| [redacted] | **Statistical Analysis Plan (SAP)** | croma est.1976 |
|---|---|---|

| | |
|---|---|
| **Protocol Title:** | Randomized double blind Phase 3 study to assess the efficacy and safety of BoNT/A-DP in the treatment of glabellar lines in comparison with placebo followed by an open label extension study |
| **Protocol Number:** | CPH-303-201400 |
| **Protocol Version, Date** | 2.0, 15 July 2019 |
| [redacted]**ID:** | 3428/0007 |
| **Document Version, Date:** | FINAL 3.0, 04 November 2020 |

Prepared by:

[redacted]

On behalf of:

CROMA-PHARMA GmbH

Confidentiality statement:

- The information provided in this document is strictly confidential.
- The recipients of the SAP must not disclose the confidential information contained within this document or any related information to other persons without the permission of the sponsor.
- In addition, the recipients of the SAP must keep this confidential document in a controlled environment which prevents unauthorized access to the document.

## SIGNATURE PAGE

**Prepared at [redacted] by:**

[redacted] 05 Nov 2020 05:57:024+0000

REASON: I approve this document as author.

a05dfa7f-b536-4636-ad10-1bb0fd2ce580

[redacted] Senior Statistician II — Date (DD Mmm YYY)

**Reviewed at [redacted] by:**

[redacted] 04 Nov 2020 14:29:017+0000

REASON: I approve this document

7623e655-56d0-484b-b5a8-9c837b94677c

[redacted], Principal Biostatistician — Date (DD Mmm YYY)

**Approved at CROMA-PHARMA GmbH by:** [redacted] 09 Nov 2020 08:35:048+0000

REASON: I approve this document

88228242-7e78-4a74-808e-30e7ff340344

[redacted] Clinical Project Manager — Date (DD Mmm YYY)

## REVISION HISTORY

| Version/Date | Version Name | Section | Changes implemented |
|---|---|---|---|
| Final 1.0 / 19 November 2019 | Final | N/A | N/A |
| Final 2.0 / 21 February 2020 | Amendment 1 | 7.1<br><br>8.8.1 | 1. Correction of study day derivation.<br>2. Following Site 110 audit findings, subgroup analyses added. |
| Final 3.0 / 04 November 2020 | Amendment 2 | 3.3, 8.2, 8.5.1, 8.8.1<br>11<br><br><br>8.7 | 1. New analyses due to COVID-19.<br>2. Appendix A: clarification added for the EOS re-mapping<br>3. Baseline definition clarified for open-label part |

# TABLE OF CONTENTS

SIGNATURE PAGE ..... 2
REVISION HISTORY ..... 3
TABLE OF CONTENTS ..... 4
LIST OF ABBREVIATIONS ..... 6
1 INTRODUCTION ..... 8
2 STUDY OBJECTIVES ..... 9
2.1 Primary Objective ..... 9
2.2 Secondary Objectives ..... 9
3 STUDY DESIGN ..... 10
3.1 General Study Design ..... 10
3.2 Randomization and Blinding ..... 12
3.3 Study Treatments and Assessments ..... 13
4 STUDY ENDPOINTS ..... 20
4.1 Primary Efficacy Endpoint ..... 20
4.2 Key Secondary Efficacy Endpoint ..... 20
4.3 Additional Secondary Efficacy Endpoints ..... 20
4.4 Safety Endpoints ..... 21
5 SAMPLE SIZE AND POWER ..... 22
6 ANALYSIS POPULATIONS ..... 24
6.1 Safety Analysis Set (SAF) ..... 24
6.2 Full Analysis Set (FAS) ..... 24
6.3 Modified Full Analysis Set (MFAS) ..... 24
6.4 Per-Protocol Analysis Set (PP) ..... 24
6.5 Protocol Deviations and Exclusions from Analysis Sets ..... 25
6.6 Data Review Meeting (DRM) ..... 25
7 Statistical Considerations and analysis ..... 26
7.1 Derived Variables ..... 26
7.2 Handling of Missing Data ..... 26
7.2.1 Missing Data Analysis Methods ..... 26
7.2.2 Handling of Missing or Incomplete Dates ..... 27
7.2.2.1 Imputation Rules for Missing or Partial AE Start Date: ..... 27
7.2.2.2 Imputation Rules for Missing or Partial Medication Start/Stop Dates ..... 28
8 STATISTICAL METHODS ..... 29
8.1 General Statistical Conventions ..... 29
8.2 Subject Disposition ..... 30
8.3 Protocol Deviations ..... 30
8.4 Demographics and Baseline Characteristics ..... 31
8.4.1 Demographics ..... 31
8.4.2 Baseline Characteristics ..... 31
8.4.3 Medical History ..... 31

8.4.4 Prior and Concomitant Medications ..... 31
8.5 Extent of Exposure ..... 32
8.5.1 Duration of Study Drug Exposure ..... 32
8.5.2 Treatment Compliance ..... 32
8.6 Efficacy Analyses ..... 32
8.6.1 Analysis Methods ..... 32
8.6.1.1 Multiplicity ..... 32
8.6.1.2 Treatment by Center Interaction Analysis (multi-center study) ..... 33
8.6.2 Analysis of Primary Efficacy Endpoint ..... 33
8.6.3 Analysis of Secondary Efficacy Endpoints ..... 34
8.6.4 Additional Secondary Analyses ..... 36
8.7 Safety Analyses ..... 39
8.7.1 Treatment Exposure ..... 39
8.7.2 Adverse Events ..... 39
8.7.3 Antibody Formation ..... 41
8.7.4 Laboratory Data ..... 41
8.7.5 Vital Signs ..... 42
8.7.6 Electrocardiograms ..... 42
8.7.7 Physical Examinations ..... 42
8.8 Other Analysis ..... 42
8.8.1 Subgroups ..... 42
8.9 Interim Analysis ..... 45
9 CHANGES TO PLANNED ANALYSIS FROM STUDY PROTOCOL ..... 49
10 REFERENCES ..... 50
11 APPENDICES ..... 51

# LIST OF ABBREVIATIONS

The following abbreviations will be used within this Statistical Analysis Plan (SAP).

| Abbreviation | Definition |
|---|---|
| ADA | Anti Drug Antibody |
| AE | Adverse Event |
| AESI | Adverse Event of Special Interest |
| ATC | Anatomical Therapeutic Chemical |
| BoNT/A | Botulinum Neurotoxin A |
| BoNT/A-DP | CROMA-PHARMA GmbH's BoNT/A Drug Product registered in Korea under the name "Botulax" |
| CETS | Composite Endpoint Treatment Success |
| CFR | Code of Federal Regulations |
| CI | Confidence Interval |
| CMH | Cochran-Mantel-Haenszel |
| CRA | Clinical Research Associate |
| CRF | Case Report Form |
| CS | Clinically Significant |
| CSR | Clinical Study Report |
| DRM | Data Review Meeting |
| ECG | Electrocardiogram |
| eCRF | Electronic Case Report Form |
| EOS | End of Study |
| EU | European Union |
| FAS | Full Analysis Set |
| FDA | Food and Drug Administration |
| FWS | Facial Wrinkle Scale |
| GCP | Good Clinical Practice |
| GL-QoL | Glabellar Line Quality of Life Scale |
| GLS-I | Glabellar Line Scale for Investigators |
| GLS-S | Glabellar Line Scale for Subjects |
| IA | Interim Analysis |
| ICF | Informed Consent Form |
| ICH | International Council for Harmonization |
| IMP | Investigational Medicinal Product |
| ITT | Intent-to-treat |

| | |
|---|---|
| IWRS | Interactive Web Response System |
| LOCF | Last Observation Carried Forward |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MFAS | Modified Full Analysis Set |
| NCS | Not Clinically Significant |
| PRO | Patient Reported Outcome(s) |
| PP | Per-protocol Analysis Set |
| PT | Preferred Term |
| ROS | Review of Systems |
| SAE | Serious Adverse Event |
| SAF | Safety Analysis Set |
| SAP | Statistical Analysis Plan |
| SD | Standard Deviation |
| SDTM | Study Data Tabulation Model |
| SOC | System Organ Class |
| SOPs | Standard Operating Procedures |
| TC | Telephone Call |
| TEAE | Treatment Emergent Adverse Event |
| TFLs | Tables, Figures, and Listings |
| US | United States |
| VAS | Visual Analog Scale |

# 1 INTRODUCTION

The purpose of this SAP is to provide detailed descriptions of the statistical methods, data derivations and data displays for study protocol CPH-303-201400 Version 2.0 "Randomized double blind Phase 3 study to assess the efficacy and safety of BoNT/A-DP in the treatment of glabellar lines in comparison with placebo followed by an open label extension study" dated 15 Jul 2019 for the planned interim and final analyses. The table of contents and templates for the tables, figures, and listings (TFLs) will be produced in a separate document.

Any deviations from this SAP will be described and justified in the Clinical Study Report (CSR).

The preparation of this SAP has been based on International Conference on Harmonization (ICH) E9 guideline and the Code of Federal Regulations (CFR) 21, part 11.

All data analyses and generation of TFLs will be performed using SAS 9.4® or higher.

# 2 STUDY OBJECTIVES

## 2.1 Primary Objective

To assess the efficacy of BoNT/A-DP in reducing the severity of glabellar frown lines following treatment (compared with placebo) based on investigator and subject assessment at week four (of the first treatment cycle).

## 2.2 Secondary Objectives

1. To assess the proportion of responders at maximum frown and at rest at various time points after each treatment, based on investigator and subject assessments.
2. To assess the psychological impact of BoNT/A-DP treatment on subjects (in terms of emotional and social functioning and concerns relating to their glabellar lines) in comparison with placebo after a single treatment.
3. To assess onset of effect and the duration of effect (at maximum frown) after a single treatment with BoNT/A-DP compared with placebo (first treatment cycle), based on investigator and subject assessments.
4. To provide long term safety data of BoNT/A-DP based on multiple treatment cycles and to establish a sufficient safety database to support regulatory approval.
5. To assess subject perceptions of effect of, and satisfaction with, treatment in comparison with placebo (first treatment cycle) and during the open label extension phase.

# 3 STUDY DESIGN

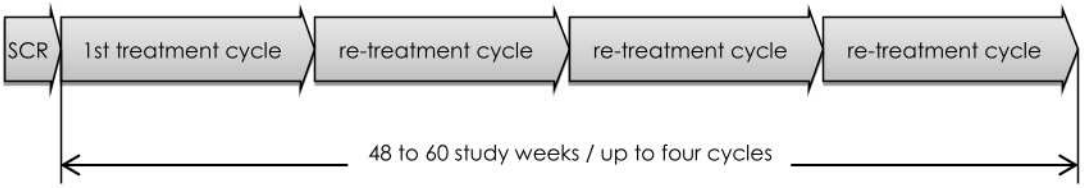
## 3.1 General Study Design

This multicenter Phase 3 study is comprised of two parts. The first part of the study is a randomized, double blind, placebo-controlled, phase which aims to demonstrate efficacy and safety of BoNT/A-DP compared with placebo. The second part is an open label extension phase to evaluate efficacy after repeat treatments and long term safety.

Subjects can receive a maximum of four treatment cycles over the duration of the study, a single treatment in the first cycle compared with placebo, and up to three subsequent treatments in the open label extension study.

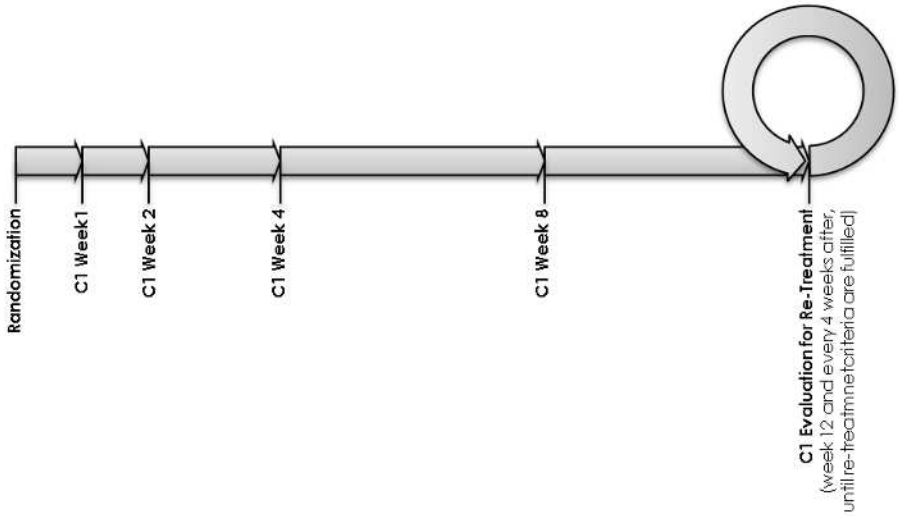
Subjects between 18 and 75 years of age, inclusive with moderate to severe glabellar frown lines at maximum frown with an important psychological impact as indicated by scores >0 on either the Emotional or Social Functioning subscales of the modified Skindex-16 (GL-QoL) will be enrolled in this study. Approximately 7 sites (United States (US) and Europe) will enroll in total 353 evaluable subjects with a variety of Fitzpatrick skin types. Applying a 15% drop out rate will result in a total number of completing subjects of about 300.

The overall study scheme is presented below:

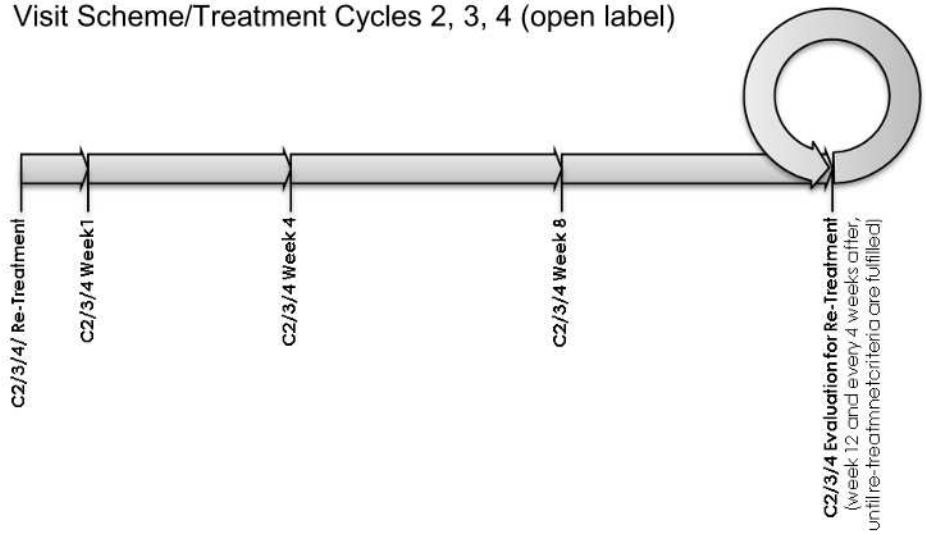



The duration of study participation for each subject will be up to 14 calendar days of screening followed by 48 to 60 weeks of treatment (re-screening will not be permitted). The first treatment cycle will be a double blinded cycle followed by up to three open label treatment cycles. Each cycle will be at least 12 weeks and can be prolonged in 4-week increments, depending on treatment effect. Re-treatment is possible until study-week 48.

Visit Scheme/Treatment Cycle 1 (double blind)




Visit Scheme/Treatment Cycles 2, 3, 4 (open label)



## 3.2 Randomization and Blinding

### 3.2.1 Randomization

This is a randomized, double blind, placebo-controlled efficacy and safety clinical study followed by an open label extension. Subjects will be randomly assigned to receive either BoNT/A-DP or placebo at a ratio of 3:1 respectively in the first treatment cycle. The 3:1 ratio in favor of active treatment versus placebo has been selected to increase the safety database size.

Randomization will be performed per study site via interactive web response system (IWRS). One unique randomization code will be assigned to each subject.

### 3.2.2 Blinding

For the first placebo controlled treatment cycle (the double blind cycle of this trial), BoNT/A-DP and placebo will be provided to the sites in glass vials. Since BoNT/A-DP-vials contain a pellet of lyophilized BoNT/A-DP, they can be discerned from placebo-vials (empty vials). However, after reconstitution by an unblinded study team member at the site, both will look identical (clear solution, comparable volume), thus maintaining the blind. The unblinded study team member preparing the Investigational Medicinal Product (IMP) at the study site must not, by any means, be involved in any other study data collection activities including adverse event (AE) assessment, case report form (CRF) completion, diary collection etc. IMP will be assigned to the subjects by IWRS with the lot number and kit number assigned corresponding to the group to which the subjects are assigned (blinded medication number and batch/lot number during the double blind phase of the study at the given site). The assigned vial will be reconstituted and will be forwarded to the investigator for injection. Specific Blinding Plans will be created at each study site during the Study Initiation visits.

### 3.2.3 Unblinding

The decision to unblind lies fully with the investigator. The randomization assignment should not be revealed before the study has been completed and the database has been cleaned and closed. The study will be unblinded using the Study Specific Unblinding Procedure (an unblinding module is standard on all blinded studies; also it is possible to grant access to regulatory unblinding users so that they can monitor the safety of the study, if required).

In case of emergency, the IMP administered to the subjects can be revealed using the unblinding function of the IWRS.

In rare emergencies, unblinding may be necessary for the clinical management of an AE. Investigators should consider unblinding only if knowledge of the administered product will have an influence on the further treatment of the AE. In such events, the investigator should make every attempt to inform the Sponsor before breaking the blind or as soon as possible after unblinding has been performed. The [redacted] or CROMA medical team is available to discuss any unblinding need. However, such discussion is not mandatory. The investigator can always unblind as per his/her discretion if the actual treatment information is considered relevant for subsequent event treatment. Once unblinding has occurred, the site should immediately contact the [redacted] or CROMA medical team. Communication of the unblinding result is considered acceptable. It is at the discretion of the investigator to continue an unblinded subject in the study. The date and time of breaking the code, the reason for breaking the code, study product administered, subject identification number and randomization code will be documented within the IWRS. Subjects for whom the blind had been broken may continue in the study as per discretion of the Investigator.

### 3.3 Study Treatments and Assessments

Subjects will participate in this study for a duration of 50 to 62 weeks from signing the informed consent form (ICF) to the End of Study visit (i.e., up to 14 calendar days for screening, followed by 48 to 60 weeks study participation).

Subjects will receive a maximum of four permitted rounds of treatment, with a minimum interval between treatments of 12 weeks.

The first treatment cycle of the study will comprise of two treatment groups as follows:

- Group A (active): BoNT/A-DP (20 units, 0.5 mL).
- Group B (placebo control): sterile, 0.9% sodium chloride (0.5 mL).

An interim analysis (IA) will be performed when all subjects complete the re-evaluation for retreatment visit at week 16 of the first treatment cycle or complete the double blind phase (whichever occurs earlier). Refer to Section 8.9 for more details regarding the IA.

Eligible subjects will be randomized at baseline (day 0) to Group A or B to receive the first treatment in a 3:1 randomization scheme, respectively. Investigators and subjects will be blinded to the treatment administered and will evaluate the severity of glabellar lines independently. The subjects must perform their assessment independently and ideally before the investigator, to ensure they are not biased by the investigator. The same investigator must assess the subject at baseline and at the visits at weeks 1, 2 and 4 in the first treatment cycle.

After a screening period of up to 14 calendar days, subjects will receive the first treatment (BoNT/A-DP or placebo) and attend for visits at 1, 2 and 4 weeks after treatment and at 4-

weekly intervals thereafter for evaluation of efficacy and safety (primary and key secondary efficacy endpoints will only be evaluated in the first treatment cycle in comparison with placebo). Re-screening will not be permitted.

The effect of botulinum toxin typically lasts a few months, hence the first treatment cycle will last at least 12 weeks and will end when the subjects qualify for re-treatment.

**Open Label Extension**

After the first treatment cycle is completed, all subjects may enter the open label extension phase and will be dosed with BoNT/A-DP (20 U) for subsequent re-treatments.

Subjects will attend for visits at one and four weeks after re-treatment and at 4-weekly intervals thereafter. Additional telephone call follow-ups will take place in the open label extension phase two and eight weeks after each re-treatment. Each re-treatment cycle will last at least 12 weeks and will end when the subject qualifies for re-treatment or with the End of Study visit. According to the study schedule, a maximum of four treatments per subject (four treatment cycles) is permitted during the study time frame, with treatments separated by a minimum of 12 weeks.

A subject may move to the next treatment cycle if more than 12 weeks have passed since the previous treatment. Starting at week 12 and at 4-weekly intervals thereafter, subjects will attend the site for the Evaluation for Re-treatment visit. If the subject has a Facial Wrinkle Scale (FWS) score (at maximum frown) of ≥2, as determined by both the subject using the Glabellar Line Scale for Subjects (GLS-S) and the investigator using the Glabellar Line Scale for Investigators (GLS-I), this visit on the same day at the same site visit will then be considered the End of Cycle visit and all other criteria required to determine eligibility for re-injection will be assessed e.g. negative urine pregnancy test (in women of child-bearing potential), lack of infection or inflammation in the planned injection area etc. Furthermore, all additional tests e.g. laboratory tests required for final cycle assessment will be performed. If the subject qualifies for re-treatment, on the same day at the same site visit, the subject will then enter the next treatment cycle and will receive the next study drug treatment as part of the Re-Treatment visit (i.e., day 0 of the next treatment cycle).

If the FWS score (at maximum frown) is assessed as ≥2 by both the investigator and the subject, but the additional criteria for reinjection are not met, e.g. relevant infection or inflammation at the injection site, the subject may attend for a visit 4 weeks thereafter. Subjects with a positive pregnancy test, or subjects who do not agree to re-treatment, will be

withdrawn from the study and the End of Study visit will be conducted at the earliest opportunity.

If at an Evaluation for Re-treatment visit the subject wants to get re-treated, but the investigator does not agree or vice versa (i.e., the FWS score is not ≥ at least 2 by both assessments), the subject will return to the site four weeks thereafter for another Evaluation for Re-treatment visit. There is no limit to the number of Evaluation for Re-treatment visits. In order to consider a visit as the End of Cycle visit, there must always be agreement between investigator and subject on a FWS score (at maximum frown) of ≥2.

**End of Study Visit**

The End of Study visit can take place in four-weekly intervals from study-week 48 until study-week 60. The last study drug re-treatment will be administered no later than week 48. For subjects receiving a re-treatment at week 48, the last cycle will end 12 weeks later at week 60 (i.e., End of Study visit).

The End of Study visit will take place at week 48 if the week 48 visit is an Evaluation for Re-treatment visit and the subject does not meet the criteria for re-treatment in case this visit is 12 or more weeks after the last treatment. As week 48 is the last possible time point for re-treatment, such subjects will continue with all additional assessments of the End of Study visit on the same day. For subjects having received their last injection eight weeks or less prior to week 48, the End of Study visit will take place 12 weeks after the last study drug treatment was given, e.g. week 52 if the subject was re-treated at week 40.

For subjects who are prematurely discontinued from the study (at any time), the End of Study visit will take place within one week of discontinuation.

A detailed description of procedures and assessments to be conducted during this study is summarized in the Scheduled of Study Assessments in Table 1 below.

**COVID-19 Pandemic**

Due to the COVID-19 pandemic, which touched the world mainly the first 6 months of 2020, it has been allowed that visits were done remotely instead of physically when site was closed.

All re-treatments were put on hold between the 16th March and 5th May 2020 due to concerns that the subject will not be able to return for Week 1 and Week 4 assessments. Consequently, when a subject was eligible for re-treatment at a visit, the injection has been delayed until the subject could go back to site (if the injection is done before or at Week 48).

Subjects have been classified as study completer if the subject was eligible for re-treatment before or at Week 48 but never re-treated due to pandemic hold. More instructions can be found in the CRF completion guidelines document.

## Table 1: Schedule of Study Assessments

| | Double-blind phase | | | | | | | | Open label phase | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cycle 1 (C1) | | | | | | | | Cycle 2 (C2)[1] | | | | | | | Cycle 3 (C3)[1] | | | | | | | | Cycle 4 (C4)[1] | | | | | | |
| **Procedures and assessments** | 1 Screening (Day -14 to -1) | 2 C1 Randomization and 1st treatmen (Baseline, Day 1) | 3 C1 Week 1 | 4 C1 Week 2 | 5 C1 Week 4 | 6 C1 Week 8 | 7 C1 Evaluation for Re-treatment[2,3] (C1 Weeks 12, 16, 20, ...) | 7 C1 End of Cycle Visit[2] | 7 C2 Re-treatment[3] | 8 C2 Week 1 | 9 C2 Week 2 (TC) | 10 C2 Week 4 | 11 C2 Week 8 (TC) | 12 C2 Evaluation for Re-treatment[2,3] (C2 Weeks 12, 16, 20, ...) | 12 C2 End of Cycle Visit[2] | 12 C3 Re-treatment[3] | 13 C3 Week 1 | 14 C3 Week 2 (TC) | 15 C3 Week 4 | 16 C3 Week 8 (TC) | 17 C3 Evaluation for Re-treatment[2,3] (C3 Weeks 12, 16, 20, ...) | 17 C3 End of Cycle Visit[2] | 17 C4 Re-treatment[3] | 18 C4 Week 1 | 19 C4 Week 2 (TC) | 20 C4 Week 4 | 21 C4 Week 8 (TC) | 22 End of Study[4] |
| Visit No*/** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 7 | 7 | 8 | 9 | 10 | 11 | 12 | 12 | 12 | 13 | 14 | 15 | 16 | 17 | 17 | 17 | 18 | 19 | 20 | 21 | 22 |
| Informed consent | x | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Consent for re-treatment | | | | | | | | | x | | | | | | | x | | | | | | | x | | | | | |
| Eligibility: Inclusion/Exclusion | x | x | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Eligibility for re-treatment | | | | | | | x | | | | | | | x | | | | | | | x | | | | | | | |
| Medical history | x | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Demographics | x | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Physical exam*** | | x | x | x | x | | | | | x | | x | | | | | x | | x | | | | | x | | x | | x |
| Pregnancy test[5] | x | x | | | | | | x | | | | | | | x | | | | | | | x | | | | | | x |
| Vital signs[6] | x | x | x | | x | | | x | | | | x | | | x | | | | x | | | x | | | | x | | x |
| Clinical laboratory assessments[7] | x[§] | | | | x | | | x | | | | x | | | x | | | | x | | | x | | | | x | | x |
| ECG | x[§] | | | | x | | | x | | | | | | | | | | | | | | | | | | | | | |
| Concomitant medication | x | x | x | x | x | x | x | | | x | x | x | x | x | | | x | x | x | x | x | | | x | x | x | x | x | x |
| FWS Investigator[8] | x | x | x | x | x | x | x | | | | | x | | x | | | | | x | | x | | | | | x | | x |
| FWS Subject self-assessment[9] | x | x | x | x | x | x | x | | | | | x | | x | | | | | x | | x | | | | | x | | x |
| Photography[10] | | x | | x | x | | x | | | | | | | | | | | | | | | | | | | | | |
| Psychological impact[11] | x | | | | x | | | x | | | | x | | | x | | | | x | | | x | | | | x | | x |
| Treatment satisfaction[12] | | | | | x | | | x | | | | x | | | x | | | | x | | | x | | | | x | | x |
| IMP administration | | x | | | | | | | x | | | | | | | x | | | | | | | x | | | | | |
| Post treatment obs./AE assessment[13] | | x | | | | | | | x | | | | | | | x | | | | | | | x | | | | | |
| AE and AESI assessment[15] | | x | x | x | x | x | x | | | x | x | x | x | x | | | x | x | x | x | x | | | x | x | x | x | x | x |
| Antibody test[14] | x[§] | | | | x | | | x | | | | x | | | x | | | | x | | | x | | | | x | | x |
| Subject diary | | D | R/D | R | | | | | | | | | | | | | | | | | | | | | | | | |

TC = telephone contact: including assessment of concomitant medication, AEs and AESIs[15] (as per AESI manual) on weeks 2 and 8 of each open label cycle (visits 9, 11, 14, 16, 19 and 21). If an AESI is reported during a telephone contact, subjects will be asked to come to the site for further assessment including a targeted physical examination as soon as possible) * Visit numbers correspond with specific treatment and assessments. If subjects are not eligible for re-treatment at the "Evaluation for Re-treatment" visit, they will return at 4 weekly intervals thereafter, until they are eligible for treatment. The subsequent visit number for the cycle will remain the same, but will be amended with "a, b, c" etc., corresponding to the time extended (at 4 weekly intervals) in that treatment cycle. For example, the first evaluation for re-treatment is at visit 7, week 12, however for subjects who are not eligible at week 12, but at week 16, this visit will be denoted “visit 7a", while "visit 7b" will denote week 20 in the first cycle and "visit 7c" will denote week 24 etc., with each additional letter corresponding to a time point 4 weeks later in that cycle.

** Time differences are measured from the previous treatment administered. A time deviation of ± 5 days is allowed for each visit, except for the week 1 and week 4 visit in each treatment cycle where a time deviation of ± 2 days is permitted. Sites must adhere to the schedule of events and visit windows and subjects must ensure they are available for those visits. However, if for practical reasons e.g. public holidays, a visit window cannot be met, a visit can be scheduled as close as possible to that visit window. Any deviation from the visit schedule and its associated time windows will still be documented as a protocol deviation.

*** Full physical examination will include neurological assessment (including extraocular movements and cranial nerves) as well as assessment for muscle weakness. In addition, if the subject reports any symptoms related to Adverse Events of Special Interest (as detailed in the AESI Manual) a focused physical examination, to evaluate these symptoms will also be undertaken.

§ The ECG performed at screening is the baseline ECG. Laboratory and Anti-drug-antibody (ADA) test samples from screening are considered baseline values.

1 Depending on the duration of treatment effect, a maximum of 4 treatments is permitted.

2 Evaluation for re-treatment takes place at the earliest at 12 weeks after the first/previous treatment. Subjects who do not qualify for re-treatment at week 12 will have the option (pending eligibility) for re-treatment at 4-weekly intervals thereafter, until they are eligible for re-treatment. Once the subject is eligible for re-treatment, the end of cycle procedures will take place and the subject can receive re-treatment (i.e. the end of cycle visit and the re-treatment visit will be conducted on the same day as the Evaluation for Re-treatment visit).

3 The latest time for re-treatment is study week 48; if a subject is not eligible for re-treatment at study week48, the End of Study visit will be completed.

4 The End of Study visit can take place in four-weekly intervals from study-week 48 until study-week 60. For subjects receiving re-treatment at week 48, the last cycle will end 12 weeks later at week 60 (=End of Study visit). End of Study visit will take place at week 48 if week 48 visit is an Evaluation for Re-treatment visit and the subject is not meeting the criteria for re-treatment. For subjects having received their last injection 8 weeks or less prior to week 48, the End of Study visit will take place 12 weeks after the last study drug treatment was given, e.g. week 52 if subject was retreated at week 40. For subjects that get prematurely discontinued from the study (at any time) the End of Study visit will take place within 1 week of discontinuation.

5 Only in women of child-bearing potential; blood serum test at screening and End of Study visit; otherwise urine dip stick.

6 Vital signs include blood pressure (diastolic /systolic) and pulse.

7 Laboratory Assessments: Hematology and Serum Chemistry, details in Section 11.15 of the study protocol..

8 Investigator's assessment of glabellar line severity at maximum frown and at rest. Assessment will be made using the 4-point FWS [GLS-I] (0=none, 1=mild, 2=moderate, 3=severe) and will be recorded in the CRF.

9 Subject's assessment of glabellar line severity at maximum frown and at rest. Assessment will be made using the 4-point FWS [GLS-S] (0=none, 1=mild, 2=moderate, 3=severe). Subject Assessment will be performed before (ideally) and always independently of investigator assessment and will be recorded in the CRF)

10 Photographs of subject's glabellar lines (at maximum frown and at rest) will be taken at C1 Randomization, C1 week 2, C1 week 4 and each C1 Evaluation for Re-Treatment visit in Cycle 1 until the subject qualifies for re-treatment in order to provide evidence of effect and confirm that the subject did not qualify for re-treatment on the preceding evaluation for re-treatment visit. Although a photograph will be taken at each Evaluation for Re-treatment visit, only the photo of the visit preceding the re-treatment visit will be reviewed by the independent reviewers.

11 Modified Skindex-16 (GL-QoL) and FACE-Q Appraisal of Lines Between Eyebrows and Age Appraisal VAS scales will be used to measure psychological impact and concerns relating to their glabellar lines, respectively (and will be recorded in the CRF).

12 Treatment satisfaction will be determined using the FACE-Q Satisfaction with Outcome Scale (will be recorded in the CRF).

13 Subjects will be monitored for AEs during 30 minutes after administration of the IMP. No additional questioning of AESIs directly post-injection.

14 Anti-drug-antibody (ADA) test. If positive, serum samples will be tested for the presence of neutralizing antibodies.

15 General, non-leading AE questioning as well as active AESI questioning at each indicated visit. The first AESI questioning will be completed at Baseline visit in order to obtain a full baseline status of any concomitant diseases prior to the first IMP injection.

AESI Questioning: active questioning by guided review of systems (ROS) as per AESI manual. If an AESI is reported, a targeted physical examination around the area of the reported AESI must follow.

AE: Adverse event; AESI: Adverse event of special interest; C: Cycle; CRF: Case report form; D: Distribution of subject diary; ECG: Electrocardiogram; FWS: Facial Wrinkle Scale; GL-QoL: Glabellar Line Quality of Life Scale; IMP: Investigational medicinal product; R:Return of subject diary; TC: telephone call; VAS: Visual analog scale.

# 4 STUDY ENDPOINTS

## 4.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the proportion of subjects among BoNT/A-DP and placebo groups with a FWS score of 0 or 1 and an improvement of ≥2 points in FWS score (at maximum frown) at the week 4 visit (of the first treatment cycle) relative to baseline (responders), based on both the investigators' and the subjects' in-clinic assessments. Thus, the primary endpoint is a composite endpoint comprising investigator and subject assessments of treatment effectiveness.

## 4.2 Key Secondary Efficacy Endpoint

The key secondary efficacy endpoint of this study is:

1. The extent of change in psychological impact (emotional and social functioning and concerns relating to their glabellar lines) at week 4 after the first treatment, in the BoNT/A-DP group in comparison with placebo, relative to baseline, as assessed by the modified Skindex-16 (Glabellar Line Quality of Life Scale [GL-QoL]) and the FACE-Q (Appraisal of Lines Between Eyebrows and Age Appraisal visual analog scales (VAS)), respectively.

## 4.3 Additional Secondary Efficacy Endpoints

The additional secondary efficacy endpoints of this study are:

2. The percentage of responders at maximum frown (as defined in 4.1 above) at week 12 (after the first treatment with BoNT/A-DP or placebo).
3. The percentage of responders at week 16 (after the first treatment).
4. The percentage of responders at week 20 or later (after the first treatment).
5. The proportion of subjects with a ≥1 point reduction in FWS score at rest at week 4 in the first treatment cycle, based separately on the investigators' and the subjects' in-clinic assessments (applicable only for subjects who have a FWS score at rest ≥1 at baseline).
6. The proportion of responders among BoNT/A-DP and placebo groups with a FWS score at maximum frown of 0 or 1 and an improvement of ≥2 points in FWS score (at maximum frown) during the first treatment cycle visit relative to baseline, based on both the investigators' and the subjects' in-clinic assessments (composite endpoint, at weeks 1, 2 and 8).
7. The proportion of subjects with ≥2 point and ≥1 point reduction in FWS score (at maximum frown) in the BoNT/A-DP and placebo groups during the first treatment cycle

visit relative to baseline, based on the independent rater's assessment of photographs (at baseline and visits 2, 4, 12, 16 and 20 weeks after treatment, within the first treatment cycle).

8. Time to onset of effect in the BoNT/A-DP and placebo groups in the first treatment cycle, as measured at weeks 1, 2 and 4 based separately on subject and investigator assessments. Onset of effect defined as at least a 1 point improvement in FWS score from baseline (at maximum frown). In addition, onset of effect will be assessed by subjects daily in the first 2 weeks after treatment, by recordings in the subject diary.
9. The extent of subject perceptions of effect of, and satisfaction with, treatment, in the BoNT/A-DP and placebo groups, during each treatment cycle, as assessed by the FACE-Q Satisfaction with Outcome Scale.
10. The proportion of subjects with a ≥1 point reduction in FWS score at rest in the BoNT/A-DP and placebo groups, relative to baseline, during the first treatment cycle, based on the independent rater's assessment of photographs.
11. The percentage of subjects with a FWS score of 0 or 1 and an improvement of ≥2 points in FWS score at maximum frown (investigator and subject assessment) at four weeks after re-treatment relative to the rating at the preceding end of cycle visit.
12. The proportion of subjects with ≥1 point reduction in FWS score (at maximum frown) in the BoNT/A-DP and placebo groups during the first treatment cycle at weeks 1, 2, 4, 8, 12, 16 and 20 relative to baseline, based on both the investigators' and the subjects' in-clinic assessments.
13. The proportion of subjects with ≥1 point reduction in FWS score (at maximum frown) in the BoNT/A-DP group during each re-treatment cycle at week 4 relative to re-treatment-baseline, based on both the investigators' and the subjects' in-clinic assessments.

### 4.4 Safety Endpoints

The safety endpoints of this study are:

1. Frequency, severity and causal relationship of AEs, Serious Adverse Events (SAEs) and Adverse Events of Special Interest (AESIs) during the entire study period.
2. Antibody formation, evaluation pre-dose before each treatment, at four weeks after each treatment and at the final study visit.
3. Safety assessments by evaluating hematology, clinical chemistry, vital signs and ECGs as per study schedule.

# 5 SAMPLE SIZE AND POWER

The primary efficacy endpoint is a composite endpoint comprising investigator and subject assessments of treatment effectiveness using the FWS. Composite endpoint treatment success (CETS) is defined as ≥2 point reduction in FWS score at maximum frown achieving a score of 0 or 1 at the visit week 4 relative to baseline based on investigator assessment and subject assessment. Thus, a subject is a CETS "Responder" only if both investigator- and subject-rated success criteria are satisfied at the visit 4 weeks after baseline.

The primary analysis of efficacy is the proportion of responders in the BoNT/A-DP treatment group compared with the placebo treatment group. For superiority testing of BoNT/A-DP compared with placebo, the following assumptions for the sample size calculation will be used:

| ***Item*** | ***Assumption*** | ***Comments*** |
|---|---|---|
| Underlying test | Fisher's Exact Test | Standard test for a response parameter for small numbers |
| Randomization scheme (BoNT/A-DP: Placebo) | 3:1 | More active treatment subjects to enlarge safety database |
| Power 1-β | 90% | Actual power is 92.3% |
| Significance level (α) | 0.025 one-sided | |
| Response Arm BoNT/A-DP | 46% | Conservative value based on the results of previous studies (BLESS I and BLESS II) |
| Response Arm Placebo | 2% | Conservative value based on the results of previous studies (BLESS I and BLESS II) |
| Software | | The sample size calculation was performed using the software nQuery Advisor® 8.2.1.0 |

Based on these assumptions, 39 subjects in the BoNT/A-DP arm and 13 subjects in the placebo arm are required, i.e., 52 subjects in total. This is a very small sample size and not sufficient for a detailed evaluation of the data, including subgroup analyses. Furthermore, the response rate could not be estimated with sufficient precision. Assuming a response rate of 46% in the BoNT/A-DP group, a two-sided confidence interval (CI) of 95% with a distance from the response rate to the CI limits of about 5.6% could be achieved with a sample size of 225 subjects in the BoNT/A-DP group. A 3:1 randomization of BoNT/A-DP (225 subjects) to placebo (75 subjects) is proposed, which is adequate for a precise estimate of response rate and for post-hoc sensitivity analyses (i.e., treatment-by-center and other subgroup analyses), also taking into account a drop-out rate.

In addition the sample size is justified in order to fulfill FDA request for additional subjects treated with BoNT/A-DP to provide an adequate Safety database.

In total, 353 evaluable subjects will be enrolled. Applying a 15% drop out rate will result in a total number of completing subjects of about 300.

# 6 ANALYSIS POPULATIONS

## 6.1 Safety Analysis Set (SAF)

All subjects who received at least one injection with study medication (independent of whether it is BoNT/A-DP or placebo) will be valid for the SAF. Within the SAF, a subject will be considered for the treatment actually received and not for the treatment assigned by randomization, if different. The SAF will be used for the evaluation of the safety assessments and for the individual subject data listings.

## 6.2 Full Analysis Set (FAS)

The FAS includes all randomized subjects, regardless of whether they received study medication. Within the FAS, a subject will be considered for the treatment assigned by randomization and not for the treatment actually received, if different, i.e., following the intent-to-treat (ITT) principle. The FAS will be used for the evaluation of the efficacy assessments. The FAS serves as the primary efficacy analysis set.

## 6.3 Modified Full Analysis Set (MFAS)

The MFAS includes all randomized subjects who received at least one injection with study medication who had a baseline (visit 1 at day 0) and at least one post-dose in-clinic assessment with the FWS by either the investigator or the subject on visits at weeks 1, 2 or 4. Within the MFAS, a subject will be considered for the treatment assigned by randomization and not for the treatment actually received, if different, i.e., following the ITT principle. The MFAS will be used for the evaluation of the efficacy assessments.

## 6.4 Per-Protocol Analysis Set (PP)

The PP includes all randomized subjects who received at least one injection with study medication who had no significant protocol deviations and an in-clinic assessment with the FWS by the investigator and the subject at baseline (day 0) and at the week 4 visit. For the PP, all subjects will be assigned to the randomized treatment. The PP will only be analyzed for main efficacy outcome measures.

### 6.5 Protocol Deviations and Exclusions from Analysis Sets

Protocol deviations and exclusions of subjects from analysis sets will be identified at the Data Review Meeting (DRM) just prior to study unblinding for the IA.

Deviations from the protocol will be classified as major or minor.

The following protocol deviations are considered a priori to be major and will lead to exclusion from the PP population:

- Subject or investigator unblinded to treatment
- Improper storage of IMP and IMP temperature excursion without proper notice to the CRA
- Incorrect treatment allocation or dose
- Wrong injection points used

### 6.6 Data Review Meeting (DRM)

For the DRM appropriate listings displaying all relevant data will be provided to the Sponsor and serve as a source for the protocol deviations discussion about the classification into major and minor. No unblinding will be done for the creation of these listings.

The following protocol deviations will be assessed on a case-by-case basis and a final decision on a possible exclusion from the PP population will be done during the DRM:

- Inclusion/exclusion criteria violated
- Use of disallowed medications
- Serious breach of GCP
- Site staff performing the study assessment not trained
- Failure to discontinue subjects where appropriate

Subjects with minor protocol deviations will not be excluded from the PP population.

Unblinding will be done after the DRM has been conducted and the minutes have been signed.

Following the IA, the protocol deviations described above will be reviewed during the remainder of the study. In addition, the following protocol deviations will also be reviewed during the remainder of the study, however these will not impact on decisions made on the analysis populations:

- Subject re-treated despite not meeting all re-treatment eligibility criteria
- Photography not collected or available for all visits.

# 7 STATISTICAL CONSIDERATIONS AND ANALYSIS

## 7.1 Derived Variables

The below table provide the list of derived variables for various duration derivations and baseline derivations applicable for this study.

| Variables | Formula |
|---|---|
| **Derivation of Duration** | |
| Study day at any visit | Date of interest – date of first dose of study drug, if date of interest is before date of first dose of study drug.<br>Date of interest – date of first dose of study drug + 1, otherwise. |
| Extent of Exposure (Days) | Last exposure date – Date of first study medication administration + 1<br>where Last exposure date = Maximum of (last study medication administration date, last visit date, end of study date). |
| Extent of Exposure (Weeks) | Extent of exposure (days)/7 |
| **Baseline Derivations** | |
| Baseline | The baseline value is defined as the last observation before treatment, i.e., pre-treatment values measured on the treatment day (baseline visit, day 0), and if missing or not evaluated, the value from screening visit. |
| Change from baseline | Post baseline value – Baseline value |
| Relative change from baseline | [(Post baseline value – Baseline value)/Baseline value]*100 |

## 7.2 Handling of Missing Data

### 7.2.1 Missing Data Analysis Methods

Analyses will be performed by visit. For responder analyses on visits at week 4 and week 12, missing in-clinic assessments (investigator or the subject) with the FWS at baseline or week 4 and week 12 visits will be assigned as being non-responders.

As sensitivity analyses, the following additional approaches for handling of missing values will be applied:

- Analysis on observed values only, i.e., missing values will be excluded from analysis.
- Last observation carried forward (LOCF).
- Tipping point analysis: each missing value will be assigned to either a response or non-response, so that all possible combinations of replacing one or more missing values within each treatment group will be analyzed. P-values will be calculated for each combination and graphically displayed using a scatter plot.

The analysis of the week 16 visit regarding the efficacy endpoint "percentage of responders at week 16 (after first treatment)" will be conducted on observed values only, i.e., missing investigator or the subject in-clinic assessments with the FWS at baseline or week 16 visit will be excluded from analysis but not assigned as being non-responders. Subjects who were re-treated before visit week 16 are considered to be non-responders. The same approach holds for the efficacy endpoint "percentage of responders at week 20 or later (after first treatment)" respectively. The analysis of the additional secondary efficacy endpoints "proportion of subjects with ≥1 point reduction in FWS score (at maximum frown) in the BoNT/A-DP and placebo groups during the first treatment cycle at week 16 and week 20 relative to baseline, based on both the investigators' and the subjects' in-clinic assessments" will be conducted in a similar manner on observed values only.

All analyses of the modified Skindex-16 (GL-QoL) Emotional and Social Functioning domains and overall scores will be conducted on observed values only, i.e., missing domain and overall scores will be excluded from analysis; no imputation of missing domain and overall scores will be performed. The same approach will be followed for the analyses of the FACE-Q scales.

The analysis of "responder" 4 weeks after re-treatment will be conducted on observed values only, i.e., subjects with missing investigator or the subject in-clinic assessments on FWS week 4 or at the preceding end of the cycle visit will be excluded from analysis.

### 7.2.2 Handling of Missing or Incomplete Dates

#### 7.2.2.1 Imputation Rules for Missing or Partial AE Start Date:

If the cycle of onset is missing, the following rules will be applied.

**If only Day of AE start date is missing:**

If the AE start year and month are the same as that for the first dose date, then:

- If the full (or partial) AE end date is NOT before the first dose date or AE end date is missing, then impute the AE start day as the day of first dose date.
- Otherwise, impute the AE start day as 1.

Otherwise, if the AE start year and month are not the same as that for the first dose date, then impute the AE start day as 1.

Compare the imputed AE start date with the first dose date to determine whether the AE is pre-treatment AE or treatment emergent adverse event (TEAE).

**If only Day and Month of AE start date are missing:**

If AE start year = first dose year, then:

- If the full (or partial) AE end date is NOT before the first dose date or AE end date is missing, then impute the AE start Month and Day as the Month and Day of first dose date;
- Otherwise, impute the AE start Month as January and the AE start Day as 1.

Otherwise, if AE start year not equal first dose year, then impute the AE start Month as January and the AE start Day as 1.

Compare the imputed AE start date with the first dose date to determine whether the AE is pre-treatment AE or TEAE.

**If Year of AE start date is missing:**

If the year of AE start is missing or AE start date is completely missing then query site with no imputation. Also compare the full (or partial) AE end date to the first dose date. If the AE end date is before the first dose date then the AE should be considered as a pre-treatment AE. Otherwise, the AE will be considered as TEAE.

#### 7.2.2.2 Imputation Rules for Missing or Partial Medication Start/Stop Dates

**Missing or partial medication start date:**

- If only Day is missing, use the first day of the month.
- If Day and Month are both missing, use the first day of the year, i.e., 01 January.
- If Day, Month and Year are all missing, use the day before the first dose date.

**Missing or partial medication stop date:**

- If only Day is missing, use the last day of the month.
- If Day and Month are both missing, use the last day of the year.
- If Day, Month and year are all missing, assign 'continuing' status to stop date

# 8 STATISTICAL METHODS

## 8.1 General Statistical Conventions

All statistical procedures will be completed using SAS version 9.4 or higher.

The statistical analysis of the IA will be performed after all subjects finalize the re-evaluation for retreatment visit at week 16 of the first treatment cycle or complete the double-blind phase (whichever occurs earlier). Unblinding will occur when the analysis populations have been determined based on a blinded data review after database closure for all data belonging to the double-blind phase up to and including the re-evaluation for retreatment visit at week 16.

The statistical analysis of the final analysis of the first treatment cycle and the open label extension phase (cycles 2 to 4) will be performed after database lock of all data up to the end of the study.

The statistical testing of the primary efficacy endpoint as well as the key secondary efficacy endpoint will be one-sided and will be performed using a significance (alpha) level of 0.025. A two-sided McNemar test (using a significance level of 0.05) will be used to compare the response proportions between week 4 re-treatment visit and at the preceding end of the cycle visit. This endpoint will be separately analyzed for each open label extension cycle. Two-sided 95% CIs will be provided when relevant.

Continuous variables will be summarized using descriptive statistics, including number of subjects (n), mean, standard deviation (SD), median, minimum and maximum. One additional decimal point for mean and median and two additional decimal points for SD will be used in addition to the number of decimal points used for the measured values.

For categorical variables, summaries will include counts of subjects and percentages in corresponding categories. Percentages will be rounded to one decimal place.

For statistical analyses “baseline” refers to the last observation before treatment, i.e., pre-treatment values measured on the treatment day (baseline visit, day 0), and if missing or not evaluated, values from screening visit will be used.

All summaries will be presented by treatment group, unless otherwise specified.

Analyses will be performed by visit, irrespective of any time window deviations.

All subject data, including those derived, will be presented in individual subject data listings. Unless otherwise stated, unscheduled visit results will be included in date/time chronological order, within subject listings only. All listings will be sorted by investigational site, subject number, date/time and visit. The treatment group will be stated on each listing. Unless

otherwise stated, data listings will be based on the SAF population. A listing with demographic data for screening failures will be presented.

### 8.2 Subject Disposition

Subject disposition information will be summarized for the double blind phase by treatment group and overall as well as separately for the open label phase. The number and percentage of subjects who are randomized, who obtained a dose of study drug, who were randomized and not treated, who were treated and not randomized, who complete the double blind treatment phase, who complete the study and who withdraw early from the study will be presented.

The primary reason for early withdrawal will also be tabulated.

Subject disposition will be listed.

The number and percentage of subjects in each analysis set will also be tabulated.

Treatment Misallocations:

If a subject was:

- Randomized but not treated, then they will be reported under their randomized treatment group. However, they will be excluded from the safety analyses as actual treatment is missing,
- Treated but not randomized, then by definition they will be excluded from the efficacy analyses since they are not randomized, but will be included in the safety analyses.
- Randomized but got incorrect treatment, then they will be reported under their randomized treatment group for all efficacy analyses, but will be reported under the treatment they actually received for all safety analyses.
- Subjects who were not treated as randomized will be excluded from the per-protocol analysis.

The visits or assessments not done as per protocol due to COVID-19 pandemic will be also listed for each subject.

### 8.3 Protocol Deviations

The number of subjects excluded from SAF, FAS, MFAS, and PP populations and reasons for exclusion will be summarized by treatment group and overall.

Analysis set membership details will be listed, including reason for exclusion from each analysis set.

All major protocol deviations identified will be summarized by treatment group and overall. Minor protocol deviations will be listed only.

A listing will include the inclusion/exclusion criteria violated at Screening and at Re-treatment on cycles 2, 3 and 4 as well as other protocol deviations identified based on data recorded on the electronic CRF (eCRF) and/or protocol deviation logs from [redacted] (based on the SAF population).

## 8.4 Demographics and Baseline Characteristics

### 8.4.1 Demographics

Age, sex, race, ethnicity, height, weight, body mass index, and Fitzpatrick skin type at baseline will be summarized descriptively by treatment group and overall using the FAS population. The FDA guideline regarding "Collection of race and ethnicity data in clinical trials" will be followed.

### 8.4.2 Baseline Characteristics

The categorical baseline characteristics such as baseline ECG, pregnancy test results and antibody test sample collection (yes/no) at screening will be summarized using frequency counts for the FAS population. Continuous baseline variables such as systolic blood pressure (mmHg), diastolic blood pressure (mmHg) and pulse rate (beats/min) will be summarised by descriptive statistics in the same way as continuous demographic variables for the FAS population.

### 8.4.3 Medical History

A summary of medical history will be presented by system organ class (SOC) and preferred term (PT) using Medical Dictionary for Regulatory Activities® (MedDRA) Version 22.0 or higher. The table will be based on the SAF population. The previous toxin treatment will be listed only.

### 8.4.4 Prior and Concomitant Medications

Medications used in this study will be coded by using the latest available version of the World Health Organization Drug Dictionary Standard or Enhanced and categorized as follows:

Prior medications and concomitant medications will be summarized descriptively using frequency tables by anatomical therapeutic chemical (ATC) class and preferred name by treatment group on the FAS population and presented separately for the following groups:

- Medication (recent) discontinued prior to baseline (day 0)

- Concomitant medication started at or after baseline, or started before baseline and were not discontinued prior to baseline

Details for imputing missing or partial start and/or stop dates of medication are described in Section 7.2.2.

## 8.5 Extent of Exposure

### 8.5.1 Duration of Study Drug Exposure

Duration of study drug exposure (in days) will be calculated as date of last study medication exposure minus date of first study medication administration + 1 day, regardless of study drug interruption. Details of study drug exposure derivation is described in Section 7.1.

Study drug exposure will be summarised by treatment group and separated for the double blind and open label phase on the SAF population using time intervals and summary statistics. Planned and actual dose as well as injections at all sites per subject and treatment cycle will be listed.

Due to the COVID-19 pandemic and the treatment hold, the duration of study drug exposure will be also derived as described above but excluding the days between the end of cycle visit and the re-treatment visit between each cycle. This will be summarized using summary statistics.

### 8.5.2 Treatment Compliance

All study procedures are to be performed under supervision at the study site, and thus, no separate procedures will be used to monitor subject compliance.

## 8.6 Efficacy Analyses

### 8.6.1 Analysis Methods

The analyses of the primary and secondary outcome measures will be based on different statistical tests such as Cochran-Mantel-Haenszel (CMH) test, Pearson Chi-Square test, Wilcoxon Rank Sum test and Fisher's exact test, if appropriate. Details for each efficacy endpoint will be provided in the following sections. Subgroup analyses will be described in Section 8.8.1.

#### 8.6.1.1 Multiplicity

The testing of the primary and the key secondary efficacy endpoints will be performed with appropriate multiplicity control based on the FAS population. The results of each test will only be considered confirmative if the previous test in the order showed a confirmatory result at a one-sided significance level of 0.025. If the one-sided p-value is larger than 0.025 for

any of the tests, the results of the subsequent tests will only be considered exploratory, and not confirmatory. The application of this hierarchical approach keeps the global significance level to 0.025 one-sided and requires no further adjustment of the significance level.

The following order of tests will be applied:

- Primary endpoint (composite endpoint)
- Key secondary endpoint 1

  1.1: The modified Skindex-16 (GL-QoL) Emotional domain
  1.2: The modified Skindex-16 (GL-QoL) Social Functioning domain
  1.3: The modified Skindex-16 (GL-QoL) Overall score
  1.4: The FACE-Q Appraisal of Lines Between Eyebrows scale
  1.5: The FACE-Q Age Appraisal VAS score

All other analyses (e.g., additional analyses of the primary and the key secondary efficacy endpoints using other analysis populations or other statistical methods, analyses of additional secondary endpoints) will only be considered exploratory.

#### 8.6.1.2 Treatment by Center Interaction Analysis (multi-center study)

Treatment-by-center interaction will be tested using the Breslow-Day test for homogeneity of the odds ratios for the primary efficacy endpoint.

### 8.6.2 Analysis of Primary Efficacy Endpoint

**Primary Analysis:**

The proportion of subjects (responders) meeting the primary endpoint with a FWS score of 0 or 1 and an improvement ≥2 points in FWS score (at maximum frown) at week 4 visit relative to baseline, based on both the investigator and the subject in-clinic assessments will be analyzed using the CMH test (with stratification variable site; small sites with less than 3 placebo subjects will be combined) using a significance level ($\alpha$) of 0.025. The null hypothesis $H_0$ given below will be tested against the alternative hypothesis $H_1$:

$H_0$: $p_{BoNT/A\text{-}DP} \leq p_{Placebo}$

$H_1$: $p_{BoNT/A\text{-}DP} > p_{Placebo}$

The CMH test will be applied by the SAS procedure Proc Freq using the General Association Statistic. The one-sided p-value will be derived by halving the two-sided p-value produced by the SAS procedure. Superiority of BoNT/A-DP over placebo will only be concluded if, besides statistical significance (one-sided p-value ≤0.025), the proportion of responders in the BoNT/A-DP treatment group is higher than for the placebo treatment group.

The FAS population will serve as the primary analysis set.

Subjects with missing investigator or subject in-clinic assessments with the FWS at baseline or visit week 4 will be assigned as being non-responders.

**Additional Exploratory Analyses on the Primary Efficacy Endpoint Variable:**

Further analyses on the primary efficacy endpoint variable are exploratory. A one-sided Pearson $\chi^2$-test will be applied. Two-sided CI of 95% for the responder rates in the BoNT/A-DP and the placebo treatment groups will be calculated using Wilson scores. As a sensitivity analysis for the primary efficacy endpoint, a tipping point analysis will be performed. Each missing value will be assigned to either a response or non-response, so that all possible combinations of replacing one or more missing values within each treatment group will be analyzed. P-values will be calculated for each combination and graphically displayed using a scatter plot. These analyses will be performed using the FAS population.

The CIs will also be calculated for the additional analyses as listed in the following:

- The primary efficacy endpoint measure using the observed values only, i.e., missing investigator or the subject in-clinic assessments with the FWS at baseline or week 4 visit will be excluded from analysis but not assigned as being non-responders. This analysis will be performed for the FAS, MFAS and PP populations.
- The primary efficacy endpoint measure using the MFAS and PP populations.
- The primary efficacy endpoint measure applying the LOCF for week 4 visit. This analysis will be done for FAS, MFAS and PP populations.

### 8.6.3 Analysis of Secondary Efficacy Endpoints

The following order of tests will be defined for the analysis of the key secondary efficacy endpoint with multiplicity control as described in 8.6.1.1:

- Primary endpoint (composite endpoint)
- Key secondary endpoint 1

  1.1: The modified Skindex-16 (GL-QoL) Emotional domain
  1.2: The modified Skindex-16 (GL-QoL) Social Functioning domain
  1.3: The modified Skindex-16 (GL-QoL) Overall score
  1.4: The FACE-Q Appraisal of Lines Between Eyebrows scale
  1.5: The FACE-Q Age Appraisal VAS score

*Analysis of key secondary efficacy endpoint 1: The extent of change in psychological impact (emotional and social functioning, and concerns relating to glabellar lines) at week 4 after*

*the first treatment, in the BoNT/A-DP group in comparison with placebo, relative to baseline, as assessed by the modified Skindex-16 (GL-QoL) and the FACE-Q Appraisal of Lines Between Eyebrows and FACE-Q Age Appraisal VAS.*

The modified Skindex-16 (GL-QoL) and FACE-Q Appraisal of Lines Between Eyebrows scale and FACE-Q Age Appraisal VAS scales will be analyzed according to the foreseen analysis approach of these tools, as described below.

**Emotional and Social Functioning**

The modified Skindex-16 (GL-QoL) Emotional and Social Functioning domains and overall scores will be derived in accordance with the Skindex-16 manual, adapted for the GL-QoL. The scores will be standardized on a scale from 0 (no impact) to 100 (maximal impact); a scale score is the average of responses to items addressing a construct. Domain and overall scores will be calculated if at least two of three social and if at least two of four emotional domain items are present. Derivation details are included in Appendix B. The absolute change from baseline (measured on the screening visit) will be calculated for the domain and overall scores and statistically compared between the treatment groups using the Wilcoxon Rank Sum test or t-test, if appropriate. Estimates of 'mild', 'moderate', and 'severe' psychological impact, and of the minimal clinical important difference, will be used to aid interpretation of scores and will be considered exploratory.

The analysis will be conducted with respect to the hierarchical order of the key secondary efficacy endpoint; the analysis using the Emotional domain first, the Social Functioning domain second, and the overall score third. All analyses will be conducted on observed values only, i.e., missing domain and overall scores will be excluded from analysis; no imputation of missing domain and overall scores will be performed.

**Concerns Relating to Glabellar Lines**

The FACE-Q Appraisal of Lines Between Eyebrows scale and FACE-Q Age Appraisal VAS scores will be derived in accordance with the developers' instructions and missing data treated accordingly. The absolute change from baseline (measured on the screening visit) will be calculated for the FACE-Q Appraisal of Lines Between Eyebrows scale and for the FACE-Q Age Appraisal VAS score and statistically compared between the treatment groups using the Wilcoxon Rank Sum test. All analyses will be conducted on observed values only, i.e., missing scale scores will be excluded from analysis; no imputation of missing scores will be performed. This analysis will be conducted with respect to the hierarchical order of key secondary efficacy endpoints as well; the analysis of FACE-Q Appraisal of Lines Between Eyebrows scale first, followed by the analysis of FACE-Q Age Appraisal VAS scores.

**Exploratory Analyses**

Exploratory analyses of modified Skindex-16 (GL-QoL) domain and overall score and FACE-Q Appraisal of Lines Between Eyebrows scale and FACE-Q Age Appraisal VAS will also be conducted using the MFAS and PP populations. The analyses described in the two previous sections will be repeated for these populations.

In addition, shift tables tables will be produced displaying the Psychological Impact of Modified Skindex-16 (GLQoL) at baseline versus week 4. The FAS, MFAS and PP populations will be used.

### 8.6.4 Additional Secondary Analyses

*Analysis of additional secondary efficacy endpoint 1: The percentage of responders at maximum frown (as defined above for the primary efficacy endpoint) at week 12 (after the first treatment with BoNT/A-DP or placebo).*

The same analyses, inclusive of the additional exploratory analyses, as described for the primary endpoint will be conducted. However, a sensitivity analysis using a tipping point analysis will not be performed.

*Analysis of additional secondary efficacy endpoint 2: The percentage of responders at maximum frown (as defined above for the primary efficacy endpoint) at week 16 (after the first treatment).*

The analysis of the week 16 visit will be conducted on observed values only, i.e., missing investigator or the subject in-clinic assessments with the FWS at baseline or week 16 visit will be excluded from analysis but not assigned as being non-responders. Subjects who were re-treated before visit week 16 are considered to be non-responders.

In addition, the same analyses, inclusive of the additional exploratory analyses, as described for the primary efficacy endpoint will be conducted as exploratory analyses. However, a sensitivity analysis using a tipping point analysis will not be performed. Subjects who were re-treated before visit week 16 are considered to be non-responders in all these analyses.

*Analysis of additional secondary efficacy endpoint 3: The percentage of responders at maximum frown (as defined above for the primary efficacy endpoint) at week 20 or later (after the first treatment).*

The analysis of percentage of responders at week 20 (after the first treatment) will be conducted on observed values only, i.e., missing investigator or the subject in-clinic

assessments on FWS at baseline or visit week 20 will be excluded from the analysis but not assigned as being non-responders. Subjects who were re-treated before visit week 20 are considered to be non-responders.

In addition, the same analyses, inclusive of the additional exploratory analyses as described for the primary efficacy endpoint, will be conducted as exploratory analyses (for select subgroups also). However, a sensitivity analysis using a tipping point analysis will not be performed.

*Analysis of additional secondary efficacy endpoint 4: The proportion of subjects with a ≥1 point reduction in FWS score at rest at week 4 in the first treatment cycle, based separately on the investigator and the subject in-clinic assessments*

In general, the same statistical analyses, inclusive of the additional exploratory analyses as described for the primary efficacy endpoint, will be conducted. However, a sensitivity analysis using a tipping point analysis will not be performed.

All analyses (inclusive the additional exploratory analyses) on the investigators' in-clinic assessment will be conducted only for subjects who have an investigators' in-clinic assessment of a FWS score at rest ≥1 at baseline. This will similarly be applied for the analyses of the subjects' in-clinic assessment.

For the primary analysis, subjects with missing investigator in-clinic assessments of FWS score at rest at baseline or week 4 visit will be assigned as being non-responders for this efficacy endpoint. This will similarly be applied for the subjects' in-clinic assessment within the primary analysis.

As a consequence, the primary analysis of this endpoint will lead to exactly the same results as the LOCF analysis (as a subject with missing week 4 is considered as a non-responder) and thus the LOCF analysis will be skipped.

Analyses of other visits after first treatment but before re-treatment up to week 48 are exploratory only based on the FAS population using descriptive statistics only. Subjects with re-treatment before the respective visit will be excluded from the analysis. Additional exploratory analyses, as described for the primary efficacy endpoint, will not be conducted, with the exception of the calculation of 95% CIs.

*Other additional secondary endpoint analyses:*

Other additional secondary efficacy endpoints will be analyzed applying the appropriate statistical method for the comparison of both treatment arms. For proportions

- of responders with a FWS score of 0 or 1 and an improvement ≥2 points in FWS score at maximum frown (composite endpoint, at weeks 1, 2 and 8),
- of subjects with ≥2 and ≥1 point reduction in FWS at maximum frown (at baseline and visits 2, 4, 12, 16 and 20 weeks after treatment, within the first treatment cycle),
- of subjects with a ≥1 point reduction in FWS at rest, relative to baseline
- of subjects with an improvement of ≥2 points in FWS at maximum frown at 4 weeks after re-treatment,
- of subjects with ≥1 point reduction in FWS score at maximum frown (at weeks 1, 2, 4, 8, 12, 16 and 20 relative to baseline,
- of subjects with ≥1 point reduction in FWS score at maximum frown at week 4 relative to re-treatment baseline,

the CMH test (with stratification variable site) and the Pearson $\chi^2$-test will be applied. In case of a total number of observations below 30, or in case of at least one cell frequency below 5, Fisher's exact test will be used instead of the $\chi^2$-test. Two-sided 95% CIs for response rates will be calculated, where appropriate. Additionally, the cumulative proportions of effects at weeks 1, 2, and 4 will be calculated.

Time to onset of effect in the first treatment cycle will be analyzed descriptively and by using the Kaplan-Meier method. Moreover, the daily assessments of the line severity when frowning will be analyzed descriptively. The FACE-Q satisfaction with outcome scale will be analyzed descriptively.

All efficacy endpoints, with the exception of "responder" 4 weeks after re-treatment, will be analyzed for the first treatment cycle only. The analysis of "responder" 4 weeks after re-treatment will be conducted on observed values only, i.e., subjects with missing investigator or the subject in-clinic assessments on FWS week 4 or at the preceding end of the cycle visit will be excluded from analysis. Subjects who were treated with placebo during the first treatment cycle are excluded from the analysis of this endpoint, but not assigned as being non-responders. A two-sided McNemar test (using a significance level of 0.05) will be used to compare the response proportions between week 4 re-treatment visit and at the preceding end of the cycle visit. These endpoints will be separately analyzed for each open label extension cycle.

In addition to the FAS population, the analyses of these endpoints will also be based on the MFAS and PP populations. For all analyses, a one-sided significance level of 0.025 will be used, if not stated otherwise. Additional exploratory subgroup analyses will be conducted as described in Section 8.8.

## 8.7 Safety Analyses

This section describes the safety analyses that will be conducted on the double blind and the open label treatment period, i.e., the safety analyses on all data collected during these treatment periods and all data collected in subjects who dropped-out during one of the treatment periods.

All definitions relative to safety endpoints are detailed in the following sections.

Safety analyses will be conducted on the SAF population and will be performed for all safety variables specified below.

All safety data will be summarized by treatment group.

The safety analyses of changes from baseline to a specific time point in safety variables (e.g., laboratory parameters, vital signs, and ECG) will only include subjects from the SAF population who have data available for both the baseline and the time point under consideration unless otherwise specified.

No statistical testing methods will be applied to statistically evaluate the differences on safety variables between treatment groups. Safety endpoint variables will be analyzed descriptively only.

An additional safety evaluation will be performed for laboratory data, vital signs, and ECG by defining cycle 2 day 0 as baseline (or last available value before drug administration of Cycle 2) for all subjects.

### 8.7.1 Treatment Exposure

A summary table displaying the number of subjects by the number of treatment cycles will be prepared. The duration of study drug exposure (calculated as described in Section 8.5.1) during each treatment cycle will be analyzed by summary statistics.

### 8.7.2 Adverse Events

All AEs will be classified by SOC and PT according to MedDRA Version 22.0 or higher.

In summaries by SOC and PT, adverse events will be sorted by decreasing frequency within each SOC and PT. In summaries by PT, AEs will be sorted by decreasing frequency within each PT.

AE summary tables will be presented for TEAEs and AESIs separately by treatment cycle and will include the number and percentage of subjects with any:

- TEAE/AESI

- TEAE/AESI related to study medication (AE will be defined as related if causality is definitely, probably or possibly or if causality assessment is missing)
- TEAE/AESI related to injection procedure (AE will be defined as related if causality is definitely, probably or possibly or if causality assessment is missing)
- Severe TEAE/AESI
- Severe TEAE/AESI related to study medication
- Severe TEAE/AESI related to injection procedure
- TEAE/AESI leading to discontinuation
- Study medication related TEAE/AESI leading to discontinuation
- Serious TEAE/AESI
- TEAE/AESI leading to death.

All TEAEs as well as all AESIs will be summarized by SOC, PT and treatment group using frequency counts and percentages (i.e., number and percentage of subjects with an event).

The number of events, as well as the number and rate of affected subjects will be reported by SOC and PT for all TEAEs and for all AESIs separately.

Adverse events will be separated to pre-treatment AEs and TEAEs. TEAEs are defined as all AEs with onset or worsening (increase in severity) after receiving first dose of study medication (independent of whether it is BoNT/A-DP or placebo). If it cannot be determined whether an AE is treatment-emergent due to missing onset cycle or a partial onset date, then it will be counted as treatment-emergent.

TEAEs and AESIs will be analyzed overall (for BoNT/A-DP or placebo), and for BoNT/A-DP treatment group additionally separately per first treatment cycle or open label extension phase: events starting on or after first treatment up to before open label extension (first cycle), and events starting on or after re-treatment until end of study (open label phase). If it cannot be determined whether an AE started during the first treatment cycle or the open label extension phase due to missing onset cycle and a partial onset date, then it will be counted as having started during the first treatment cycle. The cycle of onset as recorded in the eCRFs will be used for reported AEs by cycle where appropriate.

Events of subjects who started the first treatment with placebo but then entered the open label extension for re-treatment with BoNT/A-DP will be considered to placebo for the first cycle and to BoNT/A-DP for the open label phase.

TEAEs and AESIs (per SOC and PT) will also be summarized by seriousness, severity, relationship to study medication, and relationship to procedure using frequency counts and percentages (i.e., number and percentage of subjects with an event).

### 8.7.3 Antibody Formation

Serum samples will be tested for the presence of antibodies to botulinum toxin using an Anti Drug Antibody (ADA) assay; initially only the last sample obtained from each study subject will be analyzed together with the pre-dose baseline sample. If a sample tests positive for anti-drug antibodies, the other samples from the subject will be analyzed to determine when antibodies developed. Serum samples which test positive for binding antibodies will subsequently be tested for neutralizing activity. Blood serum for ADA tests will be collected at screening, at the visit 4 weeks after treatment and at the end of cycle visit for each treatment cycle.

The presence of antibodies will be summarized by counts and percentages of subjects by treatment group and time point. For the subgroup of subjects with a sample test positive for anti-drug antibodies, all analyzed samples will be analyzed descriptively by counts over time as well.

### 8.7.4 Laboratory Data

For the purposes of summarization in both the tables and listings, all laboratory values will be presented in SI units. If a lab value is reported using a nonnumeric qualifier e.g., less than (<) a certain value, or greater than (>) a certain value, the given numeric value will be used in the summary statistics, ignoring the nonnumeric qualifier.

Visit value and change and/or percent change from baseline (where applicable) during the treatment period will be summarized by treatment group using descriptive statistics for all laboratory parameters.

All laboratory values will be classified as normal or abnormal according to the central laboratory normal ranges and indicated as clinically significant (CS) or not clinically significant (NCS) by the investigator. Categorical variables will be summarized by counts and percentages of subjects in corresponding categories. Shift tables illustrating changes with respect to the central laboratory normal ranges and the investigator assessment between baseline and post-baseline visits will be created: number and percentage of subjects with normal, CS and NCS laboratory values.

### 8.7.5 Vital Signs

The analyses of variables for vital signs will focus on the evaluation of the change from baseline to the scheduled time points after baseline. Descriptive statistics by visit and of changes from baseline to each post-baseline visit by treatment cycle will be presented.

### 8.7.6 Electrocardiograms

Normal/abnormal shift tables illustrating changes between baseline and post-baseline visits will be created for the overall ECG interpretation by treatment group.

### 8.7.7 Physical Examinations

All physical examination data and abnormalities will be listed only.

## 8.8 Other Analysis

### 8.8.1 Subgroups

The following subgroup analyses will be performed for the primary efficacy endpoint. Subgroup analyses a) and b) will also be conducted for the key secondary efficacy endpoint and for the additional secondary efficacy endpoints 'the percentage of responders at week 12 (after the first treatment)', the percentage of responders at week 16 (after the first treatment)', 'the percentage of responders at week 20 (after the first treatment)' and 'the proportion of subjects with a ≥1 point reduction in FWS score at rest at week 4 in the first treatment cycle, based separately on the investigators' and the subjects' in-clinic assessments (applicable only for subjects who have a FWS score at rest ≥1 at baseline)':

a) Subgroup analysis by site and country.

Subgroup analyses will be conducted by site and country. Sites with less than three placebo subjects will be combined. The $\chi^2$-test will be applied for each subgroup. In case of a total number of observations below 30, or in case of at least one cell frequency below 5, Fisher's exact test will be used instead of the $\chi^2$-test. Stratified analyses applying the CMH test will also be conducted using factor country. 95% CIs will be presented, if applicable.

b) Subgroup analysis by subjects with previous use of botulinum toxin versus naïve subjects.

Subgroup analyses will be conducted for subjects with treatment with any serotype of botulinum toxin for any indication versus naïve subjects. The $\chi^2$-test or Fisher's exact test (see above corresponding criteria to use Fisher) will be applied for each subgroup. No stratified analysis using factor pre-treated/naïve subjects is currently planned. 95% CIs will be presented, if applicable.

c) Subgroup analysis by race (American Indian or Alaskan Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, Caucasian (Including Hispanic or Latino and Other)). The $\chi^2$-test or Fisher's exact test (see above corresponding criteria to use Fisher) will be applied for each subgroup. No stratified analysis using factor race is currently planned. 95% CIs will be presented, if applicable.

d) Subgroup analysis of:

- subjects with previous use of botulinum toxin by site
- naïve subjects by site

The $\chi^2$-test or Fisher's exact test (see above corresponding criteria to use Fisher) will be applied for each subgroup. No stratified analysis using site is currently planned. 95% CIs will be presented, if applicable.

e) Subgroup analysis by Fitzpatrick skin type (TYPE I, TYPE II, TYPE III, TYPE IV, TYPE V, and TYPE VI). The $\chi^2$-test or Fisher's exact test (see above corresponding criteria to use Fisher) will be applied for each subgroup. No stratified analysis using factor Fitzpatrick skin type is currently planned. 95% CIs will be presented, if applicable.

f) Subgroup analysis by sex. The $\chi^2$-test or Fisher's exact test (see above corresponding criteria to use Fisher) will be applied for each subgroup. No stratified analysis using factor sex is currently planned. 95% CIs will be presented, if applicable.

g) Subgroup analysis by ethnicity (Hispanic or Latino versus Not Hispanic or Not Latino). The $\chi^2$-test or Fisher's exact test (see above corresponding criteria to use Fisher) will be applied for each subgroup. No stratified analysis using factor ethnicity is currently planned. 95% CIs will be presented, if applicable.

h) Subgroup analysis by age groups (below 65 years, 65-74 years and 75-84 years). The $\chi^2$-test or Fisher's exact test (see above corresponding criteria to use Fisher) will be applied for each subgroup. No stratified analysis using factor age group is currently planned. 95% CIs will be presented, if applicable.

i) Subgroup analysis by additional age groups (below 65 years versus 65 years and older). The $\chi^2$-test or Fisher's exact test (see above corresponding criteria to use Fisher) will be applied for each subgroup. No stratified analysis using factor additional age group is currently planned. 95% CIs will be presented, if applicable.

Subgroups analyses will also be performed as follows for select additional secondary efficacy endpoints at select timepoints:

- The proportion of subjects with ≥1 point reduction in FWS score (at maximum frown) in the BoNT/A-DP and placebo groups during the first treatment cycle at week 4 relative to baseline, based on both the investigators' and the subjects' in-clinic assessments: all subgroups
- The proportion of subjects with ≥1 point reduction in FWS score (at maximum frown) in the BoNT/A-DP and placebo groups during the first treatment cycle at week 12, 16 and 20 relative to baseline, based on both the investigators' and the subjects' in-clinic assessments: by site, by country and by subjects with previous use of botulinum toxin versus naïve subjects.

Subgroup analyses will be performed on the FAS, MFAS and PP populations.

**COVID-19 Pandemic:**

Due to COVID-19 pandemic, the secondary endpoints 'the percentage of subjects with an improvement of ≥2 points in FWS at maximum frown at 4 weeks after re-treatment' and 'the percentage of subjects with ≥1 point reduction in FWS score at maximum frown at week 4 relative to re-treatment baseline" will be also analyzed using observed values by the FWS investigator assessment type (i.e. assessment done remotely or physically). The objective is to assess if the remote way has not impacted the investigator assessment.

These analyses will performed for the final analysis.

**Exclusion of Site 110 Subjects:**

A site audit conducted in January 2019 as pointed out a critical finding on a protocol non-compliance: an unblinded site staff performed blinded tasks on safety assessments. In order to assess the impact on the results, analyses based on FAS for primary and secondary efficacy endpoints relative to the double-blind period of the trial will be repeated excluding all the subjects of the site 110.

AE summary tables for TEAEs and AESIs, and TEAEs and AESIs summarized by SOC, PT will be also repeated for each site and by excluding all the subjects of the site 110.

These analyses will performed for the interim and final analyses.

ADA, laboratory, vital signs and ECG analyses will be also repeated by excluding all the subjects of the site 110 only for the final analysis.

Further analyses may be performed to assess the site 110 effect, if needed.

## 8.9 Interim Analysis

An IA will be performed after all subjects complete the re-evaluation for retreatment visit at week 16 of the first treatment cycle or complete the double blind phase (whichever occurs earlier).

A database closure for all data belonging to the double blind phase up to and including the C1 re-evaluation for retreatment visit at week 16 will be performed prior to this IA. All data belonging to the double blind phase up to and including the C1 re-evaluation for retreatment visit at week 16 will be checked and all queries resolved before database closure. A blinded data review will be conducted prior to unblinding based on all data of the double blind phase up to and including the C1 re-evaluation for retreatment visit at week 16 to check for protocol deviations and to allocate the subjects to the analysis sets. After database closure, data unblinding and analysis for the double blind phase up to and including the C1 re-evaluation for retreatment visit at week 16 will be performed.

Outputs corresponding to disposition, demography, baseline characteristics, AEs and AESIs along with the primary efficacy endpoint (Section 8.6.2), the key secondary efficacy endpoint (Section 8.6.3) and several additional secondary efficacy endpoints will be generated for the IA.

**Analysis of Primary and Key Secondary Efficacy Endpoints:**

Since the primary and the key secondary efficacy endpoints belong to the first 16 weeks of the double blind phase, the final analysis of these endpoints will be conducted during this IA (which is based on the final data of the double blind phase up to and including the C1 re-evaluation for retreatment visit at week 16) and no alpha adjustment needs to be done.

Sensitivity analysis for the primary efficacy endpoint:

- a tipping point analysis will be performed
- treatment-by-center interaction will be tested using the Breslow-Day test for homogeneity of the odds ratios

The above mentioned sensitivity analyses will be performed for the FAS, MFAS and PP populations.

The below listed additional analyses will also be performed for the primary efficacy endpoint:

- Using the observed values only, i.e., missing investigator or the subject in-clinic assessments with the FWS at baseline or week 4 visit will be excluded from analysis but not assigned as being non-responders. This analysis will be performed for the FAS, MFAS and PP populations.
- Applying the LOCF for the week 4 visit. This analysis will be performed for the FAS, MFAS and PP populations.
- Subgroup analyses as described in Section 8.8.1.

The above mentioned analyses will not be repeated for the final analysis.

**Analysis of Additional Secondary Efficacy Endpoints:**

The percentage of responders at week 12 (after the first treatment):

- Sensitivity analysis: treatment-by-center interaction will be tested using the Breslow-Day test for homogeneity of the odds ratios. This will be performed for the FAS, MFAS and PP populations.
- Using the observed values only, i.e., missing investigator or the subject in-clinic assessments with FWS at baseline or week 12 visit will be excluded from analysis but not assigned as being non-responders. This analysis will be performed for the FAS, MFAS and PP populations.
- Applying the LOCF for the week 12 visit. This analysis will be performed for the FAS, MFAS and PP populations.
- Subgroup analyses as described in Section 8.8.1

The above mentioned analyses will not be repeated for the final analysis.

The percentage of responders at week 16 (after the first treatment):

- Sensitivity analysis: treatment-by-center interaction will be tested using the Breslow-Day test for homogeneity of the odds ratios. This will be performed for the FAS, MFAS and PP populations.
- Using the observed values only, i.e., missing investigator or the subject in-clinic assessments with FWS at baseline or week 16 visit will be excluded from analysis but not assigned as being non-responders. This analysis will be performed for the FAS, MFAS and PP populations.
- Subgroup analyses as described in Section 8.8.1

The above mentioned analyses will not be repeated for the final analysis.

The proportion of subjects with a ≥1 point reduction in FWS score at rest at week 4 in the first treatment cycle, based separately on the investigators' and the subjects' in-clinic assessments (applicable only for subjects who have a FWS score at rest ≥1 at baseline):

- Sensitivity analysis: treatment-by-center interaction will be tested using the Breslow-Day test for homogeneity of the odds ratios. This will be performed for the FAS, MFAS and PP populations.
- Using the observed values only, i.e., missing investigator or the subject in-clinic assessments with FWS at baseline or week 4 visit will be excluded from analysis but not assigned as being non-responders. This analysis will be performed for the FAS, MFAS and PP populations.
- Subgroup analyses as described on Section 8.8.1.

The above mentioned analyses will not be repeated for the final analysis.

Other additional secondary efficacy endpoints will be analyzed applying the appropriate statistical method for the comparison of both treatment arms measured at week 16 or before.

- For the proportion of responders among BoNT/A-DP and placebo groups with a FWS score of 0 or 1 and an improvement ≥2 points in FWS score at maximum frown during the first treatment cycle visit relative to baseline, based on both the investigators' and the subjects' in-clinic assessments (composite endpoint, at weeks 1, 2 and 8), the CMH test (with stratification variable site) and the Pearson $\chi^2$-test will be applied as described in Section 8.6.4. Analyses will not be repeated for the final analysis.
- For the proportion of subjects with ≥1 point reduction in FWS score (at maximum frown) at weeks 1, 2, 4, 8, 12 and 16 relative to baseline, the CMH test (with stratification variable site) and the Pearson $\chi^2$-test will be applied as described in Section 8.6.4. Analyses will not be repeated for the final analysis however an analysis at week 20 will be performed for the final analysis.

Additionally, the cumulative proportions of effects at weeks 1, 2, and 4 will be calculated.

In addition, the below efficacy endpoint

- time to onset of effect as measured at weeks 1, 2 and 4 based separately on subject and investigator assessment. Onset of effect defined as at least a 1 point improvement in FWS score from baseline (at max. frown). In addition, onset of effect will be assessed by subjects daily in the first 2 weeks after treatment, by recordings in the subject diary;

will be analyzed descriptively and by using the Kaplan-Meier method. Analyses will not be repeated for the final analysis.

The additional secondary efficacy endpoint "extent of subject perceptions of effect of, and satisfaction with, treatment as assessed by the FACE-Q Satisfaction with Outcome Scale" will be analyzed descriptively. Analysis will be repeated for the final analysis to include all treatment cycles.

All the above tables will be supported by associated listings.

**Analysis of Safety Endpoints:**

The following tables will be generated for IA:

- Overall summary of TEAEs
- Overall summary of AESIs
- Summary of TEAEs by SOC and PT
- Summary of AESIs by SOC and PT
- Summary of TEAEs by seriousness by SOC and PT
- Summary of AESIs by seriousness by SOC and PT
- Summary of TEAEs by severity by SOC and PT
- Summary of AESIs by severity by SOC and PT
- Summary of TEAEs by relationship to study medication by SOC and PT
- Summary of AESIs by relationship to study medication by SOC and PT
- Summary of TEAEs by relationship to procedure by SOC and PT
- Summary of AESIs by relationship to procedure by SOC and PT
- Summary of related TEAEs (to study medication) by SOC and PT
- Summary of related TEAEs to procedure by SOC and PT

Tables and listings will be reproduced for the final analysis as well with the double-blind and open label phases included. All the above tables will be supported by associated listings.

In addition, the following listings will be produced for the IA:

- Inclusion/exclusion criteria at randomization; this listing will not be repeated for the final analysis.
- Medical history; these listings will not be repeated for the final analysis.
- Prior and concomitant medications; this listing will be repeated for the final analysis.

# 9 CHANGES TO PLANNED ANALYSIS FROM STUDY PROTOCOL

For the following additional secondary efficacy endpoints based on the independent rater's assessment of photographs, a worst case scenario as well as using the median value will be considered in the analyses:

- The proportion of subjects with ≥2 point and ≥1 point reduction in FWS score (at maximum frown) in the BoNT/A-DP and placebo groups during the first treatment cycle visit relative to baseline, based on the independent rater's assessment of photographs (at baseline and visits 2, 4, 12, 16 and 20 weeks after treatment, within the first treatment cycle).
- The proportion of subjects with a ≥1 point reduction in FWS score at rest in the BoNT/A-DP and placebo groups, relative to baseline, during the first treatment cycle, based on the independent rater's assessment of photographs.

The proportion of subjects with ≥1 point reduction in FWS score (at maximum frown) in the BoNT/A-DP and placebo groups during the first treatment cycle after week 20 relative to baseline, based on both the investigators' and the subjects' in-clinic assessments, will be assessed for the FAS population only. Subgroup analyses are not planned for this endpoint.

Subgroup analyses by geographic region will not be performed as this study is only being conducted in 2 countries, i.e., Austria and the US.

# 10 REFERENCES

1. ICH Topic E3: Structure and Content of Clinical Study Reports (CPMP/ICH/137/95 - adopted December 1995).
2. ICH Topic E9: Statistical Principles for Clinical Trials (CPMP/ICH/363/96 – adopted March 1998).
3. FDA: Code of Federal Regulations (CFR), part 11 - adopted 15 September 2016.

# 11 APPENDICES

## Appendix A - Visit Window

Analyses will be performed by nominal visit, irrespective of any time window deviations. For the End of Study (EOS) visit, the following criteria will be applied only for the efficacy dataset ADQS:

- The nominal visit (i.e. END OF STUDY) will be used throughout for all subjects who have not discontinued.
- For discontinued subjects, the EOS data will be mapped to the latest scheduled visit prior to EOS study day (based on ADY). If assessments are available at this visit, then data will be mapped to the visit following this preceding visit.
- Unless duplicate records are present in the database, there should be no duplicate data per visit.

The following examples illustrate the approach:

**Criteria (ii):**

| | Subjid | VISIT | VISITNUM | ADT | ADY |
|---|---|---|---|---|---|
| 1 | xxxx | SCREENING | 1 | 20MAY2019 | -11 |
| 2 | xxxx | BASELINE DAY 0 | 2 | 30MAY2019 | 1 |
| 3 | xxxx | CYCLE 1 WEEK 1 | 3 | 10JUN2019 | 11 |
| 4 | xxxx | CYCLE 1 WEEK 2 | 4 | 14JUN2019 | 15 |
| 5 | xxxx | END OF STUDY | 28 | 19JUL2016 | 50 |

For this subject, End of Study visit will be assigned to CYCLE 1 WEEK 4 visit.

**Criteria (iii):**

| | Subjid | VISIT | VISITNUM | ADT | ADY |
|---|---|---|---|---|---|
| 1 | xxxx | SCREENING | 1 | 28JUL2019 | -12 |
| 2 | xxxx | BASELINE DAY 0 | 2 | 09AUG2019 | 1 |
| 3 | xxxx | CYCLE 1 WEEK 1 | 3 | 18AUG2019 | 10 |
| 4 | xxxx | CYCLE 1 WEEK 2 | 4 | 25AUG2019 | 17 |
| 5 | xxxx | CYCLE 1 WEEK 4 | 5 | 07SEP2019 | 30 |
| 6 | xxxx | CYCLE 1 WEEK 8 | 6 | 08OCT2019 | 59 |
| 7 | xxxx | CYCLE 1 WEEK 12 | 7 | 01NOV2019 | 85 |
| 8 | xxxx | END OF STUDY | 28 | 12JAN2020 | 158 |

For this subject, End of Study visit will be assigned to CYCLE 1 WEEK 20 visit (i.e., End of study visit study day is compared versus the last possible visit study day [Study day of Cycle 1 Week 20 = 141] before End of study).

**Criteria (iii):**

| | Subjid | VISIT | VISITNUM | ADT | ADY |
|---|---|---|---|---|---|
| 1 | xxxx | SCREENING | 1 | 03AUG2019 | -6 |
| 2 | xxxx | BASELINE DAY 0 | 2 | 09AUG2019 | 1 |
| 3 | xxxx | CYCLE 1 WEEK 1 | 3 | 16AUG2019 | 8 |
| 4 | xxxx | CYCLE 1 WEEK 2 | 4 | 23AUG2019 | 15 |
| 5 | xxxx | CYCLE 1 WEEK 4 | 5 | 06SEP2019 | 29 |
| 6 | xxxx | CYCLE 1 WEEK 8 | 6 | 14OCT2019 | 67 |
| 7 | xxxx | CYCLE 1 WEEK 12 | 7 | 01NOV2019 | 85 |
| 8 | xxxx | END OF STUDY | 28 | 14DEC2019 | 128 |

For this subject, End of Study visit will be assigned to CYCLE 1 WEEK 16 visit (i.e., End of study visit study day is compared versus the last possible visit study day [Study day of Cycle 1 Week 16 = 113] before End of study).

Note: The PP population will be based on the CRF visits i.e. if EOS visit is allocated to CYCLE 1 WEEK 4, the subject will still be excluded from PP population for having CYCLE 1 WEEK 4 missing.

### Numbering of Visits

Visit numbers correspond with specific treatment and assessments as outlined in the protocol Schedule of Events, Section 2.1. If subjects are not eligible for re-treatment at the "Evaluation for Re-treatment" visit, they will return at 4-weekly intervals thereafter, until they are eligible for treatment. The subsequent visit number for the cycle will remain the same, but will be amended with "a, b, c" etc., corresponding to the time extended (at 4-weekly intervals) in that treatment cycle. For example, the first evaluation for re-treatment is at visit 7, week 12; however for subjects who are not eligible at week 12, but at week 16, this visit will be denoted visit "7a", while "visit 7b" will denote week 20 in the first cycle and "visit 7c" will denote week 24 etc., with each additional letter corresponding to a time point 4 weeks later in that cycle. This approach will enable the visit numbers to be standardized, while allowing flexibility to assess the duration of efficacy in the first treatment cycle and in the open label extension.

This visit numbering will be done at SDTM mapping level.

## Appendix B – Modified Skindex-16 GL-QoL Derivations

The modified skindex-16 GL-QoL is comprised of 2 domains: Emotional and Social Functioning.

The Emotional domain is comprised of 4 items:

[redacted]

The Social Functioning domain is comprised of 3 items:

[redacted]

Each item can be scored as follows:

[redacted]

Derivation of Standardized Scores:

Step 1: a numeric score will be applied to each item response as follows:

| Response | Numeric Score |
|---|---|
| [redacted] | |

Step 3: derive overall standardized score and domain standardized scores as follows:

Overall standardized score: average of item rescaled standardized scores [this assumes at least 2 of the 4 items have responses in the Emotional domain and at least 2 of the 3 items have responses in the Social Functioning domain]

Emotional standardized score: average of emotional item rescaled standardized scores [this assumes at least 2 of the 4 items have responses in the Emotional domain]

Social Functioning standardized score: average of social functioning item rescaled standardized scores [this assumes at least 2 of the 3 items have responses in the Social Functioning domain]

Overall standardized scores and domain standardized scores will be derived to 1 decimal place.

*Example:*

[redacted]

Derivation of Total Score:

[redacted]

Derivation of Estimates of 'Mild', 'Moderate', and 'Severe' Psychological Impact:
The total score will be used to further derive the categories 'mild', 'moderate' and 'severe' psychological impact as follows:

[redacted]

## Appendix C – FACE-Q Appraisal of Lines Between Eyebrows and Age Appraisal VAS Derivations

[redacted]

Each item can be scored as follows:

[redacted]

Derivation of Transformed Score:

Step 1: a numeric score will be applied to each item response as follows:

| Response | Numeric Score |
|---|---|
| [redacted] | |

Step 2: if there are at least 4 items with responses, derive a total score as follows:

7 items with responses: sum the numeric scores of all 7 items [total score range 7 to 28, inclusive].
4, 5 or 6 items with responses: calculate the mean of the responses and use this as the numeric score for missing items and sum the numeric scores accordingly. Round the total score down to the nearest integer for a conservative approach. Higher scores reflect a better outcome.

*Example:*
*5 items with responses and 2 items with missing responses:*

| Item Number | Response | Numeric Score |
|---|---|---|
| [redacted] | | |

Step 3: use the conversion table below to convert the total score into a score from 0 (worst) to 100 (best):

[redacted]

Age Appraisal VAS:

The VAS ranges from -15 (I look 15 years younger) to +15 (I look 15 years older), inclusive as follows:

[redacted]