Statistical Analysis Plan

Title:	A prospective, double-blind, randomized, placebo-controlled, multicenter study to investigate the efficacy and safety of NT 201 doses in the treatment of platysma prominence
Merz Study Number:	M602011078 / NCT05773053
SAP for	Final Analysis
Sponsor:	Merz Aesthetics GmbH
Version:	1.0
Date:	27-MAR-2024
Author:	PPD Statistician, PPD PPD

CONFIDENTIAL AND PROPRIETARY

The contents of this document are confidential and proprietary of Merz Aesthetics GmbH.

Unauthorized use, disclosure or reproduction is strictly prohibited. This document or parts thereof may not be disclosed to parties not associated with the clinical investigation without the prior written consent of Merz Aesthetics GmbH.

CCI

SIGNATURE PAGE

- I confirm that this Statistical Analysis Plan accurately describes the planned statistical analyses to the best of my knowledge and was finalized before database close.

PPD

Date (dd-mmm-yyyy)	Signature
Date (dd-mmm-yyyy)	Signature
Date (dd-mmm-yyyy)	Signature

Table of Contents

SIGNATURE PAGE 2
List of Abbreviations 4
1 General and Technical Aspects 5
2 Clinical Trial Design and Objectives 5
2.1 Clinical Trial Design 5
2.2 Objectives 6
3 Determination of Sample Size 6
4 Analysis Sets 7
5 Endpoints and Main Estimand for Analysis 7
5.1 Efficacy Endpoints 7
5.1.1 Primary Efficacy Endpoint 7
5.1.2 Secondary Efficacy Endpoint 8
CCI .8
5.2 Safety Endpoints 9
5.2.1 Primary Safety Endpoint 9
5.2.2 Secondary Safety Endpoint 9
CCI .9
5.3 Other Endpoints 10
6 Statistical Analysis Methods 10
6.1 Efficacy Endpoints 10
6.1.1 Primary Efficacy Endpoint 11
6.1.2 Secondary Efficacy Endpoint 13
CCI .13
6.2 Safety Endpoints 15
6.2.1 Primary Safety Endpoint 16
6.2.2 Secondary Safety Endpoint 16
CCI 18
CCI 19
6.4 Special Statistical/Analytical Issues 21
6.4.1 Discontinuations and Missing Data 21
6.4.2 Interim Analyses 21
6.4.3 Multiple Comparisons/Multiplicity 22
6.4.4 Examination of Subgroups 22
6.4.5 Pooling of Sites 22
6.4.6 Intercurrent Events 22
CCI .22
8 References 22
9 Appendix 24
CCI 24
24
24
25
27

List of Abbreviations

ADaM	Analysis Data Model
AE	Adverse event
ANCOVA	Analysis of covariance
ATC	Anatomical Therapeutic Chemical classification system of the World Health Organization
BMI	Body mass index
BoNT	Botulinum-Neurotoxin
CI	Confidence interval
CRO	Contract research organization
EOS	End of study
CCI	[redacted]
IIT	Investigator-initiated trial
IP	Investigational product
ITT	Intention to treat set
LS	Least square
MAPS-D	Merz Aesthetic Platysma Scale – Dynamic
CCI	[redacted]
MedDRA	Medical Dictionary for Regulatory Activities
N	Number of non-missing observations
CCI	[redacted]
PT	Preferred term
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System®
SDTM	Study Data Tabulation Model
SES	Safety evaluation set
SOC	System organ class
TEAE	Treatment emergent adverse event
TEAESI	Treatment emergent adverse event of special interest
TESAE	Treatment emergent serious adverse event
TFLs	Table, Figures, and Listings
U	Unit
US	United States
WHODD	World Health Organization Drug Dictionary

1 General and Technical Aspects

The objective of this statistical analysis plan (SAP) is to specify the statistical analyses with appropriate detail and precision to serve as a guideline for statistical programming and creation of tables, figures, and listings for clinical trial protocol, M602011078, dated 13-APR-2023 (version 2.0).

All programs will be written using SAS version 9.4 or higher. A preferred font size of 9 points, minimum font size of 8 points with a unique font size for the whole document required will be used for the tables and figures in Section 14. For listings, a standard font size of 9 points will be used to produce the output in A4 format. Individual SAS programs will be written for all tables, figures, and listings. All outputs will be transferred into or directly created as PDF files. These PDF files will be generated as needed to populate the subsections of Section 14 and Section 16.2 for the clinical study report. Separate documents will be created for sections 14.1, 14.2,14.3 and sections 16.2.1 to 16.2.8. Each output file will include the corresponding table of contents, preceding the content of the file.

The Merz standard Table, Figures, and Listings (TFLs) for drugs, version 8.0, dated 06-Apr-2023 will be applied and adapted to trial specific requirements as laid down in the clinical trial protocol and any amendments. These mock TFLs will serve as study-specific output specifications for statistical programming.

Special attention will be paid to planning and performance of quality control measures. Risk scores based on assessments of complexity and impact of errors and quality control measures for statistical programming (including analysis datasets, TFLs, SDTMs, ADaMs as applicable) will be documented in the quality control plan for the creation of statistical output.

2 Clinical Trial Design and Objectives

2.1 Clinical Trial Design

This is a prospective, double-blind, randomized, placebo-controlled, parallel group, multicenter, phase 2 clinical study. A total of 68 subjects will be randomized at seven sites in the US. Subjects enrolled will be adults, CCI [redacted] at the time of screening, with a score from moderate (score 3) to severe (score 4) on the validated five-point Merz Aesthetic Platysma Scale-Dynamic (at maximum contraction; MAPS-D), as assessed live by the investigator and the subject.

An equal number of study subjects will be randomly assigned with a balance ratio (1:1:1:1) to one of four groups, and all randomized subjects will receive a single treatment cycle on the platysma (Table 1). Each subject will receive one intramuscular treatment CCI [redacted] of NT 201 per injection point (given as injections of equal amount CCI [redacted]) or an equivalent volume of placebo (given as injections of equal amount CCI [redacted]).

CCI [redacted]

Study subjects will have eight on-site study visits, including screening (V1), treatment (V2), and post-treatment visits at Week 1 (V3), Week 2 (V4), Week 4 (V5), Week 9 (V6), Week 13 (V7), and Week 17 (V8, end of study [EOS] visit). Subjects will receive one post-treatment phone call at 24 to 72 hours after treatment to evaluate safety. If a subject reports a safety concern during a phone call, an unscheduled visit will be scheduled if necessary to address safety concerns. The planned duration of participation for individual subjects will be 17 weeks plus the individual duration of screening (up to 14 days) for each subject.

2.2 Objectives

The primary objective is to assess the efficacy of different NT 201 doses for treatment of moderate to severe platysma prominence by demonstrating superiority of NT 201 versus placebo.

The secondary objective is to assess the safety of different NT 201 doses for treatment of moderate to severe platysma prominence.

3 Determination of Sample Size

A total sample size of 68 randomized subjects (i.e., 17 subjects per treatment group) was chosen in order to have at least 64 evaluable subjects at the primary endpoint visit (V4, Week 2) assuming a drop-out rate of maximal 5% within the first two weeks. The sample of 64 evaluable subjects will have a power of at least 90% under the following assumptions:

- Mean change NT 201 = -1.7 (assumption for NT 201 Groups 1, 2, and 3);
- Mean change placebo (assumption for Group 4) = -0.5;
- Standard deviation = 1;
- Two-sided, alpha = 0.05;
- Randomization ratio 1:1:1:1; and
- Power at least 90%

An estimated total number of 64 subjects will provide 90.7% power to show a statistically significant difference between CCI dose group (Group 3 with N = 16 subjects) and the placebo (Group 4 with N = 16 subjects) at an alpha level of 0.05 (two-sided t-test, Statistical Analysis System software® Version 9.4). Further dose groups CCI vs. placebo and CCI vs. placebo will be tested in a hierarchical manner and no alpha adjustment will be performed.

Prager and colleagues [1] conducted an investigator-initiated trial (IIT) in which 23 subjects were treated with CCI of NT 201 in total. The mean change from baseline in mean scores at maximum tension on the published 5-point scale for platysmal bands [2] to Week 2 was -1.696. As no data for higher doses is available, a mean change from baseline of approximately -1.7 was assumed for all NT 201 treatment groups (Groups 1, 2, and 3) in this sample size estimation even if it might be even stronger especially in Group 2 CCI and Group 3 CCI As there was no data available for placebo and for standard deviation in Prager's study the assumption for mean change in placebo group and for standard deviation are based on clinical experience in other indications.

The sample size has been determined for the primary objective of the phase 2 study which is assessing the efficacy of different NT 201 doses for treatment of moderate to severe platysma prominence by demonstrating superiority of NT 201 versus placebo. The sample size of 68 randomized subjects is also considered adequate for the secondary objective of the phase 2 study which is assessing the safety of different NT 201 doses for treatment of moderate to severe platysma prominence. Seventeen (17) subjects per dose group is considered a sufficient number

of subjects to have a first safety assessment of each dose without exposing a too large number of subjects to risk and discomfort of potential AEs.

4 Analysis Sets

The following analysis sets will be defined for the statistical analysis of this clinical trial:

Safety Evaluation Set (SES)

The SES is the subset of all subjects who were exposed to IP at least once. Subjects in the SES analyses will be analyzed as treated.

Intention to Treat Set (ITT)

The ITT is the subset of all randomized subjects. Subjects in ITT analyses will be analyzed as randomized.

Modified ITT

The modified ITT is the same subset as the ITT, but it will be analyzed as treated. It will be used as sensitivity analysis for primary and secondary efficacy endpoint only.

Target population of main primary efficacy estimands

The target population of main primary efficacy estimands is the subset of all randomized subjects who are CCI [redacted] at screening visit and who have a moderate to severe MAPS-D score at maximum contraction, as assessed live by the investigator and subject at baseline, and who fulfill the protocol inclusion/exclusion criteria. Subjects in target population of main primary efficacy estimands analyses will be analyzed as randomized.

5 Endpoints and Main Estimand for Analysis

5.1 Efficacy Endpoints

5.1.1 Primary Efficacy Endpoint

The primary efficacy endpoint is:

- Change from baseline on the MAPS-D score at maximum contraction, as assessed live by the investigator at V4 (Week 2).

Note that the MAPS-D investigator's assessment conducted at the screening visit will be used as the baseline value for MAPS-D scores.

The main primary estimand is defined as follows in CCI [redacted]

CCI [redacted]

CCI

[redacted]

5.1.2 Secondary Efficacy Endpoint

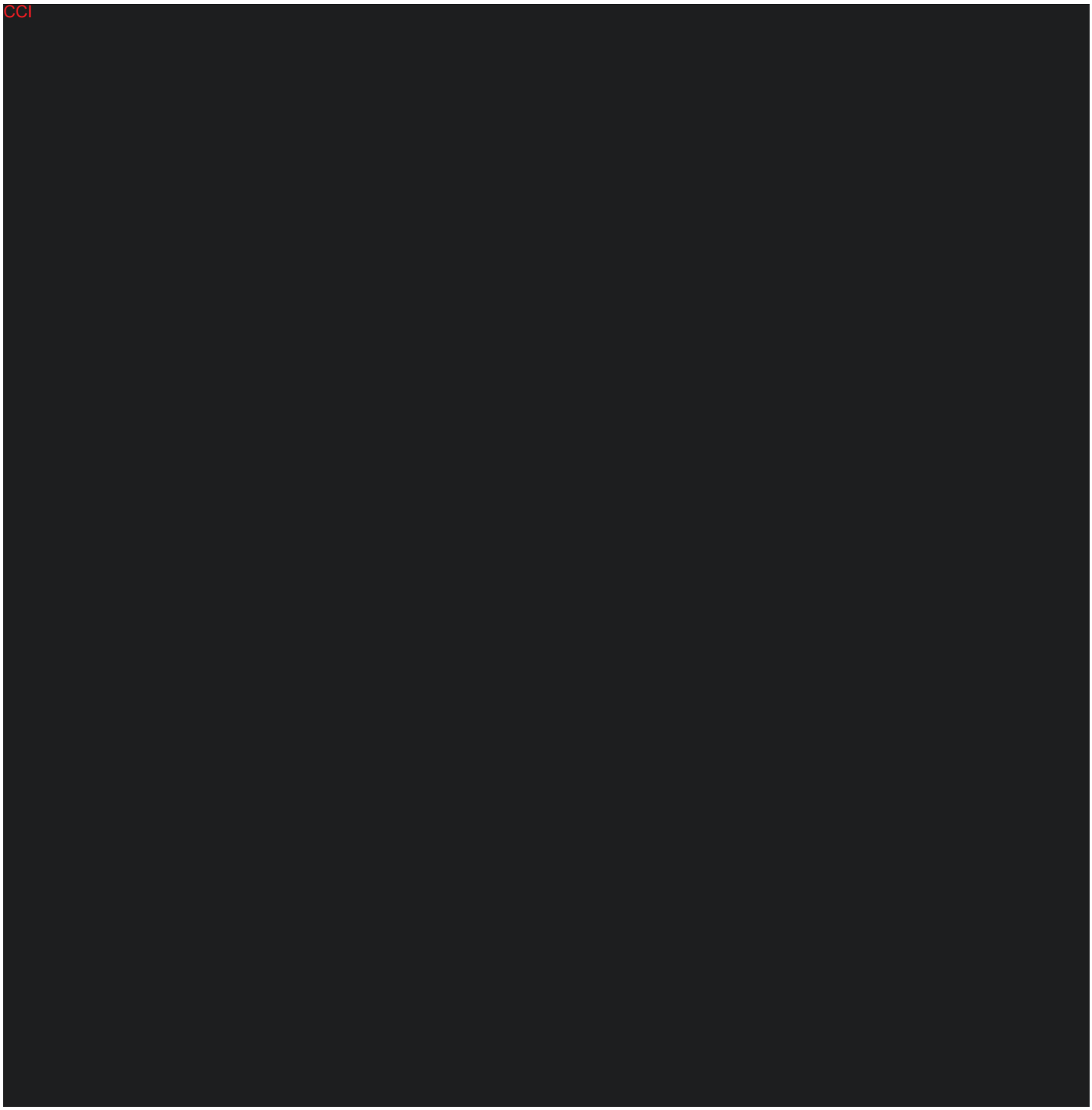
The secondary efficacy endpoint is:

- Change from baseline on the MAPS-D score at maximum contraction, as assessed by the subject at V4 (Week 2).

Note that the MAPS-D subject's assessment conducted at the screening visit will be used as the baseline value for MAPS-D scores.

CCI

[redacted]



5.2 Safety Endpoints

5.2.1 Primary Safety Endpoint

Not applicable.

5.2.2 Secondary Safety Endpoint

The secondary safety endpoint is:

- Incidence of related treatment emergent adverse events (related TEAEs).





6 Statistical Analysis Methods

The analysis for end-of study visit V8 will be based on data of subjects regularly completing the study at week 17 and of subjects prematurely discontinuing the study at any timepoint.

Metric descriptive statistics will comprise number of observed values (n obs), number of imputed values (n imp) (if applicable), mean, standard deviation, median, quartiles, minimum, and maximum. Mean, quartiles, and median will be reported to one decimal place more than the data were collected, for the standard deviation two decimal places more will be displayed; for derived data, the number of decimal places will explicitly be given in the sections below.

Frequency tables for qualitative variables will display absolute and percent frequencies (n, %). Percentages will be calculated using the denominator of all subjects in a specified analysis set within each treatment group. The denominator will be specified in a footnote to the tables for clarification if not otherwise obvious. Percentages will be reported to one decimal place.

If not otherwise specified, statistical tests will be conducted two-sided at type I error rate 5%. P-values will be reported to four decimal places (e.g., p=0.0375). P-values below 0.0001 will be presented as '<0.0001'. Confidence intervals will be two-sided 95%.

Treatment labels for summaries will be displayed as follows:

NT 201	NT 201	NT 201	Total NT 201 *Only in section 14.1 and 14.3*	Placebo	Total *Only in section 14.1*

Total NT 201 will include all subjects in NT 201 [redacted] NT 201 [redacted] and NT 201 [redacted] groups; total NT 201 group will be presented for all safety summaries.



6.1 Efficacy Endpoints

All efficacy analyses will be based on the ITT unless otherwise specified. Efficacy analyses will be performed per randomized treatment group. For primary and secondary efficacy endpoints only, a sensitivity analysis will be performed on the modified ITT per actual treatment group. Statistical tests will be two-sided hypothesis tests for between-treatment differences in general. Confidence limits and p-values will be given, where appropriate.

For all efficacy analysis, baseline is defined as the last non-missing value prior to the administration of IP (since MAPS scores are only taken at Screening and not at V2 that means that screening visit will be used as the baseline value for MAPS-D CCI).

Change from baseline at a visit/timepoint will be calculated using the following formula:

Change from baseline = result at visit/timepoint – baseline result.

In case baseline result is missing, change from baseline will not be calculated.

6.1.1 Primary Efficacy Endpoint

Change from baseline on the MAPS-D score at maximum contraction, as assessed live by the investigator at Week 2 will be used to test the difference of each NT 201 group to placebo.

6.1.1.1 Primary analysis

MAPS-D scores will be assigned to the following numeric values for analysis:

CCI



A baseline-adjusted ANCOVA will be applied to the target population of main primary efficacy estimands CCI with change from baseline to Week 2 on the MAPS-D score at maximum contraction (assessed by the investigator) as dependent variable. Randomized treatment group and study site will be used as fixed factors and MAPS-D at baseline and age at screening as covariates.

LS means of each treatment group along with 95% CI from the above model will be provided. Also, the LS means difference between each NT 201 group and placebo and their 95% CI estimates will be provided. The p-value from the two-sided test for the significance of the difference between each NT 201 group and placebo at $\alpha = 0.05$ will also be provided.

A similar SAS code to the following statements will be employed for this analysis:

CCI



The amount of missingness of the primary efficacy endpoint is expected to be low. In case the primary efficacy endpoint would have been missing for more than 10%, multiple imputation (MI) using the SAS procedure PROC MI with “1000” iterations and seed “602011078” would have been applied on the ITT. The VAR statement would have been considered the following variables:

randomized treatment group, study site, age at screening, and MAPS-D at baseline, Week 1 and Week 2. Due to the lack of information about the missing data pattern in MAPS-D values over time, the fully conditional specification option would have been chosen for the model. All missing MAPS-D values for the post baseline visits (Week 1 and Week 2) would have been replaced with the imputed values.



A higher value on the MAPS-D means higher severity, thus negative changes from baseline are favored. Therefore, superiority of NT 201 group over placebo will be demonstrated at type I error level $\alpha = 0.05$ if the LS mean difference between NT 201 group and placebo is less than zero and the corresponding p-value is less than 0.05.

For primary efficacy endpoint the three NT 201 dose groups will be tested against placebo in a hierarchical manner as described in Section 6.4.3 in order to avoid any inflation in the type 1 error (α) that might arise while considering multiple hypotheses simultaneously.

No pooling of sites is planned for primary analysis.

6.1.1.2 Sensitivity analysis

The primary analysis of the primary efficacy endpoint will be repeated for the ITT with the observed cases (that means without any imputations for the missing values) and for modified ITT (i.e. using actual instead of randomized treatment group) with observed cases. This will serve as sensitivity purposes.

6.1.1.3 Subgroup analyses

Subjects will be classified into subgroups based on the baseline MAPS-D score (assessed by investigator) as follows:

- Moderate group – Subjects with "Moderate" MAPS-D score at baseline
- Severe group – Subjects with "Severe" MAPS-D score at baseline

Metric descriptive statistics of the change from baseline to Week 2 on MAPS-D scores at maximum contraction, assessed by the investigator, will be provided by treatment group for target population of main primary efficacy estimands CCI for ITT and for modified ITT for each subgroup. No other subgroup analyses will be performed.

6.1.2 Secondary Efficacy Endpoint

Change from baseline on the MAPS-D score at maximum contraction, as assessed by the subject at Week 2 is the key secondary endpoint and will be used to test the difference of each NT 201 group to placebo. The MAPS-D scores will be assigned to numeric values as mentioned in section 6.1.1.1.

A baseline-adjusted ANCOVA will be applied to ITT with change from baseline to Week 2 on the MAPS-D score at maximum contraction (assessed by the subject) as dependent variable and randomized treatment group and study site as fixed factors and MAPS-D at baseline and age at screening as covariates.

LS means of each treatment group along with 95% CI from the above model will be provided. Also, the LS means difference between each NT 201 group and placebo and their 95% CI estimates will be provided. The p-value from the two-sided test for the significance of the difference between each NT 201 group and placebo at $\alpha = 0.05$ will also be provided. SAS codes similar to the ones provided in section 6.1.1.1 will be used for analysis.

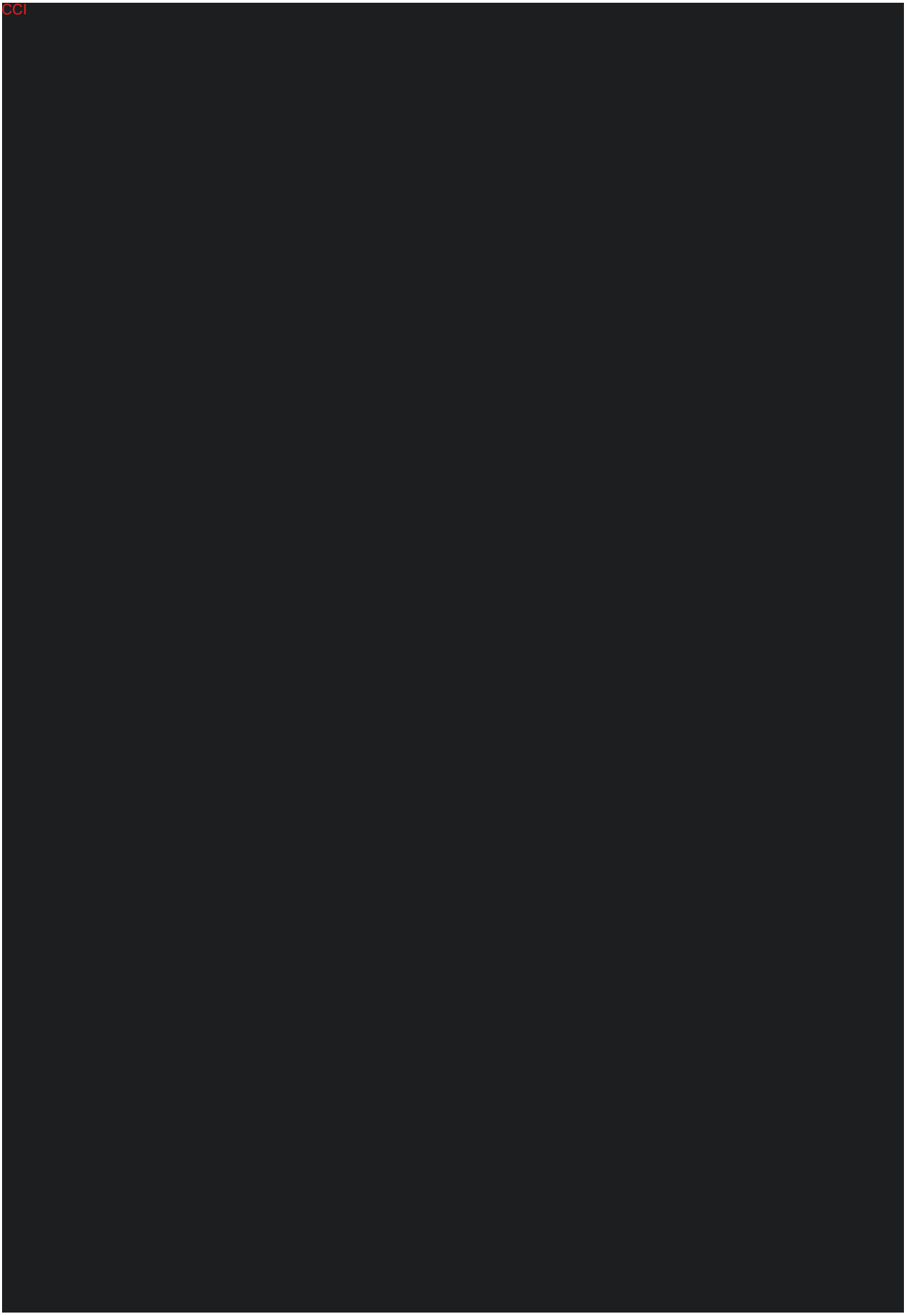
Missing values will not be imputed (i.e., all values as observed will be used for analysis).

The analysis of the secondary efficacy endpoint will be repeated for modified ITT (i.e. using actual instead of randomized treatment group) with observed cases. This will serve as sensitivity purposes.

CCI



CCI



CCI

[redacted]

6.2 Safety Endpoints

All safety analyses will be performed on the SES.

All safety tables will be presented by actual treatment group and total actual NT 201 group.

TEAEs are defined as AEs that had onset or worsened on or after date and time of first dose of study treatment. See Appendix 5: Separation of TEAEs from Non-TEAEs for detailed description of separation of TEAEs from Non-TEAEs.

Calculation of time to onset/duration of AEs (days):

- Time to onset of an AE is defined as start date of AE - date of first administration of study drug [+ 1 day for TEAEs]
- The duration will be calculated as stop date - onset date + 1 day.

In case of missing intensity or missing causal relationship of an AE the worst-case principle will be applied, i.e., a missing intensity will be set to "severe" and a missing causal relationship will be set to "related". Missing outcome of an AE will be set to "unknown".

If an AE worsens between start and end of this AE, it will be considered as new AE entry with a new record starting with the date of the worsening (start of new AE = worsening date, stop date of milder AE episode = date of AE worsening).

All AEs will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA) version in effect at the time the database is closed. Only TEAEs will be analyzed in most of the tables. Only in special table for public disclosure where specified all AEs, TEAEs and non-TEAEs will be analyzed.

For all safety analyses, baseline is defined as the last non-missing value prior to the administration of IP.

Change from baseline at a visit/timepoint will be calculated using the following formula:

Change from baseline = Result at visit/timepoint – Baseline result; In case baseline result is missing, change from baseline will not be calculated.

Absolute values and changes from baseline of laboratory evaluations and vital signs will be analyzed by metric statistics by visit.

For all summaries by visit, only the results from scheduled visits will be included. In case of repeat assessments, the results from those assessments will be discarded from analysis. Results from unscheduled visits and repeat assessments will only be listed.

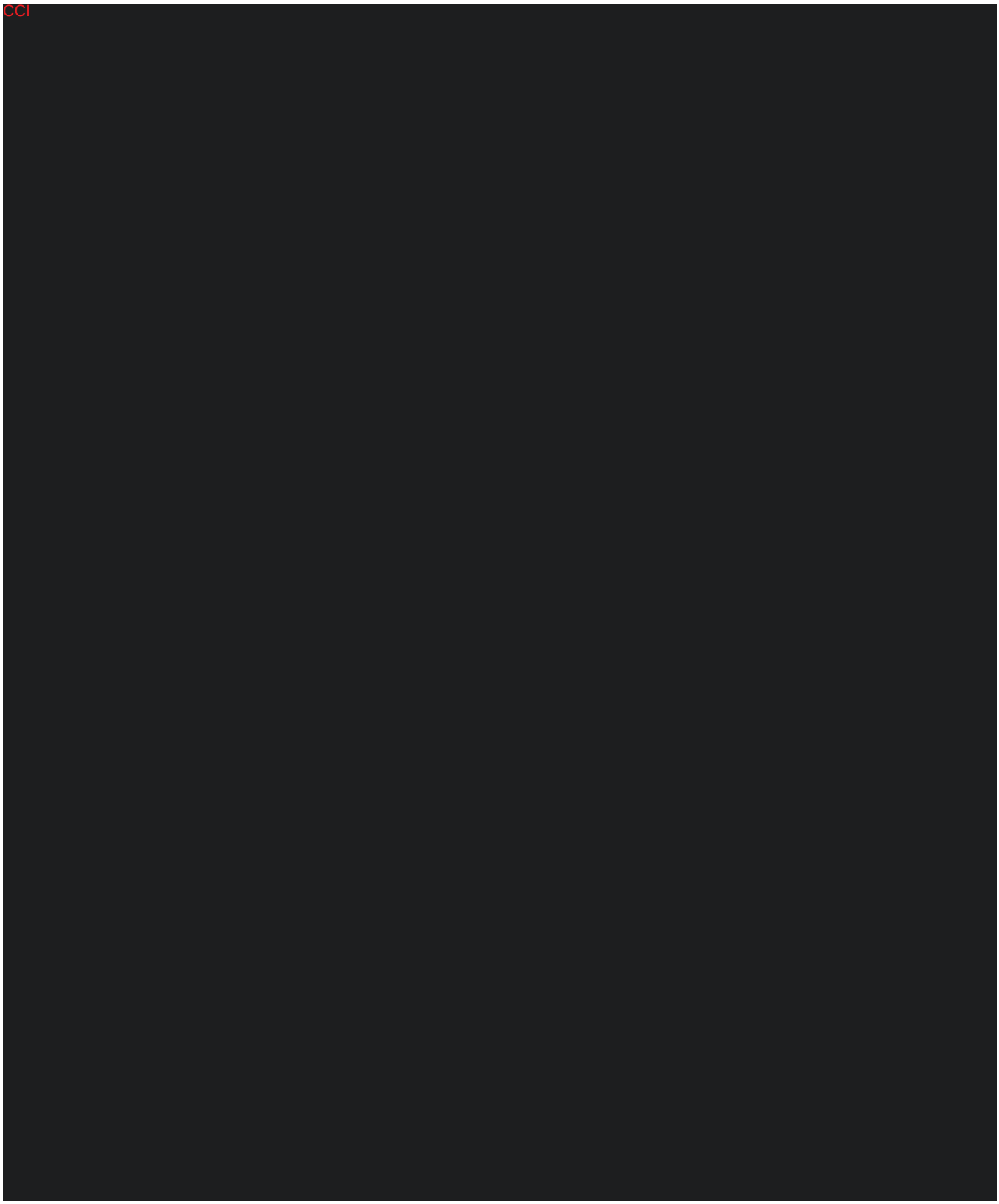
6.2.1 Primary Safety Endpoint

Not Applicable.

6.2.2 Secondary Safety Endpoint

Summary of number and percentage of subjects with related TEAEs along with number of related TEAEs will be provided. Furthermore, summaries will be provided for number and percentage of subjects with related TEAEs and number of related TEAEs on the MedDRA system organ class (SOC) level and on the preferred term (PT) level.

The AEs listed in CCI are defined as adverse event of special interest (AESIs) for this study.


CCI

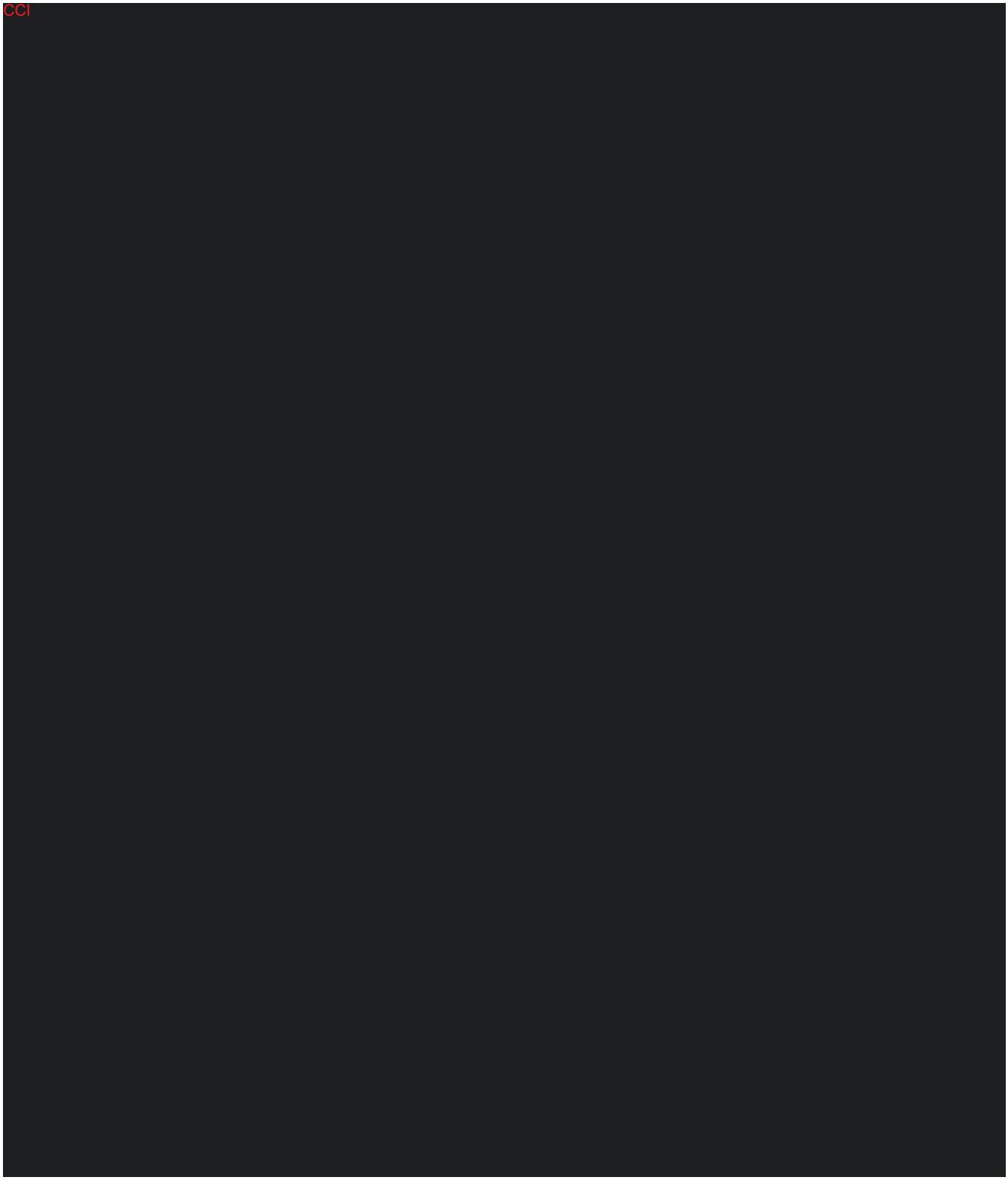
CCI

CCI

CCI

[redacted]

CCI



6.4 Special Statistical/Analytical Issues

6.4.1 Discontinuations and Missing Data

Discontinued subjects will not be replaced.

For missing imputation techniques of the primary efficacy endpoint refer to Section 6.1.1.1.

For imputation of missing causal relationship, intensity or outcome of an AE refer to Section 6.2.

6.4.2 Interim Analyses

Not applicable.

6.4.3 Multiple Comparisons/Multiplicity

The family-wise two-sided type I error level $\alpha = 0.05$ for comparisons of NT 201 dose groups and placebo on the primary efficacy endpoint will be controlled by a hierarchical test procedure based on two-sided tests with $\alpha = 0.05$.

For primary efficacy endpoint the three NT 201 dose groups will be tested against placebo in a hierarchical manner as follows:

- Only if NT 201 CCI vs. placebo is statistically significant, NT 201 CCI vs. placebo will be tested in a confirmatory way;
- If NT 201 CCI vs. placebo is also statistically significant, then NT 201 CCI vs. placebo will be tested in a confirmatory way; and
- If one of the tests is not statistically significant vs. placebo the subsequent test will be interpreted in an exploratory manner.

The hierarchical test procedure will stop once statistical significance could not be reached. All further tests will be of descriptive nature only.

6.4.4 Examination of Subgroups

Subgroup analyses for the primary efficacy endpoint are specified in Section 6.1.1.3.

6.4.5 Pooling of Sites

No pooling of sites is planned for analysis.

6.4.6 Intercurrent Events

ICEs according to ICH E9 (R1) [3] are prespecified for primary efficacy estimands (see section 5.1.1) and will be addressed by the handling strategies described in CCI.

All relevant ICE for primary efficacy estimand will be listed together with corresponding handling strategy.

CCI

[redacted]

8 References

[1] W Prager, EK Bee, I Havermann, I Zschocke. IncobotulinumtoxinA for the treatment of platysmal bands: a single-arm, prospective proof-of-concept clinical study. Dermatol Surg. 2015;41 Suppl 1:S88-92.

[2] TL Geister, B Bleßmann-Gurk, B Rzany, L Harrington, R Görtelmeyer,R Pooth. Validated assessment scale for platysmal bands. Dermatol Surg. 2013;39(8):1217-25.

[3] ICH E9 (R1): Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials, 20 November 2019, adopted by CHMP, 30 January 2020, issued as EMA/CHMP/ICH/436221/2017

9 Appendix

CCI

