Statistical Analysis Plan for Study M21-606

A Phase 2, Multicenter, Randomized, Placebo-controlled Study to Evaluate the Safety and Efficacy of OnabotulinumtoxinA X for Forehead Lines

Date: 12 September 2022

Version 2.0

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 OnabotA X Study M21-606 Forehead Lines: OnabotulinumtoxinA X in the Treatment of Moderate to Severe Forehead Lines (FHL).

Study M21-606 examines the safety and efficacy of OnabotA X in subjects with moderate to severe forehead lines.

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 objective of this study is to evaluate the safety and efficacy of 3 doses of OnabotA X for the treatment of moderate to severe FHL.

The null and alternate hypotheses are the following:

H_0: $P_T = P_P$ (The treatment response probability between each OnabotA X treatment group and the placebo group is equal)

H_A: $P_T \neq P_P$ (The treatment response probability between each OnabotA X treatment group and the placebo group is not equal)

Estimand attributes of the primary efficacy endpoint are detailed in Table 1 (see Section 9.3.2). In addition, the attribute of treatment is a single dose of active treatment or placebo.

Differences in proportion of subjects who would achieve ≥ 1-grade improvement from baseline on the investigator-rated FWS-FHL at maximum contraction at Day 30 had they

not missed FWS-FHL assessments, for each active treatment group ([redacted]) in comparison with placebo in the ITT population.

Any missing assessments will be assumed to be missing at random and imputed using multiple imputation. Statistical significance of the difference will be tested using Cochran-Mantel-Haenszel (CMH) test stratified by baseline investigator-rated FWS-FHL at maximum contraction.

The clinical hypothesis is that OnabotA X has an acceptable safety profile when administered to the frontalis muscles and glabellar complex in subjects with moderate to severe FHL and moderate to severe Glabellar Line (GL).

No estimand is defined for the safety evaluations.

2.2 Study Design Overview

The schematic of the study is shown in Figure 1.

[redacted]

2.3 Treatment Assignment and Blinding

Subjects will be randomized at Day 1 to 4 treatment groups in a 1:1:1:1 ratio to receive OnabotA X [redacted] or [redacted] (placebo) in the frontalis. All randomized

subjects will receive a fixed dose of [redacted] OnabotA X in the glabellar complex, followed by injections in the frontalis. Randomization will be stratified by the investigator-assessed baseline severity of FHL (FWS-FHL) at maximum contraction and investigator site.

2.4 Sample Size Determination

Approximately 120 subjects will be randomized into the study in a 1:1:1:1 ratio (30 in each group) yielding approximately 90 subjects receiving OnabotA X in the frontalis muscle for treatment of FHL. The sample size of 120 is chosen empirically to allow for an adequate safety data of subjects treated with OnabotA X for this indication.

The primary efficacy endpoint is the achievement of ≥ 1-grade improvement from baseline on the investigator-rated FWS-FHL at maximum contraction at Day 30. [redacted] Assuming [redacted] the sample size of N = 30 per group will provide 97% power.

All calculations are based on a 2-sided type I error rate of 0.05. [redacted].

3.0 Endpoints

3.1 Primary Endpoint(s)

The primary efficacy endpoint is the achievement of ≥ 1-grade improvement from baseline on the investigator-rated Facial Wrinkle Scale (FWS-FHL) at maximum contraction (also known as eyebrow elevation or surprise) at Day 30.

The FWS-FHL is scored on a 0-3 range (0 = None, 1 = Mild, 2 = Moderate, 3 = Severe).

3.2 Secondary Endpoint(s)

No secondary endpoints are identified in this study.

3.3 Other Efficacy Endpoint(s)

[redacted]

3.4 Safety Endpoint(s)

The safety endpoints/measures for this study include:

- Adverse events
- Vital signs
- Neurologic assessment

4.0 Analysis Populations

The following population sets will be used for the analyses.

The Intent to Treat (ITT) Population includes all randomized subjects. The ITT Population will be used for all efficacy analyses.

The Safety population consists of all subjects who received at least 1 dose 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, randomized, 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 for overall and each treatment group:

- Subjects enrolled (randomized) in the study;
- Subjects who were treated;
- Subjects who completed the study;
- Subjects who prematurely discontinued the study;
  - Reasons for discontinuation will be summarized

Subjects in each analysis population will also be summarized.

6.0 Study Duration and Treatment Exposure

For the Safety population, study duration will be summarized for each treatment group and overall. Study duration will be summarized with descriptive statistics (mean and standard deviation, median, Q1, Q3, minimum, and maximum). The number of subjects exposed for specific period of time ([redacted]) will also be summarized.

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

Demographics, baseline characteristics, medical history, prior and concomitant medications, and concurrent procedure will be summarized for the ITT population 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, Q1, Q3, minimum, and maximum).

7.1 Demographics and Baseline Characteristics

Continuous demographic variables include age. Categorical demographic variables include sex, ethnicity, race, age group [redacted]).

Baseline characteristics include weight, height, body mass index (BMI), nicotine use, alcohol use, Fitzpatrick skin type, investigator-rated and subject-rated FWS-FHL and FWS-GL severity at maximum contraction and at rest, [redacted].

7.2 Medical History

Medical history 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. 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 preferred term).

7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name. A prior medication is defined as any medication taken prior to the date of the study drug administration. A concomitant medication is defined as any medication that started prior to the study drug administration and continued to be taken after the study drug administration or any medication that started after the study drug administration. 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.

7.4 Concurrent Procedure

All procedures undergone on or after Study Day 1 visit through the exit visit will be considered concurrent procedures. Concurrent procedure will be summarized by MedDRA high level term and preferred term. The MedDRA high level term will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each high level term. A listing will be generated for concurrent procedures.

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

The primary endpoint will be analyzed based on the ITT population, 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.
- Subjects who are randomized but prematurely discontinued the study before assessment of the primary endpoint will be considered as part of the ITT population.
- 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 data for any other reason for the primary endpoint will count as though they hypothetically continued in the study.

9.0 Efficacy Analyses

9.1 General Considerations

All efficacy analyses will be conducted in the ITT Population. For the efficacy analyses, subjects are summarized based on the randomized treatment assigned.

The number and percentage (or proportion) of the subjects who are responders will be summarized. Continuous variables will be summarized by number of subjects with

observed values (n), mean, SD, median, 1st and 3rd quartiles (Q1, Q3), minimum, and maximum. All efficacy results will be summarized by treatment group.

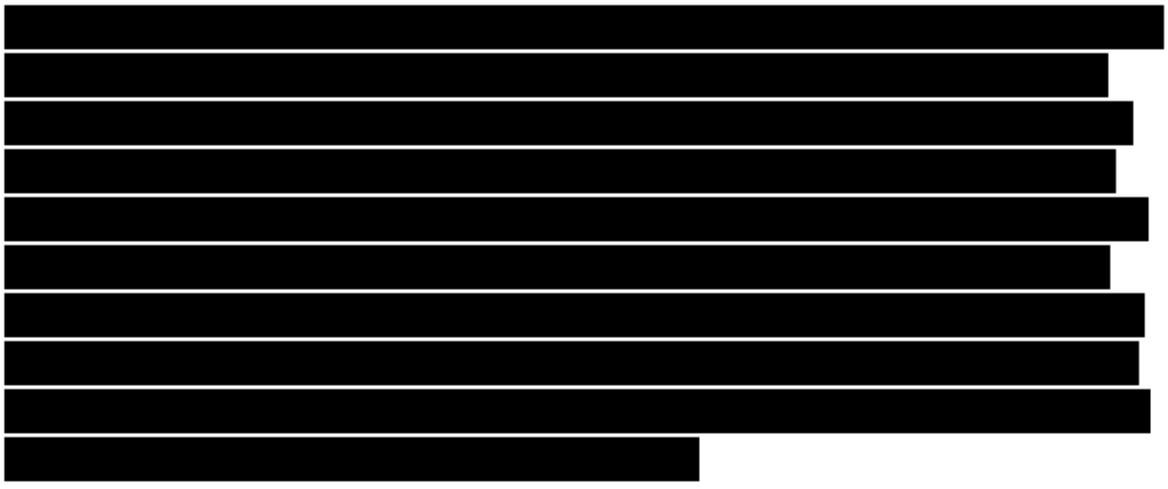
Unless otherwise specified, any subject who is randomized based on a wrong stratum will be analyzed according to the actual stratum the subject belongs to.

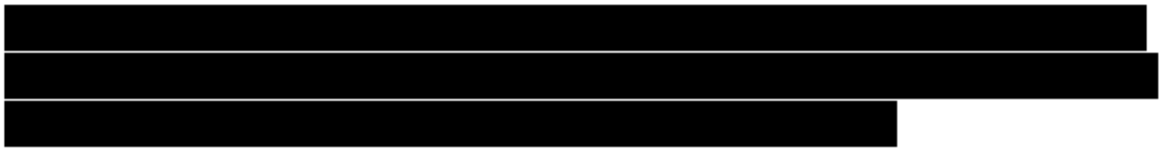
"Baseline" refers to the last non-missing observation before the first administration of study drug or randomization if no study drug is given.

9.2 Handling of Missing Data

Missing data will be imputed using multiple imputation (MI) method for the primary and secondary endpoints.

SAS Proc MI procedure will be used to generate 5 imputation data sets. Seed for all Proc MI procedure is pre-specified as 721606.





After imputation, the changes from baseline values will be calculated. The responder status based on raw values or change from baseline values will then be derived for each post baseline visit.

Each of the 5 imputation data sets will be analyzed individually, then combined to generate the final inferences as described in Section 9.3.2.

9.3 Primary Efficacy Endpoint(s) and Analyses

9.3.1 Primary Efficacy Endpoint(s)

The primary efficacy endpoint is achievement of ≥ 1-grade improvement from baseline on the investigator-rated FWS-FHL at maximum contraction at Day 30.

9.3.2 Main Analysis of Primary Efficacy Endpoint(s)

The attributes of the estimands corresponding to the primary efficacy endpoint are summarized in Table 1.

Table 1. Summary of the Estimand Attributes of the Primary Efficacy Endpoint(s)

	Attributes of the Estimand				
Estimand Label	**Treatment**	**Endpoint**	**Population**	**Handling of Intercurrent Events**	**Statistical Summary**
Hypothetical estimand for primary endpoint	OnabotA X or placebo	Achievement of ≥ 1-grade improvement from baseline on the investigator-rated FWS-FHL at maximum contraction at Day 30	ITT (All randomized)	Subjects who discontinue study prior to Day 30 assessments, or who do not have FWS-FHL assessments will be included in the analysis as a hypothetical scenario in which they had not missed the FWS-FHL assessments	Response rates and rate differences between each active treatment group and placebo, after MI using CMH test stratified by baseline investigator-rated FWS-FHL at maximum contraction

FWS-FHL = Facial Winkle Scale - Forehead Line; CMH = Cochran-Mantel-Haenszel; ITT = intent-to-treat; MI = multiple imputation; OnabotA X = onabotulinumtoxinA X

The evaluation of the equality of the proportions of responders will be based on Cochran-Mantel-Haenszel (CMH) test stratified by baseline investigator-rated FWS-FHL at maximum contraction. Wald confidence intervals for proportions of responders and difference in the proportion of responders will be presented. The Breslow-Day homogeneity of the odds-ratio test will be performed to test the treatment-by-investigator-rated baseline FWS-FHL severity at maximum contraction interaction.

To obtain pooled CMH p-value, the Wilson-Hilferty transformation will be used.[1,2]

[redacted]

[redacted]

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

A sensitivity analysis will be performed for the primary efficacy variable using observed data.

9.4 Secondary Efficacy Endpoints and Analyses

No secondary endpoints are identified in this study.

[redacted]

9.6 Efficacy Subgroup Analyses

Not applicable.

10.0 Safety Analyses

10.1 General Considerations

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

For safety analyses "baseline" refers to the last non-missing observation before the first administration of study drug unless otherwise noted.

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 are defined as any AE with the onset that is after the study drug administration. Events where the onset date is the same as the study drug start date are assumed to be treatment-emergent, unless known to have started prior to study drug administration. All treatment-emergent AEs will be summarized overall (i.e., any preferred term), as well as by primary MedDRA SOC and Preferred Term. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

10.2.2 dverse 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 treatment-emergent AE
- Any treatment-emergent AE related to study treatment according to the investigator
    - Any treatment-emergent AE related to study procedure according to the investigator
    - Any treatment-emergent AE related to study drug according to the investigator

- Any severe treatment-emergent AE
  - Any severe treatment-emergent AE related to study treatment according to the investigator
- Any serious treatment-emergent AE
  - Any serious treatment-emergent AE related to study treatment according to the investigator
- Any treatment-emergent AE leading to death
- Any PDSOT TEAEs
- All deaths

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

Treatment-emergent AEs will be summarized by SOC and PT; by maximum severity by SOC and PT; and by subject number and SOC and PT. Specific treatment-emergent AEs will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same AE occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

In addition, treatment-emergent adverse events will be summarized by PT and sorted by decreasing frequency for the overall group. Study drug related treatment-emergent adverse events and study procedure related treatment-emergent adverse events will also be summarized by SOC and PT.

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

SAEs (including deaths) will be summarized by SOC and PT and 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 treatment-emergent AEs associated with PDSOT will be tabulated by PT; in addition, all PDSOT AEs will be listed by subject.

10.3 Analysis of Vital Signs

Vital sign measurements of systolic and diastolic blood pressure, pulse rate, respiratory rate, and body temperature will be summarized by treatment group.

[redacted]

Vital sign variables will be evaluated based on PCI criteria ([redacted]). For each vital sign PCI criterion, the number and percentage of subjects who have a vital sign value meeting the criteria will be summarized. [redacted]

Listings will be provided to summarize subject-level vital sign data for subjects meeting PCI criteria.

10.4 Safety Subgroup Analyses

Not applicable.

10.5 Other Safety Analyses

Neurologic Assessment: The neurologic assessment, which comprises a Focused Symptoms Questionnaire and Focused Neurologic Examination, will be conducted throughout the study. A listing of neurologic assessments will be provided.

11.0 Other Analyses

Not applicable.

12.0 Interim Analyses

An interim analysis is planned to occur when the majority of subjects have completed the [redacted] visit or exited from the study. [redacted]

The primary and selected other efficacy analyses will be performed for each interim analysis, as well as summaries of all safety variables. Further details are provided in DMC Charter.

12.1 Data Monitoring Committee

An independent DMC will be instituted to review interim safety and efficacy data to provide a dose recommendation for future clinical studies. [redacted]

The DMC will review the results of the interim analyses and make a recommendation in accordance with the DMC Charter.

A separate DMC charter prepared outside of the protocol describes the roles and responsibilities of the DMC members, frequency and scope of the data reviews, and expectations for blinded communications.

13.0 Overall Type-I Error Control

Analyses will be conducted using a Type I error rate of α=0.05 for each OnabotA X group. The p-values will be presented as unadjusted p-values that are deemed statistically significant if $p \leq 0.05$.

14.0 Version History

Table 2. SAP Version History Summary

Version	Date of Approval	Summary
1.0	04 March 2022	Initial version
2.0		Section 3.3: [redacted] Section 4.0, changed the safety analysis set to be analyzed per randomization due to the difficulty of determine the actual treatment Section 9.2: removed race from the multiple imputation model, since the majority of subjects are White, and including the race would cause convergence problem Section 12.0 and 12.1 Removed the second possible interim snapshot, updated the interim to occur after the majority of subjects have completed the [redacted] visits instead of all subjects completed [redacted] visits Table B-1 MedDRA preferred terms evaluated for PDSOT, removed the term "Extraocular muscles paresis", added the term "Opthalmoplegia"

15.0 References

1. EB W, MM H. The distribution of chi-square. Proc Natl Acad Sci U S A. 1931;17(12):684-8.
2. O'Kelly M, Ratitch B. Clinical trials with missing data: a guide for practitioners. First ed: John Wiley & Sons, Ltd; 2014.

Appendix A. Protocol Deviations

The number and percentage of unique subjects reporting significant protocol deviations will be summarized in total and by treatment group for all randomized or treated subjects. The number and percentage of unique subjects with protocol deviation categories below will also be summarized in total and by treatment group.

- Subject entered into the study even though s/he did not satisfy entry criteria.
- Subject received wrong treatment or incorrect dose of study.
- Subject took prohibited concomitant medication or concurrent procedure.

Appendix B. Definition of Possible Distant Spread of Toxin

Table B-1. MedDRA Preferred Terms Evaluated for PDSOT

Accommodation disorder	Hyporeflexia
Aspiration	Hypotonia
Bell's Palsy	Ileus paralytic
Botulism	Muscular weakness
Bradycardia	Opthalmoplegia
Bulbar palsy	Paralysis
Constipation	Paresis cranial nerve
Cranial nerve palsies multiple	Pelvic floor muscle weakness
Cranial nerve paralysis	Peripheral nerve palsy
Diaphragmatic paralysis	Peripheral paralysis
Diplopia	Pneumonia aspiration
Dry mouth	Pupillary reflex impaired
Dysarthria	Respiratory arrest
Dysphagia	Respiratory depression
Dysphonia	Respiratory failure
Dyspnoea	Speech disorder
Eyelid function disorder	Urinary retention
Eyelid ptosis	Vision blurred
Facial paralysis	Vocal cord paralysis
Facial paresis	Vocal cord paresis

Note: Table is based on MedDRA 25.0; the actual list used for analysis will be based on the MedDRA version in use at the time of database lock.