Statistical Analysis Plan for Study M21-310

A Phase 3 Multicenter, Randomized, Double-blind, Placebo-controlled Study to Evaluate the Safety and Efficacy of BOTOX® (Botulinum Toxin Type A) Purified Neurotoxin Complex for the Treatment of Platysma Prominence

Date: 20 June 2023

Version 4.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for Study M21-310: A Phase 3 Multicenter, Randomized, Double-blind, Placebo-controlled Study to Evaluate the Safety and Efficacy of BOTOX (Botulinum Toxin Type A) Purified Neurotoxin Complex for the Treatment of Platysma Prominence.

Study M21-310 examines the safety and efficacy of BOTOX or botulinum toxin type A in subjects.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

Unless noted otherwise, all analyses will be performed using SAS Version 9.4 (SAS Institute Inc., Cary, NC 27513) or later under the Linux operating system.

2.0 Study Design and Objectives

2.1 Objectives, Hypotheses and Estimands

The study objective is to compare the safety and efficacy of BOTOX with placebo for the treatment of platysma prominence in adult subjects with moderate to severe platysma prominence at maximum contraction.

The clinical hypothesis is that BOTOX is a safe and effective treatment for the temporary improvement in the appearance associated with moderate and severe platysma prominence in adults.

Estimands: Primary Endpoints

The attributes of the estimands corresponding to the primary efficacy endpoints are summarized in Table 1 for the United States (US) and the European Union (EU) regulatory agencies. In addition, the attribute of treatment is a single dose of BOTOX [redacted] or placebo.

Table 1. Summary of the Estimand Attributes of the Primary Efficacy Endpoints

	Attributes of the Estimand			
Estimand Label	**Variable (Endpoint)**	**Population**	**Handling of Intercurrent Events**	**Statistical Summary**
Hypothetical estimand for primary composite endpoint [redacted]	Achievement of Grade 1 or 2 (Minimal or Mild) and at least a 2-grade improvement from baseline based on both investigator's assessment using the C-APPS and subject's self-assessment using the P-APPS at maximum contraction at Day 14	intent-to-treat (ITT) (all randomized)	Subjects who discontinue study prior to Day 14 assessments or who do not have Day 14 C-APPS or P-APPS assessments will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Response rates and rate differences between BOTOX and placebo treatment groups; Cochran-Mantel-Haenszel (CMH) test [redacted]
Hypothetical estimand for coprimary endpoints [redacted]	Achievement of at least a 2-grade improvement from baseline based on the following: • Investigator's assessment using C-APPS at maximum contraction at Day 14 • Subject's self-assessment using P-APPS at maximum contraction at Day 14	Modified intent-to-treat (mITT) [redacted]	Subjects who discontinue study prior to Day 14 assessments or who do not have Day 14 C-APPS or P-APPS assessments will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Response rates and rate differences between BOTOX and placebo treatment groups; CMH test [redacted]

Estimands: Secondary Endpoints

The attributes of the estimands corresponding to the secondary efficacy endpoints are summarized in Table 2. Treatment is the same as for the primary efficacy endpoints. For the US FDA variables/endpoints, the population is the ITT population (all randomized); for the EU variables/endpoints, the population is the mITT population [redacted]. The variables/endpoints listed have the same handling of intercurrent events and statistical

summary (including population-level summary and analysis methods) within their respective analysis populations for the US FDA and EU, per estimand label.

Table 2. Summary of the Estimand Attributes of the Secondary Efficacy Endpoints

	Attributes of the Estimand		
Estimand Label	**Variables (Endpoints)**	**Handling of Intercurrent Events**	**Statistical Summary**
Hypothetical estimand for secondary categorical endpoint [redacted]	Achievement of Grade 1 or 2 (Minimal or Mild) according to investigator's assessment using C-APPS at maximum contraction over time*	Subjects who discontinue study or who do not have C-APPS assessments will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Response rates and rate differences between BOTOX and placebo treatment groups; CMH test [redacted]
Hypothetical estimand for secondary categorical endpoint [redacted]	Achievement of Grade 1 or 2 (Minimal or Mild) according to subject's self-assessment using P-APPS at maximum contraction over time*	Subjects who discontinue study or who do not have P-APPS assessments will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Response rates and rate differences between BOTOX and placebo treatment groups; CMH test [redacted]
Hypothetical estimand for secondary categorical endpoint [redacted]	Achievement of a rating of Minimal or Mild according to subject's self-assessment using P-APPS at maximum contraction at Day 14	Subjects who discontinue study prior to Day 14 assessments or who do not have Day 14 P-APPS assessments will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Response rates and rate differences between BOTOX and placebo treatment groups; CMH test [redacted]

Estimand Label	Attributes of the Estimand		
	Variables (Endpoints)	**Handling of Intercurrent Events**	**Statistical Summary**
Hypothetical estimand for secondary categorical endpoints [redacted]	Responses of *Satisfied* or *Very satisfied* on the ANLFQ: Satisfaction (Follow-up) Item 5 (effect of treatment) at Day 14	Subjects who discontinue study prior to Day 14 assessments or who do not have Day 14 ANLFQ: Satisfaction (Follow-up) Item 5 assessments will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Response rates and rate differences between BOTOX and placebo treatment groups; CMH test [redacted]
Hypothetical estimand for secondary categorical endpoints [redacted]	Responses of *Not at all bothered* or *A little bothered* on the BAS-PP Item 2 (jawline) at Day 14	Subjects who discontinue study prior to Day 14 assessments or who do not have Day 14 BAS-PP Item 2 assessments will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Response rates and rate differences between BOTOX and placebo treatment groups; CMH test [redacted]
Hypothetical estimand for secondary categorical endpoints [redacted]	Responses of *Not at all bothered* or *A little bothered* on the BAS-PP Item 1 (vertical neck bands) at Day 14	Subjects who discontinue study prior to Day 14 assessments or who do not have Day 14 BAS-PP Item 1 assessments will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Response rates and rate differences between BOTOX and placebo treatment groups; CMH test [redacted]
Hypothetical estimand for secondary continuous endpoints [redacted]	Change from baseline on the ANLFQ: Impacts summary score at Day 14	Subjects who discontinue study prior to Day 14 assessment of ANLFQ: Impacts summary score will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Mean change from baseline and mean differences between BOTOX and placebo treatment groups; analysis of variance [redacted]

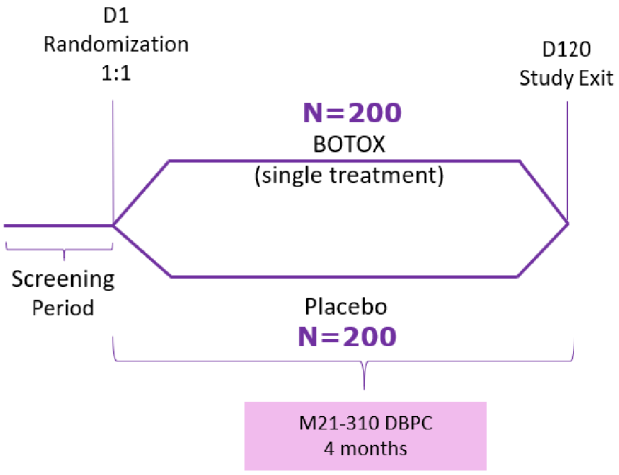
	Attributes of the Estimand		
Estimand Label	**Variables (Endpoints)**	**Handling of Intercurrent Events**	**Statistical Summary**
Hypothetical estimand for secondary continuous endpoints [redacted]	Change from baseline on the ANLFQ: Impacts summary score at Days 30, 60, and 90	Subjects who discontinue study prior to applicable timepoints or who do not have Day 30, 60, and 90 ANLFQ: Impacts summary score will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Mean change from baseline and mean differences between BOTOX and placebo treatment groups; analysis of variance [redacted]
Hypothetical estimand for secondary categorical endpoint [redacted]	Achievement of at least a 1-grade improvement from baseline based on the investigator's assessment using C-APPS at maximum contraction over time*	Subjects who discontinue study or who do not have C-APPS assessments will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Response rates and rate differences between BOTOX and placebo treatment groups; CMH test [redacted]
Hypothetical estimand for secondary categorical endpoint [redacted]	Achievement of at least a 1-grade improvement from baseline based on the subject's self-assessment using P-APPS at maximum contraction over time*	Subjects who discontinue study or who do not have P-APPS assessments will be included in the analysis as a hypothetical scenario in which the assessments were conducted as per protocol	Response rates and rate differences between BOTOX and placebo treatment groups; CMH test [redacted]

[redacted]

2.2 Study Design Overview

The schematic of the study is shown in Figure 1 for this study (DBPC [double-blind placebo-controlled]).

Figure 1. Study Schematic



2.3 Treatment Assignment and Blinding

Subjects will be randomized to BOTOX or placebo in a 1:1 ratio. Randomization will be stratified at each investigational site by baseline C-APPS on Day 1 [redacted] The same block will not be shared across sites or by baseline C-APPS.

2.4 Sample Size Determination

Approximately 400 subjects will be randomized into the study in a 1:1 ratio yielding approximately 200 subjects in the BOTOX group and 200 subjects in the placebo group in the treatment period. [redacted]

The primary analysis for FDA will be performed on the ITT population. For the US FDA primary analysis, an estimated sample size of 360 subjects will give a power of greater than 99% to detect a difference in responder rates between the BOTOX and placebo

groups, assuming a 10% dropout rate by Day 14. [redacted]

The primary analysis for EU regulatory agencies will only be performed on the mITT population, [redacted] a sample size of 180 will result in a power of greater than 99% to detect a difference between BOTOX and placebo in the proportion of subjects achieving at least a 2-grade improvement from baseline according to the investigator's assessment using C-APPS at maximum contraction at Day 14 and according to the subject's self-assessment using P-APPS at maximum contraction at Day 14 (i.e., coprimary endpoints). [redacted]

Sample size calculations were based on the two group χ^2 two-sided test of equal proportions using nQuery (Version 7.0) at a 0.05 level of significance.

3.0 Endpoints

3.1 Primary Endpoints

For the FDA, the primary composite efficacy endpoint is achievement of Grade 1 or 2 (Minimal or Mild) and at least a 2-grade improvement from baseline based on both investigator's assessment using the C-APPS and subject's self-assessment using P-APPS at maximum contraction at Day 14, that is, Grade 1 or 2 with at least a 2-grade improvement on the C-APPS on Day 14 AND Grade 1 or 2 with at least a 2-grade improvement on the P-APPS on Day 14.

For EU regulatory agencies, the coprimary efficacy endpoints are the achievement of at least a 2-grade improvement from baseline based on the following: (1) investigator's assessment using C-APPS at maximum contraction at Day 14, and (2) subject's self-assessment using P-APPS at maximum contraction at Day 14.

3.2 Secondary Endpoints

For US FDA:

- Achievement of Grade 1 or 2 (Minimal or Mild) according to investigator's assessment using C-APPS at maximum contraction over time
- Achievement of Grade 1 or 2 (Minimal or Mild) according to subject's self-assessment using P-APPS at maximum contraction over time
- Responses of *Satisfied* or *Very satisfied* on the ANLFQ: Satisfaction (Follow-up) Item 5 (effect of treatment) at Day 14
- Responses of *Not at all bothered* or *A little bothered* on the BAS-PP Item 2 (jawline) at Day 14Responses of *Not at all bothered* or *A little bothered* on the BAS-PP Item 1 (vertical neck bands) at Day 14
- Change from baseline on the ANLFQ: Impacts summary score at Day 14
- Achievement of at least a 1-grade improvement from baseline based on investigator's assessment using C-APPS at maximum contraction over time
- Achievement of at least a 1-grade improvement from baseline based on the subject's self-assessment using P-APPS at maximum contraction over time

For EU regulatory agencies:

- Achievement of a rating of Minimal or Mild according to subject's self-assessment using P-APPS at maximum contraction at Day 14
- Responses of *Satisfied* or *Very satisfied* on the ANLFQ: Satisfaction (Follow-up) Item 5 (effect of treatment) at Day 14
- Responses of *Not at all bothered* or *A little bothered* on the BAS-PP Item 2 (jawline) at Day 14

- Responses of *Not at all bothered* or *A little bothered* on the BAS-PP Item 1 (vertical neck bands) at Day 14
- Change from baseline on the ANLFQ: Impacts summary score at Day 14
- Change from baseline on the ANLFQ: Impacts summary score at Days 30, 60, and 90

3.3 Other Efficacy Endpoints

3.4 Safety Endpoint

The safety endpoint is the incidence of AEs for the entire study population.

3.5 Additional Endpoints

Not applicable.

4.0 Analysis Populations

The following population sets will be used for the analyses.

The Intent-to-Treat (ITT) Population includes all randomized subjects. Subjects will be included in the analysis according to the treatment groups to which they were randomized.

Baseline analyses and efficacy analyses for US FDA will be performed on the ITT population, consisting of all randomized subjects.

For EU regulatory agencies only, baseline analyses and efficacy analyses will be performed on the mITT population, [redacted] Subjects will be included in the analysis according to the treatment groups to which they were randomized.

The Safety Analysis Set consists of all subjects who were treated with at least 1 dose of study drug. All safety analyses will be performed with subjects analyzed by their actual treatment received. The Safety Analysis Set will be used for all safety analyses.

5.0 Subject Disposition

The total number of subjects who were screened, enrolled (randomized), and treated will be summarized for ITT and mITT populations separately. Reasons for exclusion, including screen failure, will be summarized.

A summary of subject accountability will be provided where the number of subjects in each of the following categories will be summarized for each treatment group:

- Subjects enrolled (randomized) in the study;
- Subjects treated with at least one dose of study drug;
- Subjects who completed protocol-specified treatment;
- Subjects who prematurely discontinued study (with reason);
- Subjects in each analysis population.

For end of study participation, the number and percentage of subjects who completed the protocol defined follow-up period will be summarized overall and by treatment group.

6.0 Study Drug Duration and Compliance

For the safety population, the number of subjects treated will be presented by treatment group. If a subject does not receive the full dose, this will be indicated; significant deviations to dosing will be reported.

7.0 Demographics, Baseline Characteristics, Medical History, and Prior/Concomitant Medications

Demographics, baseline or disease characteristics, medical history, and prior and concomitant medications will be summarized for the ITT and mITT populations overall and by treatment group. Categorical variables will be summarized with the number and percentage of subjects; percentages will be calculated based on the number of non-missing observations. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum, and maximum).

7.1 Demographics and Baseline Characteristics

Continuous demographic variables include age, weight, height, and body mass index (BMI). Categorical demographic and baseline variables include sex; ethnicity; region (North America, Europe); race; age group [redacted]; BMI [redacted] C-APPS grade; P-APPS grade; Fitzpatrick Skin Phototype; BAS-PP Items 1 and 2; ANLFQ impact, satisfaction, and treatment expectation summary scores; and PGIS.

7.2 Medical History and Prior and Concurrent Procedures

Medical history data will be reported and summarized separately from prior procedures data for the ITT and mITT populations. Medical history data and prior procedure data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term) will be summarized overall and by treatment group. Similarly, the number and percentage of subjects in each prior procedures category will be summarized overall and by treatment group by high level term (HLT) and PT. The HLT will be presented in alphabetical order, and the preferred terms (PT) will be presented in alphabetical order within each SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (HLT or preferred term).

In addition, any concurrent procedures, defined as any procedure performed on or after the date of first treatment, will be summarized by HLT and PT in each treatment group for the ITT and mITT populations.

7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name for the ITT and mITT populations. A prior medication is defined as any medication taken prior to the date of the first dose of study drug. A concomitant medication is defined as any medication that started prior to the date of the first dose of study drug and continued to be taken after the first dose of study drug or any medication that started on or after the date of the dose of study drug, but not started after the date of the last visit assessment. The number and percentage of subjects taking medications will be summarized by generic drug name based on the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications.

8.0 Handling of Potential Intercurrent Events for the Primary and Key Secondary Endpoints

The primary composite endpoint for the FDA and the coprimary endpoints for the EU regulatory agencies (defined in Section 3.1) will be analyzed based on the ITT population and the mITT population, respectively, and the following methods will be used to address potential intercurrent events:

- Subjects who did not receive any dose of study drug but are randomized will still be included in the ITT population. [redacted]
- Subjects who are randomized but prematurely discontinued the study before assessment of the primary endpoints will be considered as part of the ITT population. [redacted]
- Subjects who die before assessment of the primary endpoint will count as though they hypothetically continued in the study.
- Subjects who are lost to follow-up and are missing data for the primary endpoint will count as though they hypothetically continued in the study.
- Subjects who are missing assessments or data due to the COVID-19 pandemic and are missing data for the primary endpoint will count as though they hypothetically continued in the study.
- Subjects who are missing data for any other reason for the primary endpoint will count as though they hypothetically continued in the study.

The secondary endpoints (defined in Section 3.2) will be analyzed based on the same populations as above with similar methods for addressing potential intercurrent events.

9.0 Efficacy Analyses

9.1 General Considerations

Efficacy analyses for primary and secondary endpoints will be conducted in the ITT Population for the US and the mITT Population for the EU regulatory agencies. All other efficacy analyses will be conducted using the ITT population.

Analyses will be performed using a gated hierarchical testing procedure to preserve a family-wise Type I error rate of $\alpha = 0.05$.

To control the family-wise Type I error rate at 0.05 for multiplicity across the primary and secondary endpoints, a hierarchical testing strategy[1] will be used. For the US, the secondary endpoints assessed "over time" (see Section 3.2) will be excluded from the hierarchical testing strategy. [redacted] Nominal p-values will be provided thereafter.

The final analysis will be performed after all ongoing subjects have completed the study and the database has been locked. This will be the only analysis for the primary composite, coprimary and secondary efficacy endpoints, as well as all other efficacy endpoints.

Unless otherwise specified, categorical variables will be analyzed using Cochran-Mantel-Haenszel (CMH) test, [redacted]. Continuous variables will

be analyzed using analysis of covariance (ANCOVA) [redacted]

Unless otherwise specified, any subject who is randomized based on a wrong stratum will be analyzed according to the subject's actual stratum.

"Baseline" refers to the last non-missing observation before the first administration of study drug or randomization if no study drug is given. The change from baseline values will be computed as the post-baseline value minus the baseline value.

9.2 Handling of Missing Data

Missing data will be imputed using the following methods for the efficacy analyses:

- Multiple Imputation: The Multiple Imputation approach will be used as a primary analysis for the primary endpoint, as well as for all secondary endpoints. For the primary endpoint (US: primary composite endpoint; EU: coprimary endpoints) and the primary and secondary measures through Day 120 [redacted]

- Non-responder Imputation (NRI): The NRI analysis will categorize any subject who does not have evaluation during a specific visit window as a non-responder for that visit. The NRI will be one of the sensitivity analysis approaches in the analyses of categorical variables for efficacy.
- As Observed (AO): The AO analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the AO analysis for that visit. AO will include all values collected in the study and will be used for sensitivity analyses of efficacy endpoints.

9.3 Primary Efficacy Endpoints and Analyses

9.3.1 Primary Efficacy Endpoints

Clinician Allergan Platysma Prominence Scale (C-APPS): The sponsor developed the C-APPS as a clinician's assessment tool for evaluation of platysma prominence at maximum contraction.

Prior to enrolling subjects, investigators (or sub-investigators when applicable) will be trained in grading platysma prominence severity using the C-APPS.

The C-APPS is to be performed by the same trained and qualified (MD/DO) evaluator throughout the study [redacted]

Platysma prominence [redacted] will be assessed live, independently, at maximum contraction by the investigator using the C-APPS (1 = Minimal, 2 = Mild, 3 = Moderate, 4 = Severe, 5 = Extreme).

There are no COVID-19 protocol modifications for the C-APPS procedure. If a site visit is missed at which this assessment was planned, the assessment will be skipped for that timepoint.

Participant Allergan Platysma Prominence Scale (P-APPS): The P-APPS has been developed as a subject self-assessment tool for evaluation of platysma prominence at maximum contraction [redacted]

Subjects will use the P-APPS (1 = Minimal, 2 = Mild, 3 = Moderate, 4 = Severe, 5 = Extreme) [redacted] to self-assess the severity of their left and right platysma prominence at maximum contraction (left and right ratings captured separately).

There are no COVID-19 protocol modifications for the P-APPS procedure. If a site visit is missed at which this assessment was planned, the assessment will be skipped for that timepoint.

[redacted]

[redacted]

9.3.2 Main Analysis of Primary Efficacy Endpoints

The attributes of the estimands corresponding to the primary composite and coprimary efficacy endpoints for the US and EU regulatory agencies, respectively, are summarized in Table 1 (see Section 2.1).

Analysis of the primary endpoint will be conducted on the ITT population based on treatment as randomized for the US FDA analyses and on the mITT population based on treatment as randomized for the EU regulatory agencies.

US FDA: The primary responder composite endpoint for an achievement of Grade 1 or 2 (Minimal or Mild) and at least a 2-grade improvement in both C-APPS and P-APPS will be analyzed using the CMH method [redacted]

EU regulatory agencies: The coprimary responder endpoints for at least a 2-grade improvement in the C-APPS and P-APPS will be analyzed separately using the CMH method [redacted]

The evaluation of the equality of the proportions of responders for the primary endpoint at Day 14 will be based on the CMH test [redacted] The 95% CI for the treatment

responder rate differences will be constructed using the normal approximation to the binomial distribution.

9.3.3 Sensitivity and Supplementary Analyses of the Primary Efficacy Endpoints

Sensitivity analyses of the primary efficacy variables will be performed to establish their consistency and robustness as well as to further characterize the extent of subjects' responses. As-observed data analysis, as well as missing as non-responder (using NRI) analysis [redacted].

9.4 Secondary Efficacy Endpoints and Analyses

9.4.1 Secondary Efficacy Endpoints

Bother Assessment Scale - Platysma Prominence (BAS-PP): BAS-PP is a 2-item PRO measure that asks subjects to rate how bothered they are by the appearance of their vertical neck bands (Item 1) and jawline (Item 2) [redacted]

Appearance of Neck and Lower Face Questionnaire (ANLFQ): ANLFQ-Impacts is a 7-item PRO measure that assesses the psychosocial impact of the appearance of the neck and lower face from the subject perspective. [redacted] Scores from the individual items are summed [redacted] where higher scores indicate greater negative impact from the appearance of the neck and lower face.

ANLFQ-Satisfaction is a PRO measure that assesses treatment expectations and satisfaction with platysma prominence from the subject perspective. [redacted]

[redacted]

[redacted]

9.4.2 Main Analyses of Secondary Efficacy Endpoints

The attributes of the estimands corresponding to the secondary efficacy endpoints are summarized in Table 2 (see Section 2.1).

Analysis of the secondary endpoints will be conducted on the ITT population based on treatment as randomized for the US FDA analyses and on the mITT population based on treatment as randomized for the EU regulatory agencies.

For the secondary endpoints, the proportion of responders will be analyzed using the CMH test [redacted]

9.4.3 Sensitivity and Supplementary Analyses for Secondary Efficacy Endpoints

Additional sensitivity analyses may be performed as needed to assess impact of missing data for secondary endpoints due to a pandemic or natural disaster.

9.4.4 Supportive Secondary Efficacy Endpoints and Analyses

Not applicable.

9.5 Additional Efficacy Analyses

[redacted]

[redacted]

9.6 Efficacy Subgroup Analyses

Efficacy analyses for the primary composite and coprimary endpoints will be performed by subgroups of sex, age group [redacted], BMI [redacted] race [redacted] ethnicity [redacted] region [redacted] and baseline C-APPS at maximum contraction [redacted]

10.0 Safety Analyses

10.1 General Considerations

Safety data will be summarized for the Safety Analysis Set. Safety summaries will be presented by treatment group. For the safety analysis, subjects are assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized.

The safety parameters will include incidence of AEs and change from baseline in vital signs. For each safety parameter evaluating change from baseline, the last nonmissing safety assessment before study drug administration will be used as the baseline for all analyses of that parameter.

10.2 Adverse Events

Adverse events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) according to the version of the MedDRA coding dictionary used for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

10.2.1 Treatment-Emergent Adverse Events

Treatment-emergent AEs (TEAEs) are defined as any AE with the onset that is after the first dose of study drug. Events where the onset date is the same as the study drug start date are assumed to be treatment-emergent. All TEAEs will be summarized overall, as well as by primary MedDRA SOC and PT. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

The number and percentage of subjects experiencing treatment-emergent AEs will be summarized.

10.2.2 Adverse Event Overview

An overview of AEs will be presented consisting of the number and percentage of subjects experiencing at least one event for each of the following AE categories:

- Any TEAE
- Any TEAE related to study treatment according to the investigator
  - Any TEAE related to study procedure according to the investigator
  - Any TEAE related to study drug according to the investigator
- Any mild TEAE
  - Any mild TEAE related to study treatment according to the investigator

- Any moderate TEAE
  - Any moderate TEAE related to study treatment according to the investigator
- Any severe TEAE
  - Any severe TEAE related to study treatment according to the investigator
- Any serious TEAE
  - Any serious TEAE related to study treatment according to the investigator
- Any TEAE leading to discontinuation
- Any treatment-emergent adverse event of special interest (AESI)
- Any possible distant spread of toxin (PDSOT) TEAE
- All deaths
- Any TEAE leading to death

10.2.3 Treatment-Emergent Adverse Events by SOC and/or PT

TEAEs will be summarized by SOC and PT; by maximum relationship to study drug as assessed by the investigator (e.g., reasonable possibility or no reasonable possibility) and SOC and PT; by maximum severity and SOC and PT; and by subject number and SOC and PT. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to study drug will be reported.

In addition, TEAEs will be summarized by PT and sorted by decreasing frequency for the total active group.

10.2.4 Treatment-Emergent Adverse Events per Patient-Years of Exposure

Not applicable.

10.2.5 SAEs (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation

Serious AEs (including deaths) and AEs leading to study drug discontinuation will be summarized by SOC and PT and in listing format.

10.2.6 Adverse Events of Special Interest and Possible Distant Spread of Toxin Adverse Events

Treatment-emergent adverse events of special interest (AESIs) will be summarized by PT. [redacted]

Possible distant spread of toxin (PDSOT) TEAEs will be summarized by PT. [redacted]

Tabular listings of AESIs and PDSOT TEAEs will be provided.

10.3 Analysis of Laboratory Data

Not applicable.

10.4 Analysis of Vital Signs

Vital sign measurements of systolic and diastolic blood pressure, pulse rate, and respiratory rate, and changes from baseline at each assessment timepoint will be summarized by treatment group. In addition, potentially clinically important vital sign values will be summarized for systolic and diastolic blood pressure, pulse rate and temperature.

Table 3. Criteria for Potentially Clinically Important Vital Sign Values

Parameter	Flag	Criteria[a]	
		Observed Value	Change From Baseline
Sitting systolic blood pressure, mm Hg	High	[redacted]	
	Low		
Sitting diastolic blood pressure, mm Hg	High		
	Low		
Heart rate, bpm	High		
	Low		
Temperature, degrees C	High		
	Low		

bpm = beats per minute; C = Celsius

a. Except for temperature, a postbaseline value is considered potentially clinically important if it meets both the observed-value and the change-from-baseline criteria; for temperature, only the observed-value criterion is needed to be potentially clinically important.

10.5 Safety Subgroup Analyses

Safety analyses for the summary of overall AEs will be performed by subgroups of sex, age group [redacted] race [redacted] ethnicity [redacted] and region [redacted]

10.6 Other Safety Analyses

Not applicable.

11.0 Other Analyses

Not applicable.

12.0 Interim Analyses

No interim analysis is planned for this study.

12.1 Data Monitoring Committee

No data monitoring committee (DMC) will be used for this study.

13.0 Overall Type-I Error Control

To control the family-wise Type I error rate at 0.05 for multiplicity across the primary and secondary analyses [redacted]

Thus, the hierarchical testing order follows for US FDA:

1. Primary composite: Achievement of Grade 1 or 2 (Minimal or Mild) and at least a 2-grade improvement from baseline based on both investigator's assessment using the C-APPS and subject's self-assessment using P-APPS at maximum contraction at Day 14
2. Secondary: Responses of *Satisfied* or *Very satisfied* on the ANLFQ: Satisfaction (Follow-up) Item 5 (effect of treatment) at Day 14
3. Secondary: Responses of *Not at all bothered* or *A little bothered* on the BAS-PP Item 2 (jawline) at Day 14
4. Secondary: Responses of *Not at all bothered* or *A little bothered* on the BAS-PP Item 1 (vertical neck bands) at Day 14

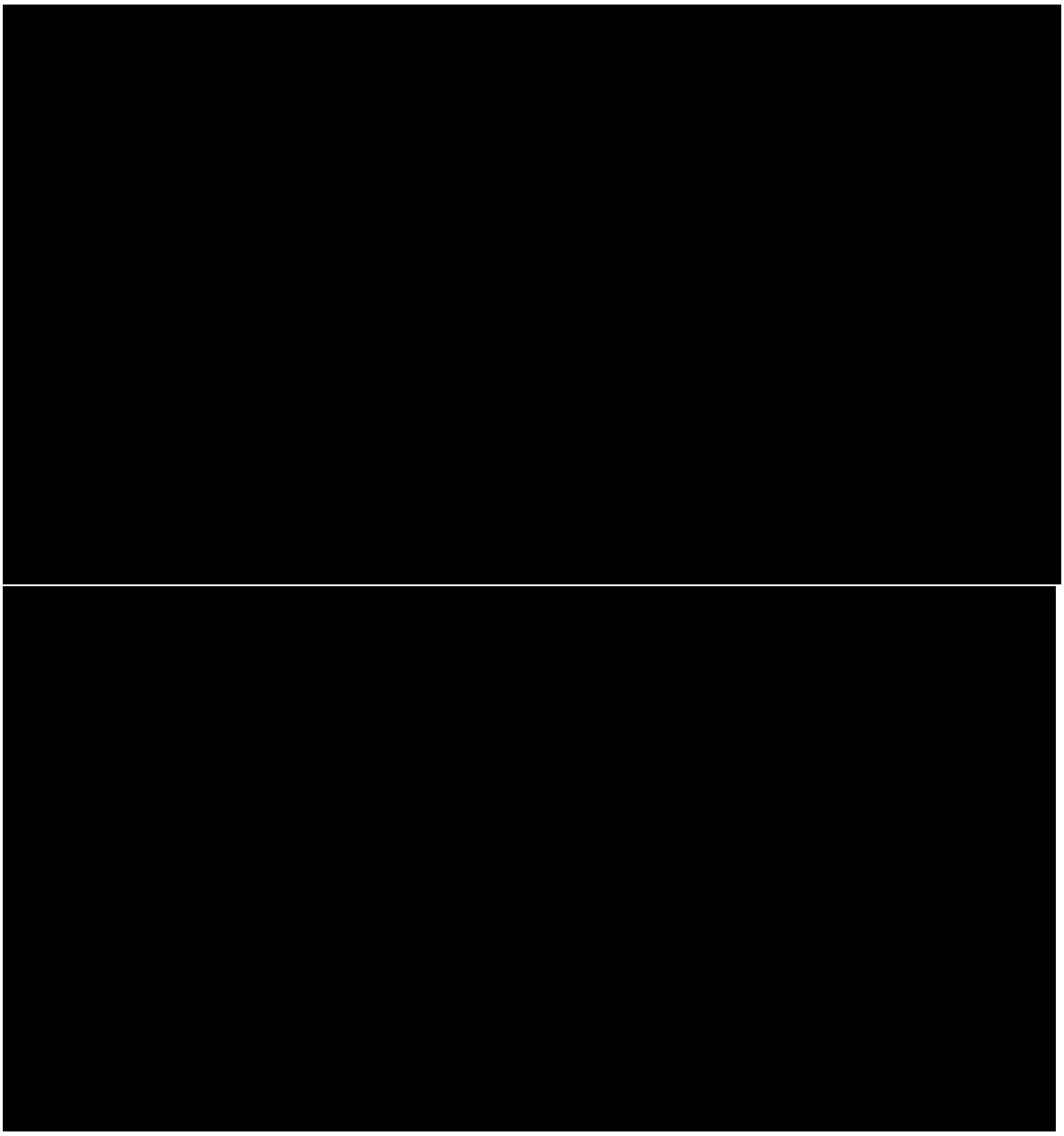
5. Secondary: Change from baseline on the ANLFQ: Impacts summary score at Day 14

The hierarchical testing order follows for EU regulatory agencies:

1. Coprimary: Achievement of at least a 2-grade improvement from baseline based on the following:
   - Investigator's assessment using C-APPS at maximum contraction at Day 14
   - Subject's self-assessment using P-APPS at maximum contraction at Day 14
2. Secondary: Achievement of a rating of Minimal or Mild according to subject's self-assessment using P-APPS at maximum contraction at Day 14
3. Secondary: Responses of *Satisfied* or *Very satisfied* on the ANLFQ: Satisfaction (Follow-up) Item 5 (effect of treatment) at Day 14
4. Secondary: Responses of *Not at all bothered* or *A little bothered* on the BAS-PP Item 2 (jawline) at Day 14
5. Secondary: Responses of *Not at all bothered* or *A little bothered* on the BAS-PP Item 1 (vertical neck bands) at Day 14
6. Secondary: Change from baseline on the ANLFQ: Impacts summary score at Day 14
7. Secondary: Change from baseline on the ANLFQ: Impacts summary score at Days 30, 60, and 90

14.0 Version History

Table 4. SAP Version History Summary



15.0 References

Appendix A. Protocol Deviations

The number and percentage of subjects who reported at least one of the following protocol deviation categories will be provided.

- Subject entered into the study even though did not satisfy entry criteria.
- Subject developed withdrawal criteria during the study and was not withdrawn.
- Subject received wrong treatment or incorrect dose of study treatment.
- Subject received prohibited concomitant medication or procedure.