negative. Immunogenicity results of neutralizing antibodies (NABs) against BOTOX will be reported as positive, negative, or inconclusive. For NABs, the total negatives including patients with negative neutralizing antibody results as well as with negative binding antibody results will be reported.

Immunogenicity findings will be tabulated with the number and percent of patients at each visit separately for each treatment group. Percentages will be based on the number of treated patients with interpretable antibody assays in each treatment group at the specified visit. Antibody titers will be summarized by median, minimum, and maximum.

Depending on the timing and availability of data, immunogenicity data may be locked separately, at a different time point following the database lock for the main clinical data.

6.8 Other Safety Analyses

6.8.1 Physical Examination

Physical examination will be performed at screening and study exit. Abnormal findings will be listed.

6.8.2 Pregnancy Test

The urine pregnancy test will be performed for females of childbearing potential at screening, day 1 prior to injection and at study exit. Positive pregnancy test results (any pregnancies will be reported in the clinical study report) will be listed.

7 Interim Analyses

No interim analysis is planned for this study.

8 Pharmacokinetic Data Analyses

There will be no pharmacokinetic data to be analyzed from this study.

9 Health Outcomes Data Analyses

There will be no health outcomes data to be analyzed from this study.

10 Analysis for US FDA HFD-550

Analyses are as described in all preceding sections except for the items described below.

10.1 Primary Efficacy Analyses

There are 2 co-primary endpoints: the average grade change from baseline in the MAS-B ankle score with knee extended at weeks 4 and 6 and the average CGI by Physician at weeks 4 and 6. For both co-primary endpoints, the statistical null hypothesis is that both the higher dose (8 U/kg) and the lower dose (4 U/kg) have the same mean as placebo. The alternative hypothesis is that at least one study dose has a different mean than placebo. In addition to the primary statistical analyses cited in Section 5.2.1 for the average grade change from baseline in MAS-B ankle score with knee extended at weeks 4 and 6, the average CGI by Physician at weeks 4 and 6 will be analyzed using the same statistical method with a same MMRM model in Section 5.3.1.

The Hochberg procedure (Hochberg and Tamhane, 1987) will be used to control the family-wise type I error rate. With 2 doses and 2 co-primary endpoints in each dose, the maximum p-value for the 2 co-primary endpoints in each dose will be used in the Hochberg procedure. We first define the following values:

p11: p value for Botox 8 U/kg vs placebo comparing MAS-B

p12: p value for Botox 4 U/kg vs placebo comparing MAS-B

p21: p value for Botox 8 U/kg vs placebo comparing CGI

p22: p value for Botox 4 U/kg vs placebo comparing CGI

p1 = max(p11, p21), p2 = max(p12, p22).

p1 and p2 will be sorted in an increasing order to get $p(1) \leq p(2)$. The following decision rule will then be applied:

Step 1: If $p(2) \leq 0.05$, both doses are considered efficacious; otherwise go to step 2.

Step 2: If $p(1) \leq 0.025$, its corresponding dose is considered efficacious; otherwise go to step 3.

Step 3: Neither dose is considered efficacious.

For sensitivity analysis purposes to supplement the co-primary analyses, ANCOVA using MI for missing values will be conducted for the co-primary endpoints of MAS-B ankle score with knee extended and CGI. The model will include baseline MAS-B ankle score with knee extended as a covariate and factors of age group, treatment group, study center, and previous botulinum toxin exposure. A summary of the percentages of missing MAS-B and CGI scores by visit will also be provided. Additional sensitivity analyses for the MAS-B ankle score with knee extended changes from baseline and CGI scores at each assessment timepoint will also be performed using the same ANCOVA with observed data, and linear contrasts will be used to estimate the treatment differences between each study dose and placebo.

10.2 Secondary Efficacy Analyses

The secondary efficacy analyses will be the same as the statistical analyses cited in Section 5.3 except for the analyses for CGI by Physician which is described in the Section 10.1.

11 Data Collected but not Analyzed

The date and signature of investigator and system required variables of electronic data capture (EDC) and IVRS/IWRS will not be analyzed.

12 Deviations from Protocol

There are no deviations from protocol amendment 3.

13 References

Hochberg Y, Tamhane A (1987). Multiple Comparison Procedures, John Wiley, New York.

14 Changes to Analyses

Changes to Analyses from Previous Version of Statistical Analysis Plan

The changes to the analyses as specified in the last approved version of the statistical analysis plan, dated 01 NOV 2016, are as follows:

- Disposition summary will be performed for all randomized patients.
- Clarification on pooling for countries with less than 12 subjects enrolled. The country will be pooled to the smallest center in the same region (alphabetical order if there is a tie).

ALLERGAN

191622-111 Statistical Analysis Plan

Date (DD/MMM/YYYY)/Time (PT)	Signed by:	Justification
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]