Statistical Analysis Plan

Title:	Prospective, randomized, double-blind, placebo-controlled, parallel-group trial with an open-label period to investigate the efficacy and safety of NT 201 in the unilateral and bilateral treatment of essential tremor of the upper limb
Merz Clinical Study Number:	M602011069
SAP for	Final Analysis of Efficacy and Safety
Sponsor:	Merz Pharmaceuticals GmbH
Version:	1.0
Date:	24-JAN-2024
Author: (Name, Function, and CRO)	[redacted] Biostatistician, Metronomia

Document References (for internal use only)	
SOP Number	Q-00457
Template name	tem-sap-v8.0-20220223
Document filename	11.01.01_M602011069_sap_1.0_2024-01-24.pdf

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 breaking the blind.

[redacted]
Author
Biostatistician,
Metronomia Clinical Research GmbH

Signature | Date (dd-MMM-yyyy)

[redacted]
Document Owner (Approver)
Biostatistician,
Merz Pharmaceuticals GmbH

Signature | Date (dd-MMM-yyyy)

[redacted], MD
Approver
Medical Expert,
Merz Pharmaceuticals GmbH

Signature | Date (dd-MMM-yyyy)

Table of Contents

Version History

The original version of the statistical analysis plan (SAP) for clinical trial M602011069 is based on the clinical study protocol version 3.0 dated 31-MAY-2021 and amendment 1 to CSP version 3.0 dated 26-MAY-2023.

SAP Version	Date	Change	Rationale
1.0	24-JAN-2024	Not applicable	Original version

List of Abbreviations

Abbreviation/Term	Definition/explanation
ADaM	Analysis data model according to CDISC
ADL	Activities of daily living
AE	Adverse event
AESI	Adverse event of special interest
ANCOVA	Analysis of covariance
ATC	Anatomical Therapeutic Chemical classification system of the World Health Organization
BMI	Body mass index
CI	Confidence interval
COVID-19	Coronavirus Disease 2019
CRF	Case report form used synonymously to electronic CRF (eCRF)
CRO	Contract research organization
CSP	Clinical study protocol
CSR	Clinical study report
deg	Degrees of arc
ET	Essential tremor
EOS	End-of-study
[redacted]	
eCRF	See CRF
FAS-UP	Full analysis set for the unilateral treatment period
FAS-BP	Full analysis set for the bilateral treatment period
GICS	Global Impression of Change Scale(s)
HHD	Handheld Dynamometer
IP	Investigational product
IRP	Independent rater panel
LS-Means	Least Square Means
MAR	Missingness at random
MedDRA	Medical Dictionary for Regulatory Activities
[redacted]	
n	Number of values analyzed for metric variables / Absolute frequency for qualitative variables
NA	Not applicable
NoA	North American
nimp	Number of imputed values
nmiss	Number of missing observations
OM	Observed margins
OC	Observed cases
PPS	Per protocol set
PT	Preferred term
PD	Protocol deviation
QC	Quality control
QUEST	Quality of Life in Essential Tremor
QoL	Quality of life
RMS	Root Mean Square
SAP	Statistical analysis plan
SD	Standard deviation
SDTM	Study data tabulation model according to CDISC
SE	Standard error
SES-UP	Safety evaluation set for the unilateral treatment period
SES-BP	Safety evaluation set for the bilateral treatment period
SOC	System organ class
Study	(Clinical) study. Used synonymously to (Clinical) trial.
TEAE	Treatment emergent adverse event
TEAESI	TEAEs of special interest
TETRAS	The Essential Tremor Rating Assessment Scale
TFLs	Table, Figures, and Listings
Trial	(Clinical) trial. See “Study”

Abbreviation/Term	Definition/explanation
UL	Upper limb
V	Visit
VAS	Visual analogue scale

1 Introduction

Changes to the protocol-planned analyses are described in Section 6.

1.1 General and Technical Aspects

The objective of this statistical analysis plan (SAP) is to specify the statistical analyses of efficacy and safety endpoints with appropriate detail and precision to serve as a guideline for statistical programming and creation of tables, figures, and listings for clinical study protocol (CSP) M602011069 version 3.0 (dated 31-May-2021) and amendment 1 to CSP version 3.0 (dated 26-MAY-2023). [redacted]

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 of the clinical study report (CSR). 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 provided as rtf files and also transferred into PDF files. These PDF files will be generated as needed to populate the subsections of Section 14 and Section 16.2 of the CSR. Each output file will include the corresponding table of contents, preceding the content of the file.

The Merz template 'tem-specification-tfls, version [8.0]' will be applied and adapted to trial specific requirements as laid down in the clinical study protocol and any amendments. The resulting mock Table, Figures, and Listings (TFLs) will serve as output specifications for statistical programming. TFL specifications are given in a separate external document to this SAP.

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 will be documented in quality control (QC) plans and reports for the creation of Standard data tabulation model (SDTM), Analysis data model (ADaM) data sets and TFLs. QC plans for SDTMs and ADaMs will be common documents for analyses as described in this SAP [redacted]. For QC of TFLs for analyses as described in this SAP [redacted], separate QC plans will be generated. QC reports will be created for each draft and final analysis.

1.2 Study Design

This is a prospective, randomized, double-blind, placebo-controlled, parallel-group clinical trial with an open-label period that investigates the efficacy and safety of NT 201 in the unilateral and bilateral treatment of essential tremor (ET) of the upper limb (UL). A total of approximately 75 evaluable male and female adults will be recruited in this Phase 2 clinical trial.

The clinical trial consists of two treatment cycles. The first is a double-blind unilateral treatment period (Cycle 1) that starts with the screening visit (V1), which will be performed between 21 and 3 days prior to the study baseline visit (V2). At the study baseline visit (V2), on Day 1, eligible subjects will be randomized to one of two treatment groups, NT 201 or placebo, with a randomization ratio of 2:1. The unilateral intramuscular injections will be administered into the motor dominant UL. Subjects will be followed up during six subsequent visits until the end-of-Cycle 1 visit at Week 24 after injection.

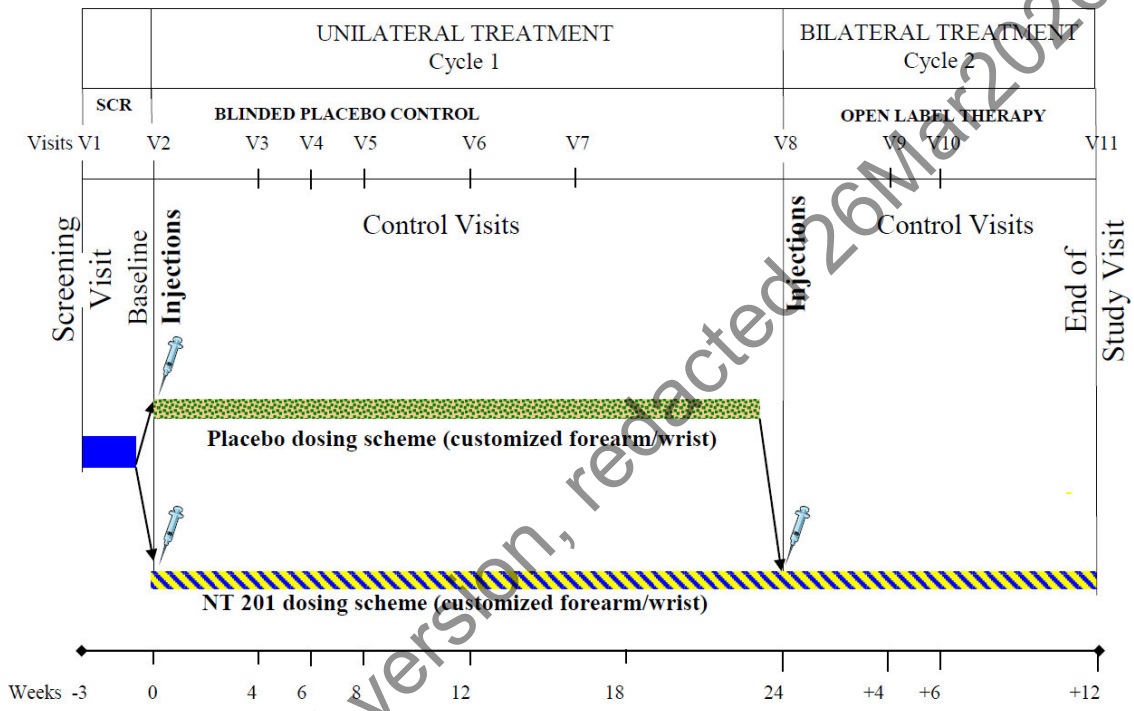
At the end of the unilateral treatment period (end-of-Cycle 1 visit), subjects will be checked for eligibility to participate in the subsequent open-label bilateral treatment period (Cycle 2). If

eligibility criteria for reinjection (see CSP Section 7.4) are met, the Cycle 2 baseline visit (Cycle 2, Day 1) will be performed on the same day, including bilateral UL treatment with NT 201. Subjects will be followed up during three subsequent visits until the end-of-study (EOS) visit (V11) at Week 12 after reinjection.

More details on study design and study treatments can be found in CSP Sections 6.1 and 8.2, respectively.

The study flow chart is shown in Figure 1.

Figure 1: Flow chart of the clinical trial assessing efficacy and safety of NT 201 in the unilateral and bilateral treatment of essential tremor of the upper limb



V = Visit; SCR = Screening.

In case of a public health emergency, e.g. due to a Coronavirus Disease 2019 (COVID-19) outbreak, in which the study or study visits for subjects in Cycle 1 were put on hold or in any way modified so that V8 could not be performed according to study plan (e.g. more than 25 weeks after Cycle 1 injection at baseline visit V2), subjects will not continue the study with Cycle 2 baseline visit V8 but will continue with an additional visit V8a when this is possible again. In this case, V8(a) will serve as Cycle 2 baseline.

2 Objectives, Endpoints, and Other Variables

2.1 Study objectives

- The primary objective of this study is to assess the efficacy of unilateral intramuscular injections of NT 201, as compared with placebo, in subjects with ET of the UL.
- Secondary objectives are the following:
    - Safety of unilateral intramuscular injections of NT 201, as compared with placebo, in subjects with ET of the UL.

- Efficacy and safety of bilateral intramuscular injections of NT 201 in subjects with ET of the UL.

2.2 Endpoints

2.2.1 Efficacy Endpoints

Efficacy endpoints specified in the subsequent subsections are based on the following clinical outcome assessments:

- Assessments of angular tremor amplitude (in degrees of arc [deg]) at the shoulder (flexion/extension, adduction/abduction), elbow (flexion/extension) and wrist (flexion/extension, radial deviation/ulnar deviation, pronation/supination) will be performed with standardized computerized kinematic analysis using the TremorTek® device (see CSP Section 9.1.1.2.1). By this means, status of UL tremor will be assessed in dependency of four different tasks. In three consecutive assessment rounds per subject and visit, Tasks I to IV will be repeated in the exact same order.
- TETRAS (Elble 2016 [2], Ondo 2018 [5]; see CSP Section 9.1.1.2.2), a validated assessment tool consisting of an activities of daily living (ADL) subscale with 12 items (to be completed by subjects, i.e. patient-reported outcome) and a Performance subscale with 9 items (to be assessed by qualified investigators, i.e. clinician-reported outcome). For an overview of TETRAS scores, defined as sum scores of specific ADL/Performance items, see Appendix 8.1. Derivation of TETRAS based efficacy endpoints is described in Section 4.1.
- Investigator's Global Impression of Change Scale (GICS) and Subject's GICS (see CSP Section 9.1.1.2.3), representing commonly used single item global outcome assessment scales to evaluate overall clinical impression of change after treatment.
- Quality of Life in Essential Tremor (QUEST; Troester 2005 [6], see CSP Section 9.1.1.2.4), a validated health-related quality of life (QoL) questionnaire comprising 30 items that inform on interference and impact of ET on five dimensions of QoL. Derivation of QUEST based efficacy endpoints, also including a total score, is described in Section 4.1.
- [redacted]

2.2.1.1 Primary Efficacy Endpoint

Change from study baseline to Week 6 (V4) in maximum tremor amplitude measurement (root mean square (RMS) deg] at wrist level during the unilateral treatment period.

The primary endpoint is based on angular tremor amplitude values recorded by the kinematic sensor for the motor-dominant UL during four tasks repeated in three assessments rounds at both study baseline (defined in Section 4) and Week 6 (V4). The task with the maximum tremor

[1] health state = pattern of scores over the five dimensions, e.g., 13333 for a subject indicating slight problems in mobility and severe problems on the other four dimensions.

amplitude at study baseline is selected. For the computation of the primary efficacy endpoint and further details on statistical analysis see Section 4.1.1.

2.2.1.2 Secondary Efficacy Endpoints

2.2.1.2.1 Key Secondary Efficacy Endpoints

- Change from study baseline to Week 6 (V4) in TETRAS Performance dominant UL score (Score no. 5 in Appendix 8.1) in the unilateral treatment period as assessed by the investigator
- Change from study baseline to Week 6 (V4) in TETRAS ADL UL score (Score no. 2 in Appendix 8.1) in the unilateral treatment period

For definition of study baseline see Section 4.

2.2.1.2.2 Other Secondary Efficacy Endpoints

Unilateral treatment period:

- Change from study baseline to Week 6 (V4) in TETRAS ADL Functional Impact score (Score no. 3 in Appendix 8.1)

  For definition of study baseline see Section 4.
- Subject´s GICS score of motor dominant UL at Week 6 (V4)
- Investigator´s GICS score of motor dominant UL at Week 6 (V4)

Bilateral treatment period:

- Change from Cycle 2 baseline to Week 6 (V10) in TETRAS Performance dominant UL score (Score no. 5 in Appendix 8.1) as assessed by the investigator[2].
- Change from Cycle 2 baseline to Week 6 (V10) in TETRAS Performance subscale score (Score no. 4 in Appendix 8.1) as assessed by the investigator[2].
- Change from Cycle 2 baseline to Week 6 (V10) in TETRAS ADL UL score (Score no. 2 in Appendix 8.1).
- Change from Cycle 2 baseline to Week 6 (V10) in TETRAS ADL Functional Impact score (Score no. 3 in Appendix 8.1).

For definition of Cycle 2 baseline see Section 4.

2.2.1.3 Other Efficacy Endpoints

Unilateral treatment period *(for comparisons of NT 201 versus placebo)*

- Time course of absolute values in the angular tremor amplitude and changes from study baseline of 1) wrist, 2) elbow, and 3) shoulder level of the treated motor dominant UL in the unilateral treatment period for the
  - Task with maximum amplitude at study baseline[3] [RMS deg]
  - Maximum amplitude from Tasks I to IV at each visit [RMS deg]

[2] For precision, "as assessed by the investigator" was added to definitions of other secondary efficacy endpoints based on TETRAS performance assessments in Cycle 2 as included in the CSP (see Section 6).
[3] Except change from study baseline to Week 6 at wrist level which is already defined as primary efficacy endpoint.

  - Mean amplitude of Tasks I to IV for each visit [RMS deg]
- Time course of absolute values and of changes from study baseline of
  - TETRAS ADL UL score (Score no. 2 in Appendix 8.1)
  - TETRAS ADL Functional Impact score (Score no. 3 in Appendix 8.1)
  - TETRAS ADL subscale score (Score no. 1 in Appendix 8.1)
  - TETRAS Performance subscale score (Score no. 4 in Appendix 8.1) as assessed by the investigator
  - TETRAS Performance dominant UL score (Score no. 5 in Appendix 8.1) as assessed by the investigator
  - TETRAS Performance dominant UL score (Score no. 5 in Appendix 8.1) as assessed by IRP (median across IRP members)
- Time course of absolute values and changes from study baseline of QUEST total score and dimension scores
- [redacted]

For definition of study baseline see Section 4.

Bilateral treatment period

- Time course of absolute values in the angular tremor amplitude and changes from study baseline and Cycle 2 baseline of 1) wrist, 2) elbow, and 3) shoulder level of the treated a) motor dominant UL, b) non-dominant UL, and c) difference between both ULs in the bilateral treatment period for the:
  - Task with maximum amplitude at Cycle 2 baseline for both ULs [RMS deg]
  - Task with maximum amplitude at study baseline for motor dominant UL [RMS deg]

  Changes from study baseline are only analyzed for motor-dominant UL. In accordance with the visit schedules for treatment Cycles 1 and 2 in the CSP (see Tables 5 and 6), angular tremor amplitude of motor-dominant UL was only assessed in Cycle 2, i.e., no study baseline is available for the non-dominant UL.
- Time course of absolute values and of changes from study baseline and Cycle 2 baseline of:
  - TETRAS ADL UL score (Score no. 2 in Appendix 8.1)
  - TETRAS ADL Functional Impact (Score no. 3 in Appendix 8.1)
  - TETRAS ADL subscale score (Score no. 1 in Appendix 8.1)
  - TETRAS Performance dominant UL score (Score no. 5 in Appendix 8.1) as assessed by the investigator[4]
  - TETRAS Performance non-dominant UL score (Score no. 6 in Appendix 8.1) as assessed by the investigator[4]
  - TETRAS Performance subscale score (Score no. 4 in Appendix 8.1) as assessed by the investigator[4]

[4] For precision, "as assessed by the investigator" was added to definitions of other efficacy endpoints based on TETRAS performance assessments in Cycle 2 as included in the CSP (see Section 6).

- Subject´s GICS score of motor dominant UL at Week 6 of Cycle 2
- Investigator´s GICS score of motor dominant UL at Week 6 of Cycle 2
- Time course of absolute values and changes from study baseline and Cycle 2 baseline of QUEST total score and dimension scores
- [redacted]

For definition of Cycle 2 baseline see Section 4.

2.2.2 Safety Endpoints

2.2.2.1 Primary Safety Endpoint

No primary safety endpoint has been defined.

2.2.2.2 Secondary Safety Endpoint

The secondary safety endpoint for both the unilateral and bilateral treatment period will be the following:

- Incidence of treatment-emergent AEs [TEAEs] related to treatment as assessed by the investigator.

2.2.2.3 Other Safety Endpoints

- Exposure to study treatment (including injected volume [mL], number of injection points) by muscle and treated UL for both study periods
- Incidence of TEAEs, TEAEs of special interest (TEAESIs), serious TEAEs, TEAEs leading to discontinuation, COVID-19 related TEAEs for both study periods.
- Time course (i.e., absolute values and changes from study baseline[5]) during the unilateral treatment period (Cycle 1) in Handheld Dynamometer (HHD) maximum grip strength (in hand of injected UL) [Newton] and in maximum grip strength difference between both hands [Newton].
- Time course (i.e., absolute values and changes from Cycle 2 baseline) during the bilateral treatment period (Cycle 2) in HHD maximum grip strength in both hands, by hand of motor-dominant/non-dominant UL [Newton].
- [redacted]
- [redacted]
- Global assessment of tolerability scale as assessed at Week 24 (V8) of unilateral treatment period and at Week 12 (V11) of bilateral treatment period.

[5] Wording in CSP Section 12.4.5.2 was “changes from applicable cycle baseline”. For consistency, “Cycle 1 baseline” is referred to as “study baseline” everywhere in this document. This does not constitute a change in planned analyses.

- Time course of vital signs (blood pressure, pulse rate) measurement (absolute values and changes from study baseline) during both study periods.
- Time course of body weight measurement and BMI (absolute values and changes from study baseline) during both study periods.
- Time course of standard clinical chemistry and hematology parameter measurements (absolute values and changes from study baseline) during both study periods.
- Time course of suicidality assessment during both study periods.

For definition of study baseline and Cycle 2 baseline see Section 4.

2.3 Other Variables

Other variables are:

- Subject disposition
    - Enrolment (including subjects screened, randomized, treated in Cycle 1, treated in Cycle 2, subjects per analysis set)
    - Study completion/ premature discontinuation of participation (completed Cycle 1/ Cycle 2/ overall, discontinued study in Cycle 1/ Cycle 2/ overall [in total and due to COVID-19 pandemic], with main reason for premature study discontinuation)
    - Visit attendance in Cycle 1 and in Cycle 2 (subjects expected, visit performed, visit not performed, study prematurely discontinued)
- Demographic data and other baseline characteristics (including gender, age [years], age category [18-64 years, 65-84 years, and ≥ 85 years], ethnicity, race, body height [cm], body weight [kg], body mass index [BMI; kg/m²])
- History of ET (age of onset of ET estimated by patient [years], duration since onset of ET (as estimated by patient) [years], family history of ET [yes/no], additional soft neurological signs with ET of uncertain significance [yes/no], motor dominant arm [left/right], tremor on other location(s) than both upper limbs: head [yes/no], face [yes/no], voice [yes/no], lower limb(s) [yes/no], at least one other location [yes/no])
- Previous and concomitant medication
- Previous and concomitant non-drug treatments (procedures)
- Medical history and concomitant diseases
- Timepoint of performance and evaluation of 12-lead ECG
- Injection guidance technique [ultrasound, electromyography, electrical stimulation] by muscle
- Injection cycle length [weeks]
- Total observation period [weeks]

2.4 Other Data

- Important protocol deviations (PDs), PDs[6] and other events related to COVID-19 pandemic
- Reasons for exclusion of subjects from analysis sets

3 Analysis Sets

The following analysis sets are defined for the statistical analysis of this study:

Safety evaluation sets

The Safety evaluation set for the unilateral treatment period (SES-UP) includes all subjects who received NT 201 or placebo during the unilateral treatment period.

The Safety evaluation set for the bilateral treatment period (SES-BP) includes all subjects who received NT 201 during the bilateral treatment period.

Full analysis sets

The Full analysis set for the unilateral treatment period (FAS-UP) is the subset of subjects in the SES-UP for whom at least one score of tremor amplitude at wrist level (injected motor-dominant UL) at study baseline and at least at one post-baseline score (V3 to V8) during the unilateral treatment period are available.

For definition of study baseline see Section 4.

The Full analysis set for the bilateral treatment period (FAS-BP) is the subset of subjects in the SES-BP for whom at least one score of tremor amplitude at wrist level at Cycle 2 baseline visit and at least at one post-baseline score (V9 to V11) of both arms during the bilateral treatment period are available.

For definition of study baseline see Section 4.

Per protocol set

The Per protocol set (PPS) is the subset of subjects in the FAS-UP without major PDs. Major PDs will be defined during the blinded data review meeting.

In any analyses on randomized subjects, the FAS-UP, FAS-BP, and the PPS, subjects will be included according to the randomized treatment group. In any analyses on the SES-UP and SES-BP, subjects will be included according to the actual treatment in Cycle 1.

4 Statistical Analyses

Descriptive statistics for metric variables will be number of values analyzed (n), number of missing values (nmiss, if any), number of imputed values (nimp, where applicable), mean, standard deviation (SD), median, quartiles, minimum, and maximum. Mean, quartiles, and median will be reported to one decimal place more than the data were collected, for the SD two decimal places more will be displayed; for calculation of derived variables, the number of decimal places is given in the sections below and the number of decimal places to be displayed for descriptive statistics of metric derived variables is defined in Appendix 8.7. For both collected and derived data, descriptive statistics for changes from study / Cycle 2 baseline, as applicable, will be displayed with the same number of decimal places.

[6] including important and non-important PDs

Frequency tables for qualitative variables will be absolute and percent frequencies (n, %). Frequency tables will include the number of missing values, where applicable. If not specified otherwise, calculation of percentages will be based on all subjects from a specified population and treatment group with non-missing data at the respective visit. 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 2-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 (CIs) will be 2-sided for 95% confidence level.

The last valid value before first administration of the investigational product (IP) will be analyzed as study baseline and the last valid value before IP administration in Cycle 2 will be analyzed as Cycle 2 baseline , but only considering data from Cycle 2 baseline visit (V8) and from control visit at Week 18 of Cycle 1 (V7).

The variable label "treatment group" will be abbreviated with "treatment" in tables, figures, and listings. For analyses of Cycle 1 and Cycle 2 data by treatment group, labels "C1 NT 201" and "C1 Placebo" will be used in respective tables. For analyses by visit the following naming convention will be used: Screening (V1), Study baseline (with "(V2)" for visit attendance table), Week 4 (V3), Week 6 (V4), Week 8 (V5), Week 12 (V6), Week 18 (V7), Week 24 (V8), Cycle 2 baseline (with "(V8(a))" for visit attendance table, tables displaying summary statistics for laboratory data, vital signs, body weight, and BMI in Cycle 2), Week 4 (V9), Week 6 (V10), EOS Week 12 (V11) (without "EOS" in visit attendance table and for efficacy endpoints based on TETRAS ADL or TETRAS Performance assessments or kinematic tremor assessment with the TremorTek® device). For subjects treated at V8, data recorded at V8 will be analyzed as "Week 24 (V8)" and as "Cycle 2 baseline". For subjects treated at V8a, V8 data will be analyzed as "Week 24 (V8)" and V8a data as "Cycle 2 baseline". For subjects discontinuing the study prior to the last scheduled visit at Week 12 in Cycle 2, efficacy endpoints based on TETRAS ADL or TETRAS Performance assessments or kinematic tremor assessment with TremorTek performed at their EOS visit (V11) will be analyzed under the study visit which is closest to the recorded V11 date. Details are described in Section 4.7.4. All other data collected at EOS visit (V11) will generally be analyzed at V11. Except of listing "Visit dates and premature study discontinuation" and listings of violation/ nonfulfillment of re-injection eligibility criteria, Visit V8 and V8a will be listed as "V8(a)". In listing "Visit dates and premature study discontinuation", the following long labels will be used for these two visits: "Week 24 (V8)" and "Cycle 2 baseline (V8a)". Label "EOS Week 12 (V11)" will be used in listings "Important dates" and "Visit dates and premature study discontinuation".

4.1 Efficacy Analyses

If not otherwise specified, efficacy analyses

- for Cycle 1 will be performed on observed cases (OC) in the FAS-UP and by treatment group
- for Cycle 2 will be performed on FAS-BP (OC), by treatment group and in total. However, changes from study baseline (as defined at beginning of Section 4) in efficacy endpoints will generally only be analyzed by treatment group, not in total. "In total" means a pooled analysis of data from both treatment groups. The label "Total" will be used in respective tables.

In case angular tremor amplitude assessments of Cycle 1 are erroneously performed on non-dominant arm, values will only be listed. In the bilateral treatment period, it will be distinguished between:

- *Category A:* endpoints that are

- general (i.e., TETRAS ADL scores, TETRAS Performance subscale scores, GICS scores, QUEST dimension and total scores, [redacted], and EQ VAS score) or
- applicable to one UL only (TETRAS Performance subscale dominant UL, TETRAS Performance subscale non-dominant UL).

- *Category B:* endpoints that are determined for both ULs (dominant and non-dominant) separately (i.e., TremorTek® angular tremor amplitude at wrist, elbows, shoulder level for the task with maximum amplitude at cycle baseline and changes thereof from study baseline and Cycle 2).

If not otherwise specified, endpoints in Category B will be summarized by UL (dominant/non-dominant).

4.1.1 Primary Efficacy Endpoint

The primary efficacy endpoint is defined in Section 2.2.1.1.

4.1.1.1 Primary analysis

The primary efficacy endpoint is based on assessments of angular tremor amplitude [deg] for the motor-dominant UL at study baseline[7] and Week 6 (V4) with standardized computerized kinematic analysis using the TremorTek® device. At each visit, these assessments are performed for four tasks, repeated in three assessments rounds (for further details, please refer to CSP Section 9.1.1.2.1). From these data, the primary efficacy endpoint will be derived as follows:

- The maximum angular tremor amplitude at wrist level will be computed in the following three steps at study baseline (defined at beginning of Section 4):
  1. For each of the four tasks and each of the three assessment rounds, individual 20-second recordings from a kinematic sensor will yield three angular tremor amplitude values in deg for flexion/extension, radial deviation/ulnar deviation and for pronation/supination levels and a single combined Root Mean Square (RMS) angular tremor amplitude value. The RMS will be calculated as (rounded by five decimal places):

$$RMS = \sqrt{\frac{x_1^2+x_2^2+x_3^2}{3}},$$

     where
     x_1 = angular tremor amplitude for flexion/extension [deg]
     x_2 = angular tremor amplitude for radial deviation/ulnar deviation [deg]
     x_3 = angular tremor amplitude for pronation/supination [deg]
  2. Calculation of the arithmetic mean from the three assessment rounds of the combined RMS for each of Tasks I to IV [RMS deg], separately for each task (rounded by five decimal places).
  3. Among Tasks I to IV selection of the task with the maximum arithmetic mean calculated during Step 2. The amplitude of this task [RMS deg] will be considered for the primary endpoint.

[7] For definition of study baseline see beginning of Section 4.

- For the selected task (see Step 3), the angular tremor amplitude at wrist level at Week 6 (V4) [RMS deg] will be derived analogously, i.e., as arithmetic mean of the combined RMS (see Steps 1 and 2 above) based on the respective data from V4.
- Change from study baseline to Week 6 [RMS deg] will be calculated by subtracting the maximum angular tremor amplitude at wrist level at study baseline from the respective value at Week 6 (V4), derived as described above.

Derivation of maximum tremor amplitude at wrist level at study baseline (as defined at beginning of Section 4) according to Steps 1 to 3 (see above) is only meaningful if the three angular tremor amplitude values for flexion/extension, radial deviation/ulnar deviation and for pronation/ supination levels are each available for all four tasks from all three assessments rounds. Likewise, maximum tremor amplitude at wrist level at Week 6 (V4) can only be calculated as described above if the three above-mentioned angular tremor amplitude values are each available for the selected task from all three assessment rounds. If any of these study baseline or Week 6 (V4) data are missing, the primary endpoint cannot be derived and will be considered as missing. Under the assumption of missingness at random (MAR), missing values will not be replaced for the primary analysis.

The confirmatory analysis of the primary efficacy endpoint will be based on the FAS-UP (OC) using least square means (LS Means) from a baseline-adjusted analysis of covariance (ANCOVA) as implemented in SAS PROC MIXED. Fixed class effects will be the treatment group (NT 201 versus placebo) and study site. Study baseline angular tremor amplitude at wrist level for the task with the maximum amplitude will be the fixed continuous covariate. Variances will be allowed to differ between treatment groups. Denominator degrees of freedom will be calculated according to the Kenward-Rodger method.

The difference in LS Means between treatment groups with associated 95% CI from this model will serve as an estimate for the size of the treatment effect. The two-sided t-test of the LS Mean treatment difference will be applied to the following testing problem at $\alpha_{2\text{-sided}}$=5%:

$$H_0\text{: } \mu_{NT201} - \mu_{Placebo} = 0 \quad \text{vs.} \quad H_1\text{: } \mu_{NT201} - \mu_{Placebo} \neq 0,$$

where $\mu_{NT201} - \mu_{Placebo}$ denotes the expected mean treatment difference on the primary efficacy endpoint, with a negative difference indicating a treatment effect in favor of NT 201.

In case of convergence issues, study sites might be pooled for statistical analysis. Prior to unblinding, no need of pooling was seen for the primary analysis (see also Section 4.7.3).

SAS code similar to the following statements will be used:

```
PROC MIXED data = dataset;
    CLASS TRT01P (REF='C1 Placebo') SITEID;
    MODEL CHG = TRT01P SITEID BASE / solution ddfm=kr;
    LSMEANS TRT01P / pdiff=control('C1 Placebo') CL alpha=0.05;
    REPEATED / group=trt01p;
    WHERE avisit eq 'Week 6 (V4)' and param eq 'Angular tremor amplitude at
    wrist level dominant UL: Task with maximum amplitude at study baseline';
RUN;
```

Here, “TRT01P” denotes the randomized treatment group, “SITEID” the study site, “CHG” the change from study baseline to Week 6 (V4) in maximum tremor amplitude measurement at wrist level [RMS deg] during the unilateral treatment period (i.e., the primary endpoint, derived as described above), and “BASE” maximum tremor amplitude measurement at wrist level at study baseline (derived as described above).

The following statistics will be displayed:

- n (number of values analyzed)
- p-values of fixed class effects and the fixed continuous covariate
- LS Means of each treatment group (including standard error [SE] and 95% CI)
- LS Mean treatment difference (including SE, 95% CI, and t-test p-value).

4.1.1.2 Sensitivity Analyses

If not specified otherwise, the exploratory sensitivity analyses described in the following will be performed on the FAS-UP (OC). Study sites might be pooled for statistical analysis to overcome convergence issues in ANCOVA models. Prior to unblinding, the need of pooling was only seen for sensitivity no. 2 (see below and Section 4.7.3).

- *Sensitivity analysis no. 1:* The ANCOVA model specified in Section 4.1.1.1 will be repeated on the PPS (OC).
- *Sensitivity analysis no. 2:* For the analysis of the treatment-by-pooled site interaction, study sites from the United States of America [USA] and Canada will be grouped together ("North American [NoA] sites") and compared vs. all remaining study sites from Poland ("Polish [Pol] sites"). The ANCOVA model specified in Section 4.1.1.1 will be modified as follows to estimate treatment effects by pooled study site: The fixed effect study site will be replaced by fixed effects for pooled study site [NoA vs. Pol] and the treatment-by-pooled study site interaction. Apart from the respective modifications in the MODEL and LSMEANS statement, the SAS code will be identical to the one provided in Section 4.1.1.1. LS-Means by treatment group and LS Mean treatment differences will not only be provided overall but also by pooled study site (NoA vs. Pol).[8]
- *Sensitivity analysis no. 3:* Gender[9] (as fixed class effect) and age (as fixed continuous covariate) will be added to the model specified in Section 4.1.1.1. The SAS code will be similar to the one provided there, with the only modification that gender[9] and age will be added to the MODEL statement and gender[9] additionally to the CLASS statement. The same statistics as those specified in Section 4.1.1.1 (with p-values for all fixed class effects and the fixed continuous covariates in the model) will be displayed.
- *Sensitivity analysis no. 4:* In case more than 10% missingness on the primary efficacy endpoint occurs in either treatment group among all treated subjects, multiple imputation of missing maximum tremor amplitude measurements at wrist level at study baseline and at Week 6 (V4)[10,11] will be performed and results for the primary efficacy endpoint from an ANCOVA

[8] In the CSP, a model "with treatment-by-site interaction to estimate treatment effects by site" was planned, with the option to pool sites for statistical analysis and the statement that "the necessity and details of pooling will be determined under double blind conditions". To avoid convergence issues due to the high number of study sites in relation to the number of subjects to be analyzed, it was decided prior to unblinding to pool sites (NoA sites vs. Pol sites) for this interaction analysis. This adaption is thus not considered as a change from analyses planned in the CSP.
[9] Standard name for variable is 'sex' in CDISC SDTM/ADaM.
[10] Note that in case the task with maximum amplitude at wrist level at study baseline cannot be derived due to missing or incomplete assessments of tremor amplitude at wrist level at study baseline, the respective value at Week 6 (V4) cannot be derived either and hence needs to be imputed as well, even if TremorTek measurements of angular tremor amplitude at wrist level are completely available at Week 6 (V4) or not.
[11] Maximum tremor amplitude at V3/V4 to be calculated for the task with maximum tremor amplitude at study baseline.

model (as described in Section 4.1.1.1) applied to 1000 completed datasets will be combined for statistical inference. This analysis will be performed on all subjects treated in Cycle 1.

In detail, the following four steps will be performed:

1. One thousand completed datasets will be generated using the FCS statement with the regression method for arbitrary missing data patterns in SAS PROC MI. Variables considered will comprise maximum tremor amplitude at wrist level from previous visits[12], treatment, study site, gender[9] [female/male], and age [years]. SAS code similar to the following will be used:

```
PROC MI DATA=dataset SEED=602011079 NIMPUTE=1000 OUT=mi_
   MINIMUM=. . . 0 0 0
   ROUND=  . . . 1 1 1;
  WHERE avisit="Week 6 (V4)" and param="Angular tremor
  amplitude at wrist level dominant UL: Task with maximum
  amplitude at study baseline";
  CLASS SITEID SEX TRT01P;
  FCS REG(BASE = TRT01P SITEID SEX AGE);
  FCS REG(V3VAL = TRT01P SITEID SEX AGE BASE);
  VAR TRT01P SITEID SEX BASE V3VAL AVAL;
RUN;
```

where
SITEID = study site
SEX = gender
TRT01P = randomized treatment group
AVAL = maximum tremor amplitude at wrist level at Week 6 (V4)[13]
V3VAL = maximum tremor amplitude at wrist level at Week 4 (V3)[13]
BASE = study baseline value for maximum tremor amplitude at wrist level

2. The primary endpoint will be calculated from observed and imputed maximum tremor amplitude at wrist level at study baseline and at Week 6 (V4) in the completed datasets.

3. For each completed dataset, the primary endpoint will be analyzed by an ANCOVA model as defined in Section 4.1.1.1 whereby “ods output diffs=diff_mi LSMeans=lsm_mi” will be added to the SAS code.

4. The results from the 1000 imputed datasets will be combined for statistical inference using PROC MIANALYZE. SAS code similar to the following will be used:

[12] including study baseline value for imputation of missing values at Week 4 (V3) and study baseline and Week 4 (V3) values for imputation of missing values at Week 6 (V4).
[13] calculated using measurements for the task with maximum tremor amplitude at study baseline

```
PROC MIANALYZE data=diff_mi;
  MODELEFFECTS estimate;
  STDERR stderr;
  ODS OUTPUT ParameterEstimates=DIFF_MIAN;
run;

PROC MIANALYZE data=lsm_mi;
  BY treatment;
  MODELEFFECTS estimate;
  STDERR stderr;
  ODS OUTPUT ParameterEstimates=LSM_MIAN;
run;
```

The following statistics will be displayed :

- n (number of values analyzed) and nimp (number of imputed values)
- LS Means of each treatment group (including SE and 95% CI)
- LS Mean treatment difference (including SE, 95% CI, and p-value).

All results from the above-described sensitivity analyses, including any CIs and p-values provided, will be interpreted descriptively only.

4.1.1.3 Supplementary Analyses

Primary endpoint data will moreover be summarized by randomized treatment group using descriptive statistics for metric variables (see beginning of Section 4.1) on the FAS-UP (OC) and on the PPS (OC).

4.1.2 Secondary Endpoints

4.1.2.1 Key Secondary Efficacy Endpoints

The two key secondary efficacy endpoints are defined in Section 2.2.1.2.1 and will derived as follows:

- The value of the respective TETRAS score (TETRAS Performance dominant UL score and TETRAS ADL UL score) will be calculated for study baseline[14] and for Week 6 (V4) as sum of all corresponding items as defined in Appendix 8.1 (see scores no. 5 and no. 2, respectively).
- Change from study baseline[14] to Week 6 (V4) will be calculated by subtracting the value of the respective TETRAS score at study baseline from the respective value at Week 6 (V4), derived as described above.
- In case an item belonging to the respective TETRAS score is missing, the respective TETRAS score value cannot be meaningfully derived for the affected visit. Hence, derivation of each of the two key secondary endpoints requires that answers to all items belonging to the respective TETRAS score are available at both study baseline and Week 6 (V4). Otherwise, the respective key secondary endpoint will not be calculated and considered as missing.

The two key secondary efficacy endpoints will be tested with ANCOVA models similar to that for the primary analysis of the primary endpoint (see Section 4.1.1.1). However, the study baseline of the analyzed TETRAS score [i.e., TETRAS Performance dominant UL score (Score no. 5 in

[14] For definition of study baseline see beginning of Section 4.

Appendix 8.1) or TETRAS ADL UL score (Score no. 2 in Appendix 8.1), as applicable] will be used for baseline adjustment. Handling of multiplicity is addressed in Section 4.7.2.

Missing data handling, pooling of study sites, exploratory sensitivity analyses (Section 4.1.1.2), and supplementary analyses (Section 4.1.1.3) will also be handled analogously to the respective analyses of the primary endpoint. For ANCOVA-based sensitivity analyses, the respectively analyzed TETRAS subscale score (see above) will be used. The multiple imputation based sensitivity analysis was planned to be performed in case of more than 10% missingness on the respective key secondary endpoint among all treated subjects. During the DRM performed on 17Jan2024 under double blind conditions, the two key secondary efficacy endpoints were found to be completely available for all treated subjects. It was hence concluded that multiple imputation is not required for key secondary endpoints.

4.1.2.2 Other Secondary Efficacy Endpoints

Other secondary endpoints are defined in Section 2.2.1.2.2. TETRAS based endpoints will be derived analogously as described in Section 4.1.2.1 for the two key secondary efficacy endpoints (see also Appendix 8.1 for assignment of TETRAS items to the different TETRAS scores). This requires that answers to all respective items are available. No missing value replacement will be performed for any other secondary efficacy endpoint. Inferential statistical results reported for other secondary efficacy endpoints will be interpreted only descriptively.

Unilateral treatment period:

Other secondary endpoints of the unilateral treatment period will be analyzed by applying ANCOVA models similar to that for the primary analysis of the primary endpoint (see Section 4.1.1.1). However, for TETRAS subscales, the study baseline of the analyzed TETRAS (subscale) score will be used for baseline adjustment. For GICS by investigator and GICS by subject, the study baseline of the TETRAS ADL subscale score will be used for baseline adjustment. The same statistics as defined in Section 4.1.1.1 will be reported.

Furthermore, descriptive statistics for metric variables will be provided by treatment group.

Bilateral treatment period:

Other secondary endpoints of the bilateral treatment period all belong to Category A, (see Section 4.1). They will be analyzed by applying ANCOVA models similar to that for the primary analysis of the primary endpoint (see Section 4.1.1.1). However, the Cycle 2 baseline of the analyzed variable will be used for baseline adjustment.

Since all subjects will receive the same active treatment (NT201) in Cycle 2, it is not intended to test the effect of treatment for any other secondary endpoint from the bilateral treatment period.

Rather, the above mentioned ANCOVA models will be used to estimate the average changes from Cycle 2 baseline by LS Means and respective 95% confidence intervals.

SAS code similar to the statements provided in Section 4.1.1.1 will be used. Only the LSMeans statement will be adapted as follows because the LS Mean treatment difference is not needed:

```
LSMEANS TRT01P / OM CL alpha=0.05;
```

The following statistics will be displayed:

- n (number of observations used in the analysis)
- p-values of fixed class effects (randomized treatment group and study site) and the fixed continuous covariate (baseline severity level of the respectively analyzed variable)
- LS Mean per treatment group with associated SE and 95% CI.

Furthermore, adequate descriptive statistics will be provided for the "total active" treatment group in Cycle 2 and by randomized treatment group.

4.1.3 Other Efficacy Endpoints

- Derived other efficacy endpoints will be computed as follows:
  - Endpoints based on assessments of angular tremor amplitude [deg] in the unilateral treatment period:
    - For the task with maximum amplitude at study baseline, derivations will be performed as follows:
      - The task with maximum amplitude at study baseline at wrist level will be determined by performing steps 1 to 3 for calculation of the primary efficacy endpoint (see Section 4.1.1.1). The same approach will be applied for selection of the task with maximum amplitude at study baseline at elbow and shoulder levels.
      - For the respectively selected task, the angular tremor amplitude at wrist/ shoulder/ elbow level will be derived analogously, i.e. as arithmetic mean of the combined RMS (see Steps 1 and 2 in Section 4.1.1.1) for each post-baseline visit in Cycle 1 [RMS deg].
      - Changes from study baseline to all post-baseline visits in Cycle 1 at wrist/ elbow/ shoulder level will be derived as described for wrist level at Week 6 (V4) [RMS deg] in Section 4.1.1.1.

      Note: In case the task with maximum amplitude at wrist/ shoulder/ elbow level at study baseline cannot be derived due to missing or incomplete assessments of tremor amplitude at corresponding level at study baseline, respective values and changes from study baseline at any post-baseline visits cannot be derived either and will as well be considered as missing. This applies regardless of whether TremorTek assessments wrist level are completely available at Week 6 (V4) or not.
    - For the maximum tremor amplitude from Tasks I to IV [RMS deg], steps 1 to 3 as for calculation of maximum tremor amplitude at wrist level at study baseline (see Section 4.1.1.1) will be performed. The selection of the task with the maximum arithmetic mean (step 3) will be performed separately for each visit and for each level (wrist/elbow/shoulder). I.e., for calculation of values and changes from study baseline, only the task with maximum tremor amplitude at respective visit will be considered.
    - For the mean amplitude of Tasks I to IV [RMS deg], step 1 and step 2 as for calculation of maximum amplitude (see Section 4.1.1.1) will be performed whereby for the calculation of RMS respective measurements of wrist, elbow or shoulder will be used. The mean of the four tasks will be calculated (rounded by five decimals). In addition to values, changes from study baseline will be calculated.
  - Endpoints based on assessments of angular tremor amplitude in the bilateral treatment period will be derived analogously as described above for respective endpoints for the unilateral treatment period, but separately by motor dominant/ non-dominant UL as well as for the difference between both ULs. For all post-baseline visits in Cycle 2, changes will be derived as both changes from study baseline[15] and as changes from Cycle 2 baseline for

[15] For definition of study baseline see beginning of Section 4.

the motor dominant UL, as applicable. Changes from study baseline are only derived for motor-dominant UL because, in accordance with the visit schedules for treatment Cycles 1 and 2 in the CSP (see Tables 5 and 6), angular tremor amplitude of motor-dominant UL was only assessed in Cycle 2, i.e., no study baseline is available for the non-dominant UL.

- As defined in Section 4.1.1.1 for the primary efficacy endpoint, derivation of endpoints based on assessments of tremor amplitude generally requires that all respective assessments are available. Otherwise, the affected endpoint will be considered as missing.
- TETRAS based endpoints:
  - Endpoints based on TETRAS Performance by investigator and TETRAS ADL will be derived analogously as described in Section 4.1.2.1 for the two key secondary efficacy endpoints (see also Appendix 8.1 for assignment of TETRAS items to the different TETRAS scores).
  - Endpoints based on TETRAS Performance assessments by IRP: Values will be derived as medians over all available scores from IRP members: First, the respective TETRAS score (Score no. 5 in Appendix 8.1) will be calculated for each IRP member. Second, the median over the respective TETRAS score from all IRP members will be computed and rounded by 3 decimals whereby median will only be calculated if score is available for at least 2 out of 3 IRP members. Changes from study baseline will be calculated based on these medians.
  - As TETRAS based endpoints will only be calculated if answers to all respective items are available.
- Endpoints based on QUEST dimension and total scores:
  - This patient-reported outcome questionnaire has 30 items that inform on interference and impact of ET on five dimensions of QoL: Physical (9 items: items 13-21), Psychosocial (9 items: items 22-30), Communication (3 items: items 1-3), Hobbies/Leisure (3 items: items 10-12), and Work/Finance (6 items: items 4-9). Most items have 5-point frequency response scale ranging from 0 ('Never') to 4 ('Always'). For the items 6, 7, 11, and 12 a 'No' or 'Yes' response is asked for (scored as 0 or 4, respectively). Some items (4, 5, 7, 8, 14) have an additional answer option "Not applicable" (no score assigned, "not applicable" items will not be considered for analysis). The QUEST scores on the five dimensions/domains are each calculated based on applicable items[16] only and expressed as a percentage of the total score possible for the respective dimension/domain, ranging from 0 to 100 (rounded by two decimals).
  - Each dimension score is hence calculated as the sum of scores for all applicable items of the respective dimension divided by the sum of possible points (=number of applicable items * 4) of the respective dimension.
  - The total score (summary index) is calculated as the mean of the five domain scores, also ranging from 0 to 100 (rounded by two decimals). A higher score represents higher dissatisfaction with health-related QoL and disability.
  - In addition to values per visit, changes from study baseline will be calculated for applicable Cycle 1 and Cycle 2 visits and changes from cycle baseline additionally for Cycle 2.

[16] Applicable items = items for which another response option than "not applicable" was chosen

- QUEST scores and respective changes from study / Cycle 2 baseline will only be calculated if answers to all respective (and applicable) items are available.

- [redacted]

All above-mentioned changes from study/ Cycle 2 baseline (defined at beginning of Section 4) will be derived by subtracting the study/ Cycle 2 baseline value from the respective post-baseline value.

No missing value replacement will be performed for any other efficacy endpoint.

All other efficacy endpoints will be analyzed by descriptive statistics only.

4.2 Pharmacodynamic Analyses

Not applicable.

4.3 Pharmacokinetic Analyses

Not applicable.

4.4 Pharmacogenetic Analyses

Not applicable.

4.5 Safety Analyses

All safety analyses for Cycle 1 will be performed by actual treatment group on SES-UP. All safety analyses for Cycle 2 will be performed on the SES-BT, by actual treatment group in Cycle 1 and in total. All safety analyses for the total observation period will be performed on SES-UP, by actual treatment group in Cycle 1 and in total.

For calculation of TEAE based incidence rates, the denominator will be the number of subjects in the respective analysis set, by actual treatment group or in total (as applicable).

AEs will be coded according to the Medical Dictionary for Regulatory Activities [MedDRA] version that is in effect at the time the database is closed.

Only TEAEs will be analyzed, which are defined as AEs with onset or worsening after the first administration of the IP. For details on separation of TEAEs and non-TEAEs see Appendix 8.5.

If in case of a public health emergency, e.g. due to a COVID-19 outbreak, the second injection visit V8a of Cycle 2 is performed later than 25 weeks after the first injection visit V2 the AEs

[17] Health state = pattern of scores over the five dimensions, e.g., 13333 for a subject indicating slight problems in mobility and severe problems on the other four dimensions.

starting or worsening later than 25 weeks after baseline visit V2 of Cycle 1 and before injection at V8a of Cycle 2 are treated as non-TEAEs.

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

For analysis, incompletely recorded or completely missing start of AEs will be estimated according to rules laid down in Appendix 8.6. End dates will not be imputed. In the AE listing, start date as recorded and as imputed for analysis will be displayed.

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 in respective treatment cycle [+ 1 day for TEAEs]
- The duration of an AE will be calculated as stop date - onset date + 1 day. The duration of an AE will be missing if no AE stop date is provided

Treatment-relatedness of AEs will be determined based on the assessment of the investigator.

If a subject has more than one intensity/ causal relationship to treatment/ outcome within a preferred term (PT) only the worst intensity/ causal relationship to treatment/ outcome will be considered for calculation of incidence rates in the frequency tables reporting incidences by intensity/ causal relationship to treatment/ outcome. Similarly, for the analysis overall, i.e., for subjects with at least one TEAE, only the worst intensity/ causal relationship to treatment/ outcome category per subject will be counted for calculating incidence rates.

The worst outcome is defined in the following order (worst outcome to best outcome):

- fatal
- unknown
- not recovered/not resolved
- recovering/resolving
- recovered/resolved with sequelae
- recovered/resolved.

Relationship to COVID-19 disease is documented in eCRF. Furthermore, the medical coding specialist will check if further AEs could be related to COVID-19 disease (but also consider information from eCRF). The medical coding specialist will classify all AEs as related/ suspected/ not related to COVID-19 disease. Both kinds of information will be added in SDTMs and displayed in by-subject data listings of AEs. For analysis and the listing of TEAEs related to COVID-19 disease, information from coding specialist will be used, whereby “suspected” AEs will be regarded as related to COVID-19 disease.

The AEs listed in Table 1 are defined as adverse event of special interest (AESIs) for this study.

Table 1: List of Adverse Events of Special Interest Possibly Indicating Toxin Spread

MedDRA Preferred Term	MedDRA Preferred Term
Accommodation disorder	IIIrd nerve paresis
Areflexia	Ileus paralytic
Aspiration	IVth nerve paresis
Botulism	Monoparesis
Bradycardia	Muscular weakness
Bulbar palsy	Paralysis
Constipation	Paraparesis
Cranial nerve palsies multiple	Paresis
Cranial nerve paralysis	Paresis cranial nerve
Diaphragmatic paralysis	Peripheral nerve palsy
Diplopia	Peripheral paralysis
Dry mouth	Pelvic floor muscle weakness
Dysarthria	Pneumonia aspiration
Dysphagia	Pupillary reflex impaired
Dysphonia	Quadriparesis
Dyspnoea	Respiratory arrest
Extraocular muscle paresis*	Respiratory depression
Eyelid function disorder	Respiratory failure
Eyelid ptosis	Speech disorder
Facial paralysis	Trigeminal nerve paresis
Facial paresis	Urinary retention
Hemiparesis	Vision blurred
Hypoglossal nerve paresis	Vocal cord paralysis
Hyporeflexia	Vocal cord paresis
Hypotonia	

*LLT Extraocular muscle paresis, PT: Ophtalmoplegia
Wording of terms is according to MedDRA version 26.1.

TEAEs of Cycle 1 are defined as TEAEs with onset or worsening before date and time of administration of IP at V8(a). For subjects who do not receive study treatment during Cycle 2 all TEAEs will be considered as TEAEs of Cycle 1. TEAEs of Cycle 2 are defined as TEAEs with onset or worsening on or after date and time of the administration of IP at V8(a).

In case of missing intensity or missing causal relationship of an AE to treatment, 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 will be set to “unknown”.

4.5.1 Primary Safety Endpoint

No primary safety endpoint has been defined.

4.5.2 Secondary Safety Endpoints

The secondary safety endpoint is defined in Section 2.2.2.2. Frequency tables for subjects with treatment-related TEAEs and number of treatment-related TEAEs will be provided.

4.5.3 Other Safety Endpoints

- Exposure to study treatment will be analyzed as follows for both treatment periods:
  - For injected volume, descriptive statistics for metric variables will be provided for by injected muscle, muscle group (total injected volume for forearm/wrist muscles, elbow muscles, shoulder muscles), and total injected UL. The number of subjects injected in respective muscle / in at least one muscle of respective muscle group (M) will be displayed as well.
    - For Cycle 1, the analysis will be performed for the injected UL, irrespective of whether the motor dominant UL (acc. to CSP) or the non-dominant UL was treated.
    - For Cycle 2, the analysis will be performed separately by motor dominant and non-dominant UL, and in total for both injected ULs, but only for muscle group and total UL (i.e., total injected volume over both ULs will not be provided separately by muscle). In case, deviating from CSP, not all subjects were treated in both ULs, the analyses will be performed on the subset of subjects treated at the respective UL / both UL(s), as applicable, in Cycle 2, and an explanation (including Subject ID, treatment group, untreated UL [motor dominant or non-dominant] of affected subjects) will be provided in a footnote.

    For both cycles, only data with injected volume larger than 0 ml will be included.

  - Number of injection points by muscle [including 0 to account for non-injected muscles for treated UL(s)], by muscle, muscle group (sum of injection points over forearm/wrist muscles / elbow muscles / shoulder muscles) and for total injected UL (total sum of injection points of injected UL).
    - For Cycle 1, the analysis will be performed for the injected UL, irrespective of whether the motor dominant UL (acc. to CSP) or the non-dominant UL was treated.
    - For Cycle 2, the analysis will be performed separately by motor dominant and non-dominant UL. In case, deviating from CSP, not all subjects in SES-BP were treated at both ULs, the analyses will be performed on the subset of subjects treated at the respective UL and an explanation (including Subject ID, treatment group, untreated UL [motor dominant or non-dominant] of affected subjects) will be provided in a footnote.

- Adverse events
  - The following summary tables for Section 14.3 of the CSR will be created for total observation period, Cycle 1 and Cycle 2 (if not stated otherwise below):
  - Overall summary of TEAEs, displaying incidences of TEAEs (in total / related to treatment), TEAESIs (in total / related to treatment), serious TEAEs (in total / related

to treatment, TEAEs leading to premature study discontinuation (in total / related to treatment), fatal TEAE (in total / related to treatment) and TEAEs related to COVID-19

- TEAEs, subjects with TEAEs and number of TEAEs by MedDRA SOC and PT
- TEAEs, subjects with TEAEs by PT
- TEAEs by worst intensity, subjects with TEAEs
- TEAEs by worst causal relationship, subjects with TEAEs
- TEAEs by worst outcome, subjects with TEAEs
- Serious TEAEs, subjects with serious TEAEs and number of serious TEAEs by SOC and PT
- Non-serious TEAEs, subjects with non-serious TEAEs and number of non-serious TEAEs by SOC and PT
- Treatment-related non-serious TEAEs
- TEAESIs, subjects with TEAESIs and number of TEAESIs by SOC and PT
- Treatment-related TEAESIs

  - The following by-subject listings of TEAEs occurring in the total observation period will be generated for Section 14.3 of the CSR and include information on treatment cycle:

- Serious TEAEs, listing of subjects
- Deaths, listing of all serious TEAEs leading to death
- TEAEs leading to premature study discontinuation of study, listing of subjects
- TEAEs of special interest, listing of subjects
- TEAEs related to COVID-19, listing of subjects.

  - In addition, a complete listing of all AEs (TEAEs and non-TEAEs) occurring in the total observation period will be provided for CSR Section 16.2, also including information on treatment cycle for TEAEs.

- HHD maximum grip strength: Descriptive statistics for HHD maximum grip strength [Newton] (values and changes from applicable cycle baseline, i.e., study baseline for Cycle 1, Cycle 2 baseline for Cycle 2) will be provided by visit for both injection cycles.

  - For Cycle 1, the above-specified analyses will be performed only for treated UL, irrespective of whether the motor dominant UL (acc. to CSP) or the non-dominant UL was treated. Data for non-injected UL will be listed. Furthermore, the intra-individual difference of changes in maximum grip strength from study baseline between hand of treated and hand of untreated UL will be analyzed. The difference will be derived as change from study baseline at hand of injected UL – change from study baseline at hand of non-injected UL.

  - For Cycle 2, maximum grip strength will be summarized separately for motor dominant and non-dominant UL. In case, deviation from CSP, not all subjects in the SES-BP were treated at both ULs, the analyses will be performed on the subset of subjects treated at the respective UL in Cycle 2 and an explanation (including Subject ID, treatment group, untreated UL [motor dominant or non-dominant] of affected subjects) will be provided in a footnote. In case either or both UL are erroneously not injected in Cycle 2, respective values and changes will only be listed.

- [redacted]

- Global assessment of tolerability as assessed at Week 24 of unilateral treatment period will be summarized descriptively by treatment group. Week 12 of bilateral treatment period will be summarized descriptively by treatment group and overall. This includes frequency tables and descriptive statistics for metric variables.
- Time course of vital signs (blood pressure, pulse rate), body weight measurement and BMI at selected visits (values and changes from study baseline in both injection cycles) will be summarized descriptively;
- Time course of standard clinical chemistry and hematology parameters measurements at selected visits (values and changes from study baseline in both injection cycles) will be summarized descriptively.

  Laboratory values from Screening Visit V1 will serve as study baseline. In case laboratory testing scheduled for Screening Visit V1 was performed after V1 but before Baseline Visit V2 (treatment visit), these values will be analyzed as study baseline. I.e., in line with definition of study baseline provided at beginning of Section 4, baseline flag is 'yes' for last record with non-missing value before first treatment in SDTM). The first valid value is analyzed for a post baseline visit. In addition to descriptive statistics, shift tables from study baseline to Week 24 (V8) and from study baseline to EOS (V11) will be provided.

  In case no exact value is given because a laboratory value is below limit of detection and thus provided as <x.x, the respective value (x.x) will be used for calculation of descriptive statistics for metric variables, and <x.x will be listed and flagged accordingly.
- Other safety data will be listed only.

For definition of study baseline and Cycle 2 baseline see beginning of Section 4.

4.6 Other Analyses

4.6.1 Other Variables

Other variables will be summarized descriptively by treatment group and in total if not stated otherwise.

- Enrollment variables will be analyzed overall and by study site. Absolute and percent frequencies will be displayed for all variables except for "Subjects screened" (absolute frequencies only).
    - "Subjects randomized", "Subjects treated in Cycle 1", "Subjects treated in Cycle 2", Subjects in FAS-UP, Subjects in FAS-BP, and Subjects in PPS will be analyzed on all randomized subjects
    - "Subjects in SES-UP" and "Subjects in SES-BP" will be analyzed based on the SES-UP.
- Variables on study completion/ premature discontinuation of participation will be summarized using frequency tables. Analyses or Cycle 1 and the total observation study will be based on the SES-UP, those for Cycle 2 on the SES-BP.
- Visit attendance variables will be analyzed overall and by treatment group for all visits defined in visit schedules for Cycle 1 and Cycle 2 the CSP (see CSP Tables 5 and 6, respectively). Frequency tables will be created based on all randomized subjects for Cycle 1 and based on SES-BP for Cycle 2.
    - For "Subjects expected at visit", absolute frequencies will be reported. All randomized subjects are expected for Screening visit (V1) and study baseline (V2). For all following visits in Cycle 1, the number of "Subjects expected at visit" will be the number of randomized subjects not yet having prematurely discontinued study participation until respective visit. Visits after premature discontinuation are generally not expected. The number of "Subjects expected at visit" is thus calculated for all post-study baseline visits in Cycle 1 as number of randomized subjects – the number of randomized subjects that discontinued study participation before the respective visit.
    - For Cycle 2 baseline, the number of subjects expected at visit will be the number of subjects in SES-BP. For all following visits in Cycle 2 [including Week 12 (V11)], the number of "Subjects expected at visit" will be the number of subjects in SES-BP not yet having prematurely discontinued study participation until respective visit. Visits after premature discontinuation are generally not expected. I.e., V11 (Week 12) will only be displayed for subjects in SES-BP.
    - For "Subjects expected at visit", percentages will be calculated relative to the number of subjects randomized for visits in Cycle 1 and relative to the number of subjects treated in Cycle 2 for visits in Cycle 2.
    - For "Visit performed" and "Visit not performed", absolute and percent frequencies will be provided. Percentages of columns will be calculated relative to the number of subjects expected at the respective visit.

      In case of premature study discontinuation, eCRF question 'Visit was performed' will automatically be set to 'No' in the eCRF for all visits after premature discontinuation of study participation (except for V11 among subjects in the SES-BP). Therefore, for 'Visit not performed' this question will not be used. The number

of subjects with 'Visit not performed' is calculated as number of "Subjects expected at visit" – number of subjects who performed visit.

- For "Study prematurely discontinued", cumulative absolute frequencies will be provided per visit. Absolute frequencies[18] will be determined as numbers of subjects who discontinued study participation before or on the respective visit, but after the preceding visit. For cumulative absolute frequencies, the number of subjects who discontinued the study before or at the respective visit will be displayed.

All performed and all not performed visits until regular study completion or premature study discontinuation (as applicable for the respective subject) will be listed by subject. Flags will also be included for visits performed virtually, visits not performed/ not performed due to COVID-19 pandemic (if applicable).

- Absolute und percent frequencies of subjects excluded from analysis sets will be reported by exclusion reason based on all randomized subjects.
- Demographic data will be summarized descriptively for SES-UP and FAS-UP.

  BMI will be calculated as follows:
  BMI [kg/m²] = weight [kg] / (height [cm] * 100)² (rounded to two decimal places)
- Variables on history of ET will be summarized descriptively for SES-UP.

  Duration since age of onset of ET [years] will be calculated as age of onset of ET estimated by patient – age as recorded at enrollment, i.e., at Screening (V1)

- Absolute and percent frequencies of previous and concomitant medication will be given on the basis of second and third ATC code levels (Anatomical Therapeutic Chemical classification system of the World Health Organization; WHO Collaborating Centre for Drug Statistics Methodology, ATC classification index with defined daily doses, 2023.) as well as by generic name for the SES-UP. Indications for previous and concomitant medication will not be coded and will only be listed. Separation of previous from concomitant medication is described in Appendix 8.3.
- Medical history, concomitant diseases, previous non-drug treatments (procedures) and concomitant non-drug treatments (procedures) will be coded using the Medical Dictionary for Regulatory Activities (MedDRA, version 26.1) and for SES-UP displayed in frequency tables with respect to SOC and PT. Separation of medical history and concomitant diseases and TEAEs is described in Appendix 8.4. Separation of previous from concomitant non-drug treatments (procedures) is described in Appendix 8.3.
- ECG data will only be listed.
- Injection cycles length and total observation period will be summarized using descriptive statistics for metric variables. Length of Cycle 1 and total observation period will be analyzed based on the SES-UP, length of Cycle 2 based on SES-BP.

  The length of an injection cycle and the total observation period will be determined as follows:

  The date of injection will be regarded as start of a treatment cycle and the day before the next injection will be regarded as end of cycle. For a discontinued last cycle, the

[18] not to be displayed, only for calculation of cumulative absolute frequencies

maximum of (date of last contact, last visit and last AE start date) will be regarded as end of cycle.

The length of an injection cycle and the total observation period in weeks will be calculated as (end date – start date + 1)/7 (rounded to one decimal), which means:

- The total observation period in weeks over the entire study will be calculated as (maximum of date of last contact, last visit and last AE start date - date of first injection +1 day)/7.
- The injection cycle length of cycle (x) in weeks will be calculated as (Date of injection in cycle (x+1) - date of injection in cycle (x))/7

  or (whenever the afore formula cannot be applied) as (Maximum of [date of last contact, last visit and last AE start date] – date of injection of last or discontinued cycle (x) +1 day)/7.

- Injection guidance technique by muscle will be analyzed using frequency tables, including categories "EMG", "Electrical stimulation", "Ultrasound" and all combinations which occurred For calculation of percent frequencies, the number of subjects in the applicable analysis set treated at respective muscle for injection guidance technique at that muscle will be used as denominator (i.e., the sum of absolute frequencies over the first three afore-listed categories). The denominator is defined as number participants in in analysis set by treatment group/ in total treated at respective muscle. It will be displayed in the table as "Muscle treated".
  - For Cycle 1, the analysis will be performed by treatment group on the SES-UP for the treated UL, irrespective of whether the motor dominant UL (acc. to CSP) or the non-dominant UL was treated.
  - For Cycle 2, the analysis will be performed by motor dominant / non-dominant UL on the SES-BP. In case, deviating from CSP, not all subjects in SES-BP were treated at both ULs, the analyses will be performed on the respectively treated subset of subjects only and an explanation (including the following details on affected subjects: Subject ID, treatment group, whether motor dominant non-dominant UL was not treated) will be provided in a footnote.

4.6.2 Other Data

Important PDs and final reasons for exclusion of subjects from analysis sets were be determined in the data review meeting (DRM) performed on 17Jan2024 prior to database close and unblinding. Classified important PDs as well as classified reasons for exclusion of subjects from analysis sets will be analyzed by absolute and percent frequencies over all randomized subjects.

Important PDs and reasons for exclusions from any analysis sets will be displayed in separate by subject listings, displaying the respective class of PD/exclusion reasons and further details. The listing of important PDs will also display information on relatedness to COVID-19 pandemic [yes/no] and the assessment as major or minor according to decisions made during the blind data review meeting (statistical judgement).

To describe the impact of COVID-19 pandemic on study conduct, an additional listing will display all (important or non-important) PDs and other events related to COVID-19 pandemic not considered as PDs (including, e.g., premature discontinuations due to COVID-19 pandemic and visits performed virtually due to COVID-19 pandemic).

4.6.3 Subgroup Analyses

Treatment effects by pooled study site will be analyzed for the primary and the two key secondary efficacy endpoints by use of an ANCOVA model with a treatment-by-pooled site interaction term. The respective analyses are specified as exploratory sensitivity analyses (see Section 4.1.1.2 for the primary efficacy endpoint, and Section 4.1.2.1 for the two key secondary efficacy endpoints). No further subgroup analyses are planned for this study.

4.7 Special Statistical / Analytical Issues

4.7.1 Interim Analyses

No interim analysis will be performed in this study.

4.7.2 Multiplicity Adjustment

[redacted]

4.7.3 Pooling of Sites

In case of convergence issues in ANCOVA models, study sites might be pooled for statistical analysis. During the data review meeting performed on 17Jan2024 under double blind conditions, pooling of study sites was only considered necessary for sensitivity analysis no. 2 of the primary and the two key secondary efficacy endpoints (see Sections 4.1.1.2 and 4.1.2.1 for details). No need for further pooling was seen during the DRM.

4.7.4 Premature discontinuations of study participation

Premature discontinuation of study participation and main reason for discontinuation will be analyzed descriptively (see Section 4.6.1). If premature discontinuation is related to a public health emergency, e.g. due to a COVID-19 outbreak, this information will be documented and listed. Incidence of TEAEs leading to study discontinuation (in total and related to treatment), if any, will be reported as part of the overall summary on TEAEs; details of respective TEAEs will be listed by subject (see Section 4.5.3). Discontinued subjects will generally not be replaced.

For subjects discontinuing the study prematurely and for whom EOS visit V11 was documented,

efficacy endpoints based on TETRAS ADL or TETRAS Performance assessments or kinematic tremor assessment with TremorTek performed at V11 will be analyzed under the study visit which is closest to the recorded V11 date. This will be based on the assessment date relative to study / Cycle 2 baseline, and considering treatment received for each cycle, using adapted visit windows as displayed in Table 2.

Table 2: Adapted Visit Windows for Assignment of V11 to Analysis Visit for subjects prematurely discontinuing the study

Analysis visit	Window Start Day	Window End Day	Window(s) applicable for
Week 4 (V3)	2	35	all subjects[1]
Week 6 (V4)	36	49	
Week 8 (V5)	50	70	
Week 12 (V6)	71	105	
Week 18 (V7)	108	147	
Week 24 (V8)	148	Open end	subjects[1] not treated in Cycle 2
Week 24 (V8)	148	Day of treatment in C2	subjects[1] treated in Cycle 2
Week 4 (V9)	C2 / Day 2	C2 / Day 35	
Week 6 (V10)	C2 / Day 36	C2 / Day 63	
Week 12 (V11)	C2 / Day 64	Open end	

[1] subjects prematurely discontinuing the study

Note that V11 EOS data will never be assigned to any (cycle) baseline visit based on the above rules.

In general, if based on the above windows, a V11 EOS assessment would be assigned to a timepoint for which a regular visit was performed and respective TETRAS ADL/ TETRAS Performance / kinematic tremor assessments are available, then the respective data from the regularly performed visit will be used for analysis, and V11 data will only be listed.

All other data collected at EOS visit (V11) will generally be analyzed at V11.

5 Sample Size Determination

The results of a previous study [4] suggest that a treatment effect of about one standard deviation might be expected. For a 5% type I error and a randomization ratio of 2:1 a total sample size of 51 analyzable subjects is required to achieve 90% power. In order to account for uncertainty in parameter estimates and drop-outs it has been decided to randomize approximately 75 subjects.

6 Changes in the Planned Analyses

Compared to CSP version 3.0 dated 31-May-2021 and amendment 1 to CSP version 3.0 dated 26-MAY-2023, the following change to planned analyses is included in this SAP:

Sections 2.2.1.2.2 and 2.2.1.3:

- For precision, “as assessed by the investigator” was added to the definition of all other secondary endpoints and all other efficacy endpoints that are based on the TETRAS performance assessments in the bilateral treatment period. In accordance with the CSP, the IRP did not perform any assessments in Cycle 2.

Section 2.2.1.3:

- In the definition of other efficacy endpoints for the bilateral treatment period (Cycle 2) based on angular tremor amplitude assessments, the word "difference" was missing in the CSP and therefore added in part c) of the definition before "between both ULs".
- The following restriction was added to definition of other efficacy endpoints for the bilateral treatment period (Cycle 2) based on angular tremor amplitude assessments:

  "Changes from study baseline are only analyzed for motor-dominant UL because, in accordance with the visit schedules for treatment Cycles 1 and 2 in the CSP (see Tables 5 and 6), angular tremor amplitude of motor-dominant UL was only assessed in Cycle 2."

Section 3:

- In the definition of the FAS-UP, "study baseline (V2)" was changed to "study baseline" and in the definition of the FAS-BP, "Cycle 2 baseline visit (V8)" was changed to "Cycle 2 baseline".
  References to the study baseline and the Cycle 2 baseline definition in Section 4 were inserted. These changes were made to ensure consistency of definitions of FAS-UP and FAS-BP with definition of efficacy endpoints specified as changes from at study baseline or changes from Cycle 2 baseline as defined in Section 4 and to avoid unnecessary removal of subjects from these two analysis sets. E.g., the change of the FAS-UP definition implies that a subject who does not have any scores of tremor amplitude at wrist level at study baseline visit (V2) and would thus be removed from the FAS-UP according to the FAS-UP definition in the CSP, will actually be included in the FAS-UP if he/she has at least one score of tremor amplitude at wrist level at screening visit (V1).
- For consistency, "injected UL" was changed to "injected motor-dominant UL" in the definition of the FAS-UP. This is in accordance with the specification in CSP Section 8.1.1 that the motor-dominant UL is to be injected in the unilateral treatment period (Cycle 1), the specification in CSP Section 9.1.1.2.1 that angular tremor amplitude (for evaluation of the primary efficacy endpoint and some other efficacy endpoints) is to be performed "for the injected UL(s), i.e., the motor-dominant UL in the unilateral and both ULs in the bilateral treatment period" and the specification of the unilateral key secondary efficacy endpoint in Section 12.3.1.2.1 ("Change from study baseline to Week 6 in TETRAS Performance dominant UL Score […] in the unilateral treatment period as assessed by the investigator.").

Sections 4.1.1.2 & 4.1.2.1:

- For sake of preciseness, "among all treated subjects" was added after "In case more than 10% missingness on the primary efficacy endpoint occurs in either treatment group" in the specification of the condition when sensitivity analysis no 4 (based on multiple imputation) needs to be performed for the primary efficacy endpoint (Section 4.1.1.2).
- Likewise, "among all treated subjects" was added after "in case of more than 10% missingness on the respective key secondary endpoint" regarding to need to perform respective sensitivity analysis for key secondary endpoints (Section 4.1.2.1).

7 References

[1] [redacted]

[2] Elble RJ. The Essential Tremor Rating Assessment Scale. Journal of Neurology & Neuromedicine. 2016;1(4):34-8.

[3] [redacted]

[4] Merz Pharmaceuticals GmbH. MRZ60201_2094_1 Clinical Study Report: Prospective, randomized, double-blind, placebo-controlled, parallel-group study of the efficacy and safety of NT 201 in the unilateral treatment of essential tremor of the upper limb. 2017

[5] Ondo W, Hashem V, LeWitt PA, et al. Comparison of the Fahn-Tolosa-Marin Clinical Rating Scale and the Essential Tremor Rating Assessment Scale. Movement Disorders Clinical Practice. 2018;5(1):60-5.

[6] Troester AI, Pahwa R, Fields JA, Tanner CM, Lyons KE. Quality of life in Essential Tremor Questionnaire (QUEST): development and initial validation. Parkinsonism Relat Disord. 2005;11(6):367-73.