Statistical Analysis Plan for Study M22-974

Subject Satisfaction and Natural Outcomes Following BOTOX® Cosmetic Treatment in Subjects with Upper Facial Lines

Date: 17 March 2025

Version 2.0

Table of Contents

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for Study M22-974: A Prospective, Multi-center, Open-label Study to Evaluate Subject Satisfaction and Natural Outcomes Following Administration of BOTOX® Cosmetic (Botulinum Toxin Type A) Purified Neurotoxin Complex in Subjects with Upper Facial Lines (Glabellar Lines, Lateral Canthal Lines, and Forehead Lines).

Study M22-974 examines subject satisfaction and natural outcomes following simultaneous treatment of multiple upper facial lines (GL, LCL, and FHL) with 64 U of BOTOX Cosmetic.

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.

2.0 Study Design and Objectives

2.1 Objectives and Hypotheses

The primary objective of this study is to evaluate subject satisfaction and natural outcomes following simultaneous treatment of multiple upper facial lines (GL, LCL, and FHL) with 64 U of BOTOX Cosmetic by Facial Line Satisfaction Questionnaire (FLSQ) Item 4 at Day 30 based on the modified Intent-to-Treat (mITT) population, which consists of all enrolled subjects who receive the total dose of 64U of BOTOX cosmetic.

2.2 Study Design Overview

This is a 6-month, prospective, multicenter, open-label study to evaluate subject satisfaction and natural outcomes following the administration of BOTOX Cosmetic 64 U for the treatment of upper facial lines.

The study will enroll approximately 100 subjects with moderate to severe GL, LCL, and FHL as assessed by the investigator. Out of the approximately 100 enrolled subjects:

- approximately 20% of the study population will be treatment naïve to any botulinum toxin of any serotype for any indication,
- approximately 20% of the study population will be self-identified as male,
- approximately 20% of the study population will be self-identified as Asian, and
- approximately 20% of the study population will have Fitzpatrick skin phototypes of IV, V, or VI.

On Day 1, after subjects are verified to meet all eligibility criteria and have completed all baseline study procedures, they will receive 64 U of BOTOX Cosmetic. Each subject will receive a total of 16 injections (0.1 mL or 4 U of BOTOX Cosmetic per injection), with 5 injections in the glabellar complex, 3 injections in the lateral canthal area on each side (6 total injections), and 5 injections in the frontalis muscle.

There are seven in-clinic visits: screening/baseline/treatment (Day 1), post-treatment follow-up visits (Days 30, 60, 90, 120, 150), and study exit (Day 180 or Early Exit). The study schema of the study is shown in Figure 1.

Figure 1. Study Schema

2.3 Treatment Assignment and Blinding

This is an open-label study. Hence, there is no subject randomization or blinding for the study treatment assignment.

2.4 Sample Size Determination

No formal sample size calculation was performed.

For a conservative expected proportion of 75% for the primary endpoint (proportion of subjects reporting *Mostly satisfied* or *Very satisfied* on the FLSQ Follow-up Item 4 at Day 30), a sample size of 100 subjects in the study is required for an adjusted logit-based 95% binomial confidence interval to have a width of 16.8% (interval is 75% - 9.1% to 75% + 7.7% with recommended correction of - 1/2).[1]

3.0 Endpoints

3.1 Primary Endpoint(s)

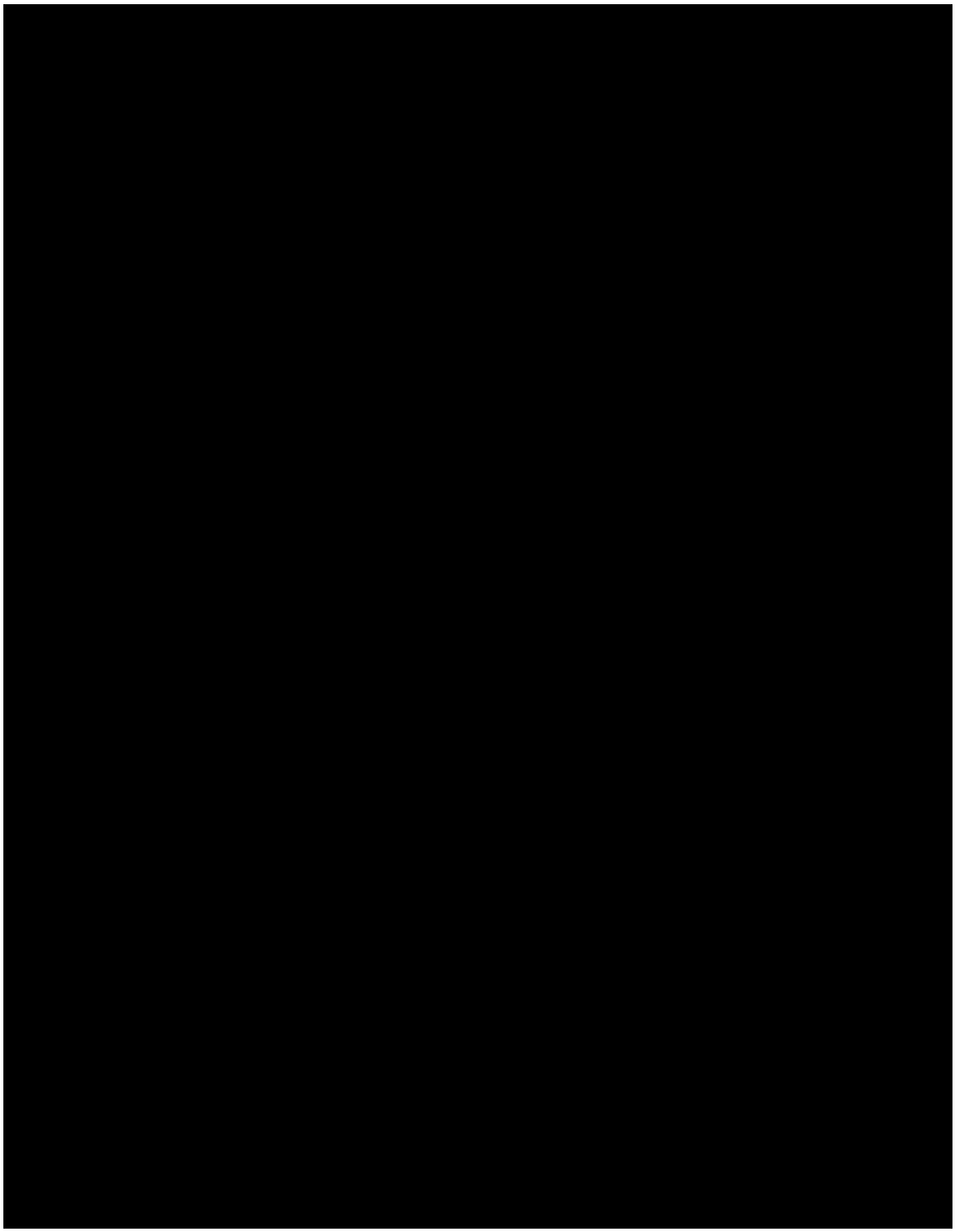
The primary endpoint is the responder status of *Mostly satisfied or Very satisfied* on the Facial Line Satisfaction Questionnaire (FLSQ) Follow-up Item 4 at Day 30.

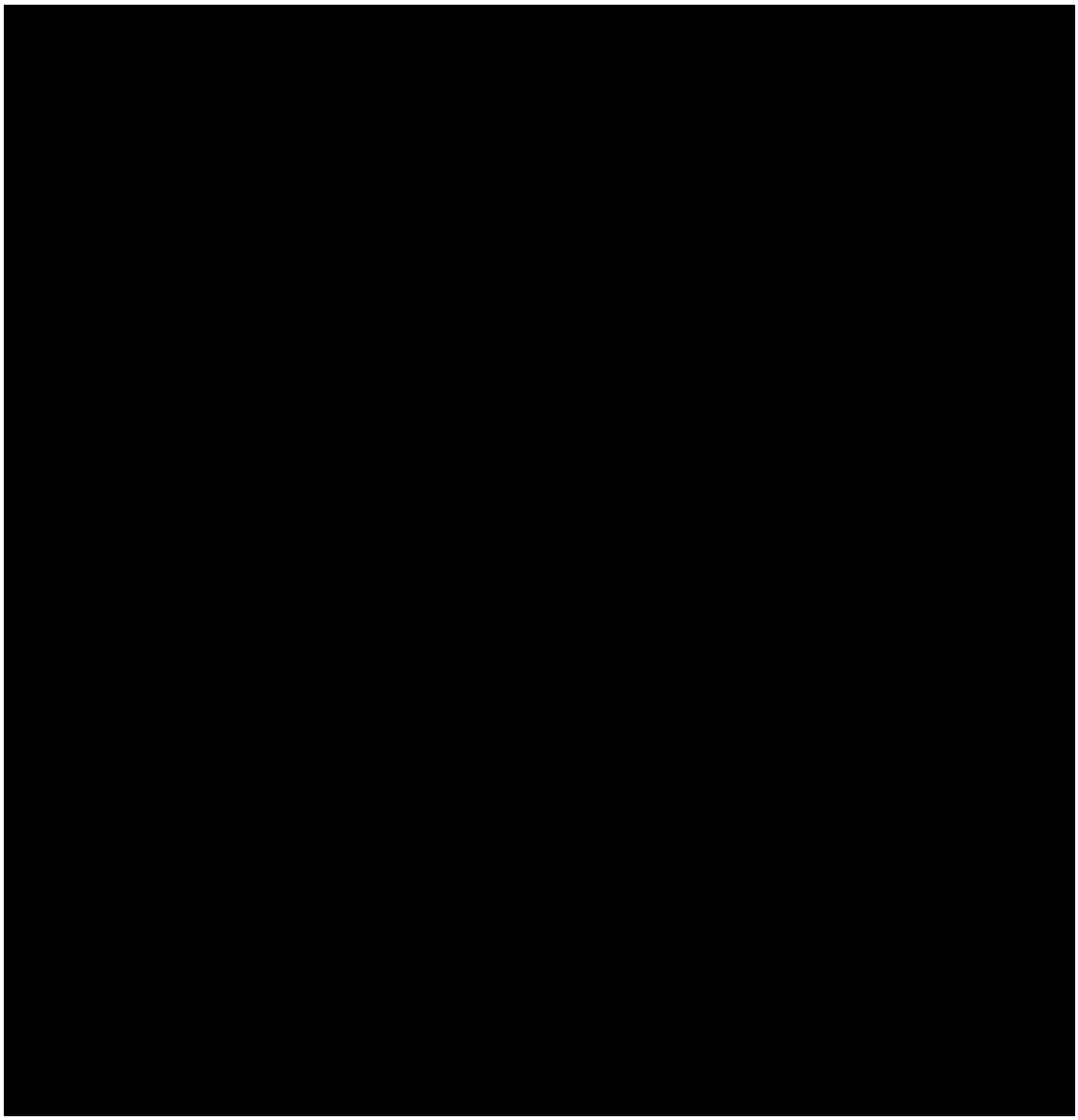
3.2 Secondary Endpoint(s)

The secondary endpoint is

- Change from baseline in subject's assessment based on Rasch-transformed score of FACE-Q Psychological Function scale at Day 30

3.3 Additional Endpoint(s)





3.4 Safety Endpoint(s)

- The safety endpoint is the incidence of adverse events (AEs).

4.0 Analysis Populations

The following population sets will be used for the analyses.

The Modified Intent to Treat (mITT) Population includes all enrolled subjects who receive the total dose of 64 U of BOTOX Cosmetic, as described in the study protocol. The mITT Population will be used for all efficacy analyses. The Safety Analysis Set (also referred to as the Safety Population) consists of all subjects who receive at least 1 injection of study drug. The safety analyses will be based on the Safety Population.

5.0 Subject Disposition

The total number of subjects who were screened, enrolled and treated 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:

- Number of subjects screened
- Number of subjects treated
- Number of subjects completed the study
- Number of subjects discontinued from the study
    - Reasons for discontinuation from the study

The number of subjects in each analysis population will also be summarized.

6.0 Study Drug Duration

For the Safety Population, study duration will be summarized with descriptive statistics (mean and standard deviation, median, Q1, Q3, minimum, and maximum). The number of subjects followed for specific periods of time (1-30 days, 31-60 days, 61-90 days, and 91-120 days, 121-150 days,151 days to 180 days, >180 days) will also be summarized. Study duration will be calculated as date of end of study participation minus date of study drug administration +1.

7.0 Subject Characteristics

Demographics, baseline characteristics, medical history, prior and concomitant medications, and prior and concurrent procedure will be summarized for the mITT population. 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

Demographic and baseline characteristics will be summarized descriptively for the mITT population. Unless otherwise specified, baseline is defined as the last non-missing value prior to the first administration of study treatment.

Continuous demographic variables include age, weight, height and BMI.

Categorical demographic variables include:

- Sex reported at birth
- Sex
- Gender identity
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Race
- Fitzpatrick skin type (I, II, III, IV, V, VI)
- Prior Botulinum Toxin Treatment

Baseline characteristics include investigator-rated and subject-rated GL severity at maximum contraction and at rest using AGLSS, investigator-rated and subject-rated LCL at maximum contraction and at rest using LCLSS, investigator-rated and subject-rated FHL at maximum contraction and at rest using FHLSS, FLSQ baseline impact domain score and FACE-Q baseline Psychological Function score.

7.2 Medical History and Prior and Concomitant Medication

Medical history data will be coded using the current Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term (PT)) will be summarized. The system organ class (SOC) will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (SOC or PT).

A prior medication is defined as any medication taken prior to the date of the first study drug administration. A concomitant medication is defined as any medication that started prior to the date of the first study drug administration and continued to be taken after the first study drug administration or any medication that started on or after the date of the first study drug administration. Prior and concomitant medications will be summarized separately for the mITT Population. The number and percentage of subjects taking prior and concomitant medications will be summarized by generic drug name, based on the World Health Organization (WHO) Drug Dictionary. The actual version of the WHO Drug Dictionary will be noted in the statistical tables.

7.3 Prior and Concurrent Procedures

All procedures undergone prior to the date of the study drug administration will be considered as prior procedures. All procedures undergone on or after date of the study drug administration through the exit visit will be considered concurrent procedures. Prior and concurrent procedures will each be summarized by MedDRA High Level Term (HLT) and PT for the mITT Population. The MedDRA HLT will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each HLT.

7.4 Protocol Deviations

Protocol deviations are defined in accordance with ICH guidelines and include but are not limited to: eligibility criteria violations, receipt of wrong treatment or incorrect dose of

study treatment, development of withdrawal criteria without being withdrawn, and the use of excluded concomitant medications. A listing of subjects with protocol deviations will be provided.

For all protocol deviation categories and across all categories, the number and percentage of randomized subjects with at least one protocol deviation will be summarized.

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

Observed data will be used without imputation, so subjects who have an intercurrent event resulting in not having FLSQ Item 4 at Day 30 for any reason will not be included in the primary efficacy endpoint. Intercurrent events include:

- Subjects who are enrolled but do not receive any study drug;
- Subjects who are enrolled and treated but prematurely discontinue the study before assessment of the primary endpoint;
- Subjects who die before assessment of primary endpoint;
- Subjects who are lost to follow-up and are missing assessment of primary endpoint;
- Subjects who are missing assessment of primary endpoint for any other reason (including COVID) not mentioned above
    - The same approach will be used for the secondary analysis.

9.0 Efficacy Analyses

9.1 General Considerations

The following considerations apply to all efficacy analyses, unless otherwise specified:

- All efficacy analyses will be conducted in the mITT population unless otherwise specified.
- Day 1 is defined as the day on which the study treatment is first received.

- Baseline for effectiveness parameters will be the last non-missing assessment prior to or on Day 1.
- Study day is always relative to Day 1.
- Change from baseline will be computed as postbaseline minus baseline.
- Continuous variables will be summarized descriptively using number of subjects with observed values (n), mean, standard deviation (SD), median, minimum (min), and maximum (max). In addition, 95% confidence intervals (CIs) will be performed using paired t-tests for change from baseline endpoints.
- Categorical variables will be summarized descriptively using number of subjects with observed values or events (n) and the percentage of subjects with observed values or events.
- Analyses based on responder rates will include a 95% CI for the responder rate using the exact binominal distribution (Clopper-Pearson approach).
- All CIs will be 2-sided 95% CIs, unless stated otherwise.

9.2 Handling of Missing Data

No imputation of missing data will be made for the efficacy analyses.

9.3 Primary Efficacy Endpoint(s) and Analyses

9.3.1 Primary Efficacy Endpoint(s)

The primary efficacy endpoint is the responder status of *Mostly satisfied or Very satisfied* on the Facial Line Satisfaction Questionnaire (FLSQ) Follow-up Item 4 at Day 30.

9.3.2 Main Analysis of Primary Efficacy Endpoint(s)

Analysis of the primary endpoint will be conducted based on mITT population. No formal statistical testing will be performed, and descriptive statistics will be tabulated. Observed data will be used without imputation.

9.3.3 Sensitivity and Supplementary Analyses of the Primary Efficacy Endpoint(s)

Not applicable.

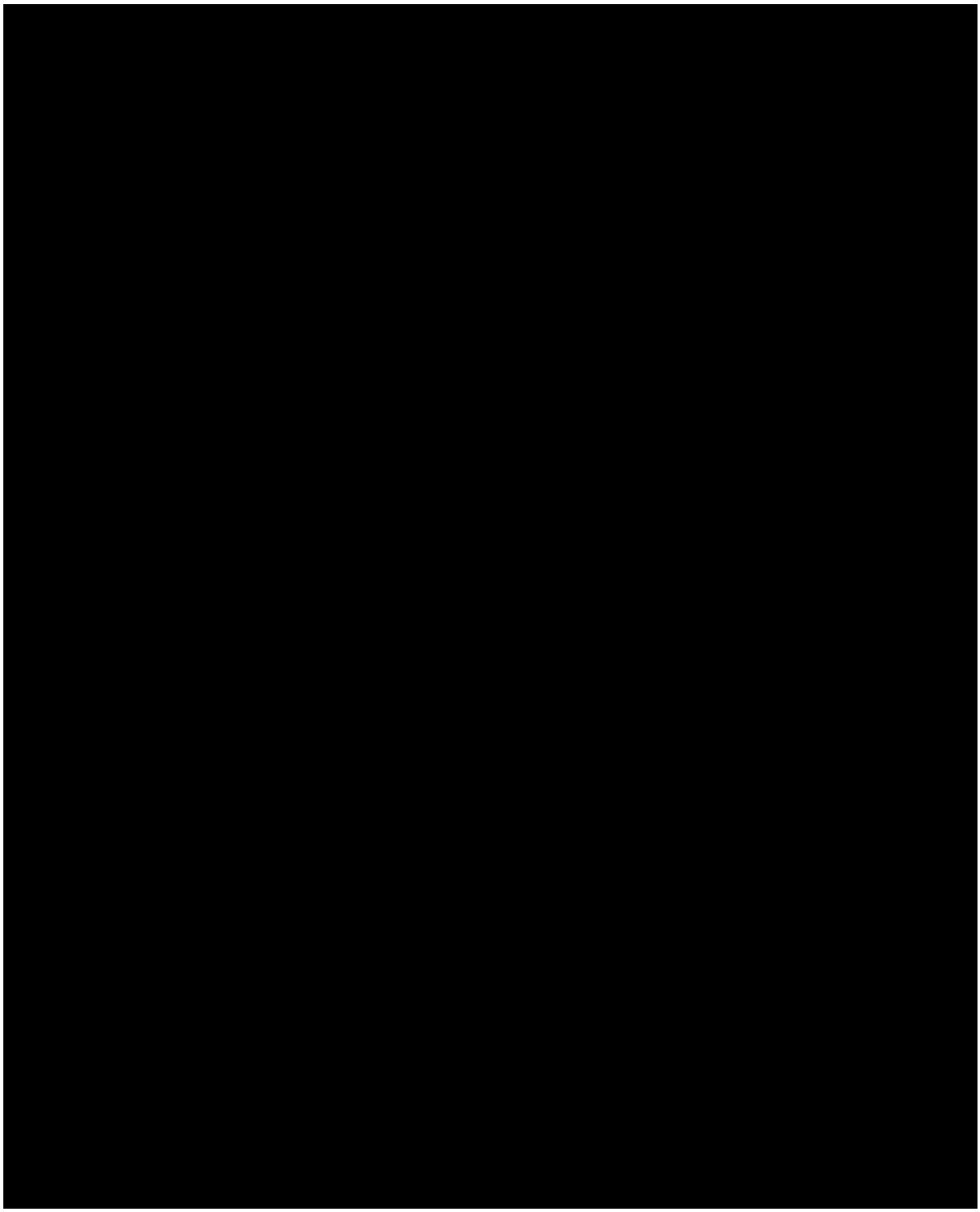
9.4 Secondary Efficacy Endpoints and Analyses

The secondary endpoint is:

- Change from baseline in subject's assessment based on Rasch-transformed score of FACE-Q Psychological Function scale at Day 30.

This endpoint will be analyzed using descriptive statistics.

9.5 Additional Efficacy Analyses



[redacted]

9.6 Efficacy Subgroup Analyses

For the primary and secondary endpoints, analyses will be conducted using the following subgroups with descriptive summary statistics provided:

- Toxin Naïve (Yes or No)
- Sex (Female, Male)
- Race (Asian, Non-Asian)
- Fitzpatrick skin type (Type I-III, Type IV-VI)

10.0 Safety Analyses

10.1 General Considerations

Safety analyses will be performed on the Safety Population. For safety analyses, "baseline" refers to the last non-missing observation before the first administration of study drug unless otherwise noted.

Adverse events (AEs) will be coded using the MedDRA. The actual version of the MedDRA coding dictionary will be noted in the statistical outputs and clinical study report.

10.2 Adverse Events

AEs will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) according to the current 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 adverse events (TEAEs) are defined as any AE with an onset after the first dose of study treatment. Events where the onset date is the same as the study treatment start date are assumed to be treatment-emergent, unless the study treatment start time and the AE start time are collected and the AE start time is prior to the study treatment start time. If an incomplete onset date is collected for an AE, then the AE will be assumed to be treatment-emergent unless there is evidence that confirms that the AE was not treatment-emergent (e.g., the known portion of the AE onset and/or the AE end date was prior to the date of the first dose of study treatment).

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 TEAEs will be summarized.

An AE will be considered a TESAE if it is a TEAE that additionally meets any serious adverse event (SAE) criterion.

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 drug according to the investigator
  - Any TEAE related to study procedure 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 death
- Any PDSOT TEAEs
- Any TEAE of special interest
- All deaths

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

Treatment-emergent adverse events will be summarized by SOC and PT; by maximum severity and SOC and PT; and by subject number and SOC and PT.

Furthermore, TEAEs will be summarized by PT and sorted by decreasing frequency.

Treatment-related TEAEs, study drug related TEAEs, and study procedure related TEAEs will also be summarized by SOC and PT.

In addition, for both possible distant spread of toxin (PDSOT) TEAEs and treatment-related TEAEs, onset after study drug administration and duration will be evaluated. TEAE onset will be reported based on number of days subsequent to the study drug administration and will be summarized with descriptive statistics. TEAE duration will be defined as (AE end date minus AE start date) + 1 day) and will be summarized with descriptive statistics. For ongoing AEs, duration will be estimated based on duration at study exit, unless otherwise specified.

10.2.4 SAEs (Including Deaths)

Serious TEAEs (including deaths) will be summarized by SOC and PT and all SAEs will be provided in listing format.

10.2.5 Potential Distant Spread of Toxin Adverse Events

To assess PDSOT, MedDRA preferred terms that may be associated with botulinum toxin effects have been identified (Appendix B). All TEAEs associated with PDSOT will be tabulated by PT; in addition, all PDSOT AEs will be listed by subject.

10.3 Analysis of Laboratory Data

Pregnancy test will be performed at Day 1 and may also be performed at any other visit, at the investigators' discretion.

10.4 Immunogenicity

Not applicable.

10.5 Analysis of Vital Signs

Not applicable.

10.6 Safety Subgroup Analyses

Not applicable.

10.7 Other Safety Analyses

Not applicable.

11.0 Other Analyses

11.1 Empirical Cumulative Distribution Function by Treatment Group for Change in Clinical Outcome Assessment (COA) Scores

To support the interpretation of meaningful treatment effect as measured by the change in FACE-Q Psychological Function Scale, empirical cumulative distribution functions (eCDFs) of the change in FACE-Q Psychological Function Scale from baseline to Day 30 will be plotted by treatment group based on observed data, with change from baseline in FACE-Q Psychological Function Scale on the horizontal axis and the cumulative percentage of subjects experiencing up to that specific level of change on the vertical axis. eCDFs provide a visual display of the cumulative percentage of subjects within each group that achieve within-patient change across the entire distribution of change, thereby illustrating the percentage of responders across the range of potentially meaningful change threshold values.

12.0 Interim Analyses

No interim analysis is planned for this study.

13.0 Overall Type-I Error Control

No statistical test will be performed. All CIs will be 2-sided 95% CIs, unless stated otherwise.

14.0 Version History

SAP Version	Date	Summary
1.0	14 March 2024	Initial version

2.0	17 March 2025	Section 3.3
		• Added 3 additional endpoints:
		  ○ Responder status of Mostly satisfied or Very satisfied on the Facial Line Satisfaction Questionnaire (FLSQ) Follow-up Item 4 at all follow-up visits
		  ○ Change from baseline in subject's assessment based on Rasch-transformed score of FACE-Q Psychological Function scale at all follow-up visits
		  ○ Achievement of ≥ 20 points improvement from baseline on FLSQ impact domain score at all follow-up visits
		Section 6.0
		• Updated study drug duration analysis grouping
		Section 7.1
		• Added 'prior Botulinum Toxin Treatment' in demographic variable list
		Section 9.1
		• Stated ‘95% confidence intervals (CIs) will be performed using paired t-tests for change from baseline endpoints’ and ‘ Analyses based on responder rates will include a 95% CI for the responder rate using the exact binominal distribution (Clopper-Pearson approach).’
		Section 9.5
		• removed 'Listings of all endpoint values will also be provided.' Only listings for primary and secondary endpoints will be provided.
		• updated 'Onset of efficacy as measured by subject self-assessment in a patient electronic diary from Day 1 up to Day 14' to 'Onset of efficacy as measured by subject self-assessment in a patient electronic diary from Day 2 to Day 15' per protocol V3.0
		• added Kaplan-Meier language for endpoint 'Onset of efficacy as measured by subject self-assessment in a patient electronic diary from Day 2 to Day 15.'
		• Added 3 additional endpoints:

		○ Responder status of Mostly satisfied or Very satisfied on the Facial Line Satisfaction Questionnaire (FLSQ) Follow-up Item 4 at all follow-up visits ○ Change from baseline in subject's assessment based on Rasch-transformed score of FACE-Q Psychological Function scale at all follow-up visits ○ Achievement of ≥ 20 points improvement from baseline on FLSQ impact domain score at all follow-up visits Section 9.6 • added subgroup analyses for primary and secondary endpoint by toxin naïve, sex, race and Fitzpatrick skin type

15.0 References

1. Edwardes M. The evaluation of confidence sets with application to binomial intervals. Stat Sin. 1998;8(2):393-409.

Appendix A. List of SAP Signatories

Name	Title	Role/Functional Area
[redacted]	[redacted]	Author
		Biostatistics
		Statistical Programming
		Medical/Scientific Monitor

Appendix B. Definition of Possible Distant Spread of Toxin

PDSOT AEs will be identified using the following search criteria for the MedDRA version being used for the study:

Area of Safety Interest	Search Criteria
PDSOT	All PTs in the CMQ (company MedDRA query): Botulinum Toxin Adverse Events Of Interest Possible Distant Spread Of Toxin (PDSOT) (BOTOX Product Specific); identified using CMQ = 12200048