Merz Pharmaceuticals GmbH

Statistical Analysis Plan

A prospective, randomized, double-blind, multicenter study to investigate the safety and duration of effect of different NT 201 dose groups following the treatment of glabellar frown lines

Phase 2

M602011015 / NCT03806933

Version 3.0, Final

Date: 14-JAN-2021

Author: [redacted]

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/database close.

TABLE OF CONTENTS

[redacted]

[redacted]

1 LIST OF ABBREVIATIONS

ADTTE	Analysis dataset for time to event
AE	Adverse event
AESI	Adverse event of special interest
ATC	Anatomical Therapeutic Chemical classification system of the World Health Organization
BDRM	Blind data review meeting
BMI	Body mass index
BoNT	Botulinum neurotoxin
CETS	Combined endpoint treatment success
CSP	Clinical study protocol
DEM	Dose escalation meeting
eCRF	Electronic case report form
FAS	Full analysis set
FWS	Facial wrinkle scale
[redacted]	
GFL	Glabellar frown lines
IP	Investigational product
IWRS	Interactive web response system
MedDRA	Medical Dictionary for Regulatory Activities
MP	Main period
n	Number of non-missing observations
n miss	Number of missing observations
NT 201	Clostridium botulinum neurotoxin type A (150 KD, free of complexing proteins) powder for solution for injection
OLEX	Open-label extension
PDF	Portable document format
PT	Preferred term
Q1	First quartile
Q3	Third quartile
SADR	Serious adverse drug reactions
SAP	Statistical analysis plan
SAS®	Statistical Analysis System software

SDTM	Study data tabulation model
SES	Safety evaluation set
SOC	System organ class
TFLs	Tables / figures / listings
TEAE	Treatment emergent adverse event
TEAESI	Treatment emergent adverse event of special interest
TESAE	Treatment emergent serious adverse event

[redacted]

2 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 study protocol (CSP) M602011015, dated 01-OCT-2018, amendments 1 and 2 to CSP, dated 22-OCT-2019 and 30-APR-2020, respectively.

All programs will be written using Statistical Analysis System Software (SAS®) version 9.4 or higher. A preferred font size of 10 points will be used for the tables and figures in section 14, corresponding to a line size of 111 digits and a page size of 42 lines for an output in A4 landscape format. For listings, a standard font size of 10 points with the line size and page size as defined above will be used to produce the output in A4 format. Individual SAS programs will be written for all tables, figures, and listings. [redacted]

3 CLINICAL STUDY DESIGN AND OBJECTIVES

3.1 Clinical Study Design

This multicenter Phase 2 clinical study consists of a Main Period (MP) and the subsequent optional open-label extension (OLEX) Period. The Main Period has a prospective, randomized, double-blind, dose-ranging, design with two treatment stages. In stage 1, the dose groups 20 U, 50 U, and 75 U are investigated. Once all subjects of stage 1 completed Day 30±7 Follow-up Observational Visit V4, a dose escalation meeting (DEM) was held. Stage 2 is performed only if no safety concerns have been raised during DEM and investigates the dose groups 20 U and 100 U. Subjects do not cross over from stage 1 to stage 2. A total number of up to 240 subjects with moderate to severe glabellar frown lines (GFL) at maximum frown as assessed by the investigator using the 4-point Facial wrinkle scale (FWS) will be randomly assigned at 9 study sites in USA and Germany.

Study procedures will be identical in both stages. The subjects will receive intramuscular injection of NT 201 according to their respective randomized dose groups administered at the Baseline Visit V2 (Day 1) of stage 1 or stage 2. The injection at Baseline Visit V2 (Day 1) of stage 1 or stage 2 and the following Observational Visits (up to V15) comprise the MP of the study.

Additionally, subjects who have relapsed to baseline status in the MP, will have the option to receive a follow-up treatment with the market approved dose of 20 U NT 201 for their GFL [redacted]

[redacted]. This injection and the subsequent follow-up visit constitute an optional OLEX period in this study.

MP:

Stage 1 of the study

In stage 1 of the study, eligible subjects, who had a successful Screening Visit V1 (Day -3 to -14), were randomized at the Baseline Visit V2 (Day 1) to one of three dose groups of 20 U (n=30), 50 U (n=60) or 75 U NT 201 (n=60) according to the randomization ratio 1:2:2.

Once all enrolled subjects for stage 1 had their Day 30±7 Follow-up Observational Visit V4 completed, a blinded evaluation of safety data was performed in a DEM. Based on the stage 1 data, a decision to proceed with stage 2 of the study with a dose group of 100 U NT 201 was made. No safety concerns were raised.

In addition to the blinded safety data evaluation (once all subjects of stage 1 completed Day 30±7 Follow-up Observational Visit V4), an interim analysis will be performed once complete 180-day data (V9) of stage 1 are available. Data will be unblinded for sponsor staff and vendor data management and biostatistics staff only with regard to stage 1 subject's assignment to treatment groups. The blind for subjects in stage 2 will be maintained. Stage 2 subjects will not be included in the 180-day interim analysis.

An interim analysis of complete stage 1 MP data will be performed as soon as data until and including the End of MP visit V15 from stage 1 subjects are completely available (except for missingness due to premature study discontinuation or intermittently missing visits). The analysis of stage 1 MP data will serve to get further insights regarding duration of effect and occurrence of TEAEs over the entire Main Period of up to 360 days. Knowledge gained from the interim analysis at End of MP of stage 1 will be useful to decide whether it is still meaningful to motivate stage 2 subjects affected by COVID-19 (see section 7.4.7) to re-assume on-site visits depending on the number of missing visits for the primary efficacy variable. Since ensuring the safety of subjects is paramount, this analysis will be valuable to balance the potential risks of further on-site visits against benefits such as avoidance of missing data and premature study discontinuations. Investigators and subjects, site staff and blinded monitor will be kept blinded throughout for both stages until final unblinding. The blind for subjects in stage 2 will be completely maintained.

Stage 2 of the study

No safety concerns were raised during DEM. Therefore, stage 2 of the study was started. After successful screening, eligible subjects are enrolled and randomized at the Baseline Visit V2 (Day 1) to one of two dose groups of 20 U (n=30) or 100 U NT 201 (n = 60) according to randomization ratio 1:2.

[redacted]

The observation period for each subject (in stage 1 and stage 2) will be at least 180 days. Subjects who have relapsed to baseline status of their GFL based on the investigator's assessment on the FWS at maximum frown up to Day 180, should complete the End of MP study visit at Day 180. Subjects who have not relapsed to baseline status of their GFL at Day 180 (V9) will continue to be observed for up to 180 additional days. For subjects who have relapsed to baseline status of their GFL between Day 180 and Day 360 the End of MP can be performed at each of the respective Observational Visits. If no relapse to baseline status is measured, the End of MP will be performed at Day 360 (V15).

Optional OLEX:

Subjects who have relapsed to baseline status of their GFL in the MP will have the option to receive a follow-up treatment with the market approved dose of 20 U NT 201 for their GFL between Day 180 and Day 360. An OLEX Follow-Up Visit V16 will be performed 30 days after the optional injection to document the occurrence of adverse events (AEs)/ adverse events of special interest (AESIs).

Study duration for each subject (in stage 1 and stage 2) will be at least 180 days, plus the individual duration of screening (up to 14 days), plus 30 days OLEX follow-up for subjects who elect to receive an optional follow-up treatment. The overall maximum study duration will be up to 390 days in case the subject will receive an optional follow-up treatment plus the individual duration of screening (up to 14 days) for each subject.

Subjects who withdraw or are withdrawn prematurely from the study will be requested to perform the End of MP study visit. Please note that subjects who receive the optional follow-up treatment but do not attend the OLEX Follow-Up Visit V16 for any reason will not be considered as discontinued prematurely from the study.

3.2 Clinical Study Objectives

The primary objective is to assess the safety and duration of effect of different NT 201 dose groups following the treatment of GFL.

The secondary objective is to assess the efficacy of different NT 201 dose groups following the treatment of GFL.

4 DETERMINATION OF SAMPLE SIZE

[redacted]

In total, 60 subjects in the high dose groups are deemed required for the objectives of the study. For the low dose group, 30 subjects suffice in case the study stops after stage 1 as the side effects and efficacy of a dose of 20 U are well characterized in previous studies.

5 ANALYSIS SETS

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

Safety Evaluation Set (SES)

The SES is the subset of all subjects who were exposed to study medication.

Analyses based on the SES will use the actual treatment that a subject received at Day 1 (V2) to define the treatment group.

[redacted]

Full Analysis Set (FAS)

The FAS is the subset of subjects in the SES for whom any efficacy variable is available (i.e., all subjects who have a baseline and at least one post-baseline value of any efficacy variable).

[redacted]

6 VARIABLES FOR ANALYSIS

6.1 Efficacy Variables

Efficacy variables defined in the following subsections are based on

- the investigator's assessment and/or the subject's self-assessment of the status of GFL at [redacted] maximum frown according to the 4-point FWS[redacted]

[redacted]

If on-site visits due to COVID-19 restrictions are not possible, efficacy live assessments by the investigator cannot be performed and will inevitably lead to missing data. Subjects who provide their consent for self-assessment of efficacy at home will be re-instructed on how to perform their self-assessments at home during the safety call.

The *FWS*[redacted] *at maximum frown* will assess the severity of GFL[redacted]

[redacted]

[redacted]

6.1.1 *Primary Efficacy Variable(s)*

The primary efficacy variable is the duration of effect as defined by time between treatment and relapse to baseline status. Effect is defined as improvement at maximum frown as assessed by the investigator on the FWS. If no improvement is observed on the FWS at maximum frown at any point after treatment, the time will be set to 0.

Remarks:

"Time between treatment and relapse to baseline status" is assessed as the time between treatment and *first occurrence of* relapse to baseline status, and "improvement […] on the FWS" means any improvement, i.e., an improvement of at least one point on the FWS.

[redacted]

In case no improvement is observed at any point after treatment, the respective subjects will be assigned an event of relapse on the day of treatment and the corresponding time will be set to 0.

6.1.2 *Secondary Efficacy Variables*

The secondary efficacy variables are:

- Duration of effect, whereby effect is defined by a score of none (0) or mild (1) at maximum frown as assessed by the investigator according to FWS

  This variable is defined as the time between treatment and the first point in time when the score is moderate (2) or severe (3) again. If no effect is observed on the FWS at any point after treatment, the time will be set to 0.

- Duration of effect whereby effect is defined by 2-point improvement from baseline at maximum frown as assessed by the investigator according to FWS

  This variable is defined as the time between treatment and the first point in time when the improvement is less than 2 points again. If no improvement (e.g. difference to baseline <=1) is observed on the FWS at any point after treatment, the time will be set to 0.

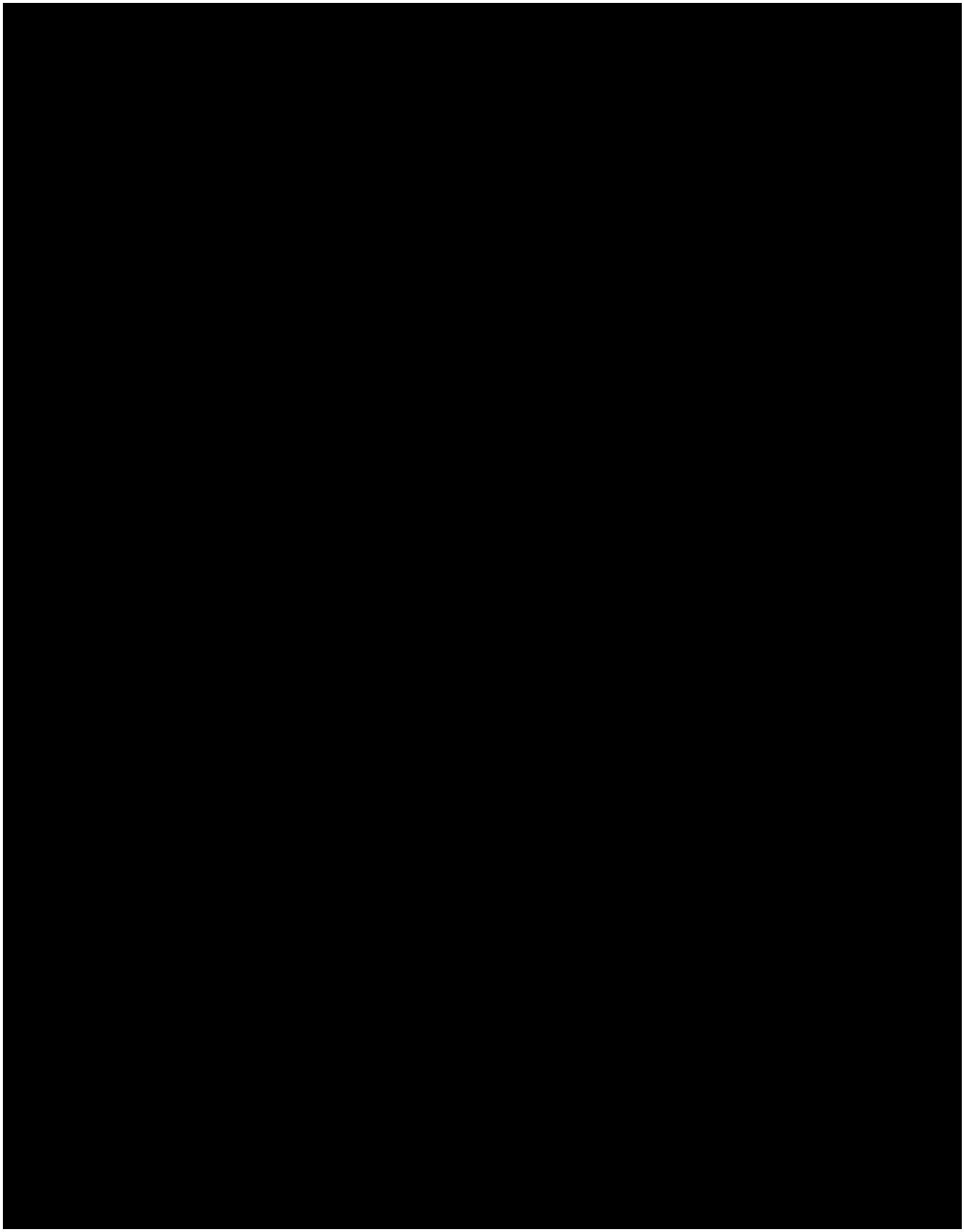
- Percentage of subjects rated as none (0) or mild (1) at maximum frown by investigator's rating on FWS at Day 180 (V9)

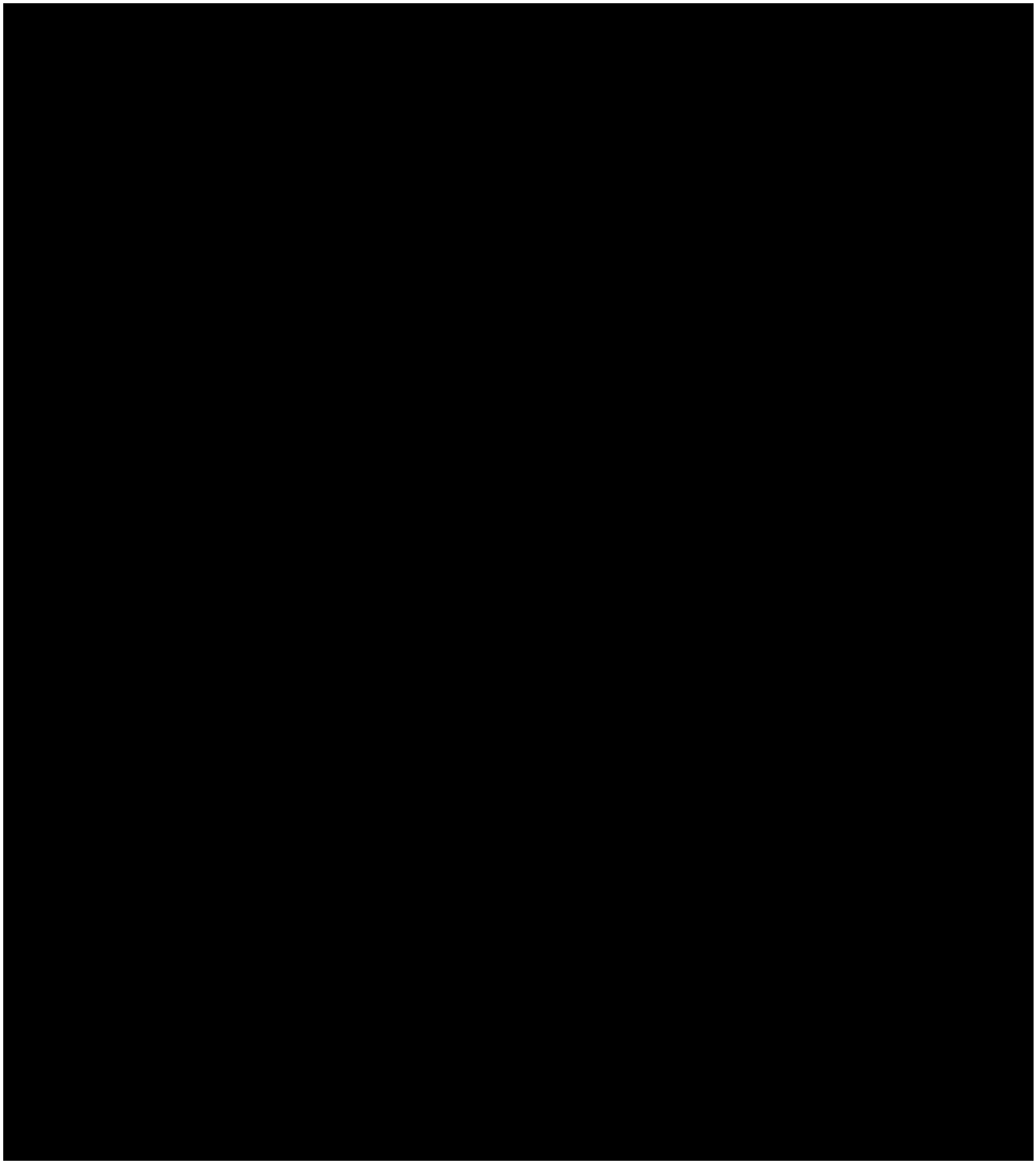
- Percentage of subjects rated as none (0) or mild (1) at maximum frown by subject's rating on FWS at Day 180 (V9)

- Percentage of subjects rated as at least 1-point improvement compared to baseline at maximum frown by investigator's rating on FWS at Day 180 (V9)

- Percentage of subjects rated as at least 1-point improvement compared to baseline at maximum frown by subject's rating on FWS Subject at Day 180 (V9)

[redacted]





6.2 Safety Variables

Per CSP version 1.0, it was foreseen to collect all safety data during on-site visits. In amendment 2 of the CSP, the following strategy to mitigate impact of COVID-19 pandemic on safety of study participants and on completeness of safety data was specified: If planned on-site visits cannot be performed due to COVID-19 situation, safety data (AEs, change in medication and non-drug treatment, and occurrence of pregnancy) will be collected via phone calls provided the subject consented. In addition, on-site visits should be performed at the earliest possible date when the subject is able to safely come to the site again.

6.2.1 *Primary Safety Variable(s)*

The primary safety variables are the occurrences of treatment emergent adverse events (TEAEs), treatment emergent serious adverse events (TESAEs), treatment emergent adverse events of special interest (TEAESIs), related TEAEs and related TESAEs by dose group for the entire study period.

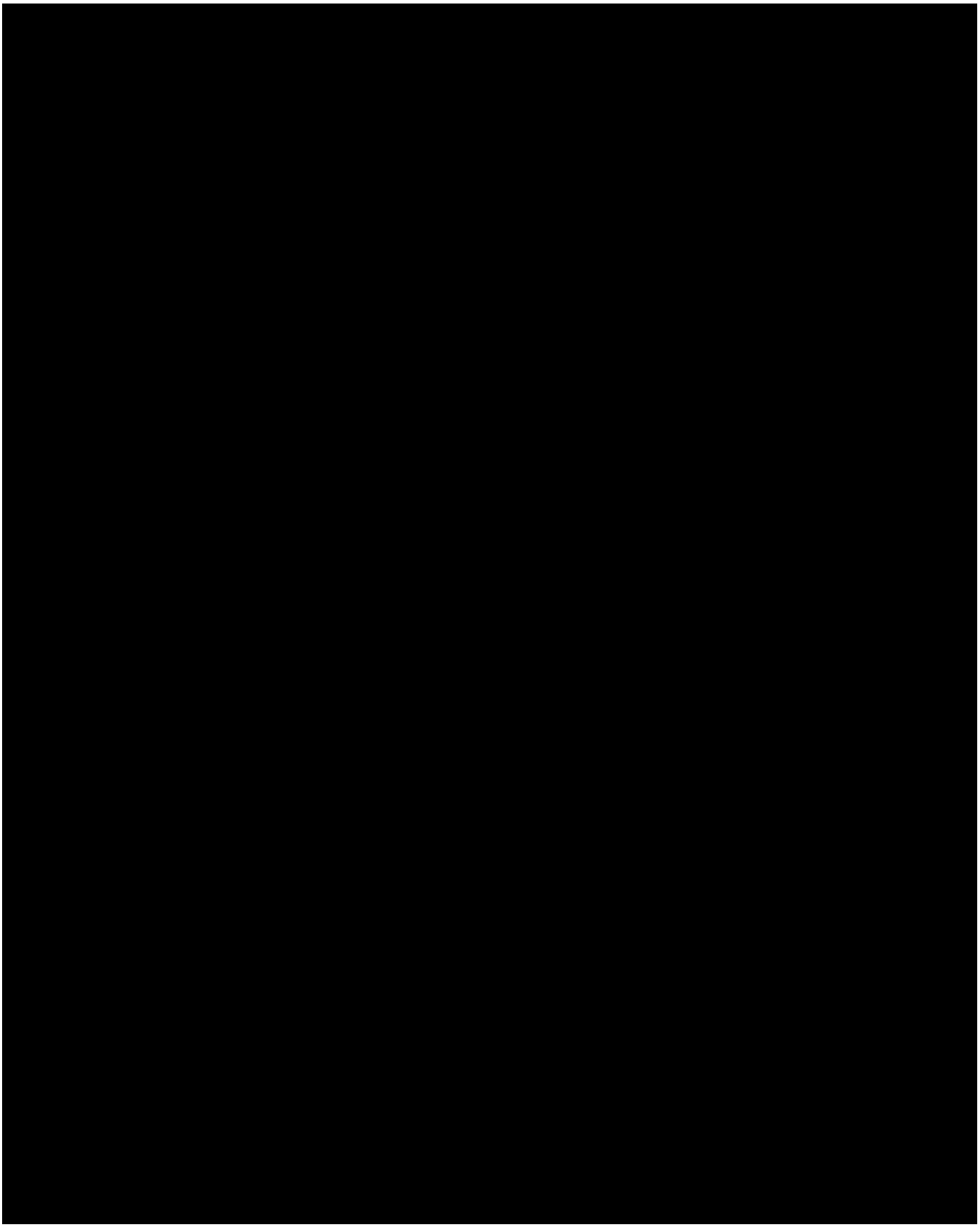
An AE is considered treatment-emergent with onset or worsening on or after date of the first administration of investigational product (IP). Each documented worsening of an AE will be counted and presented as separate AE. The start date of the AE will be the date of worsening.

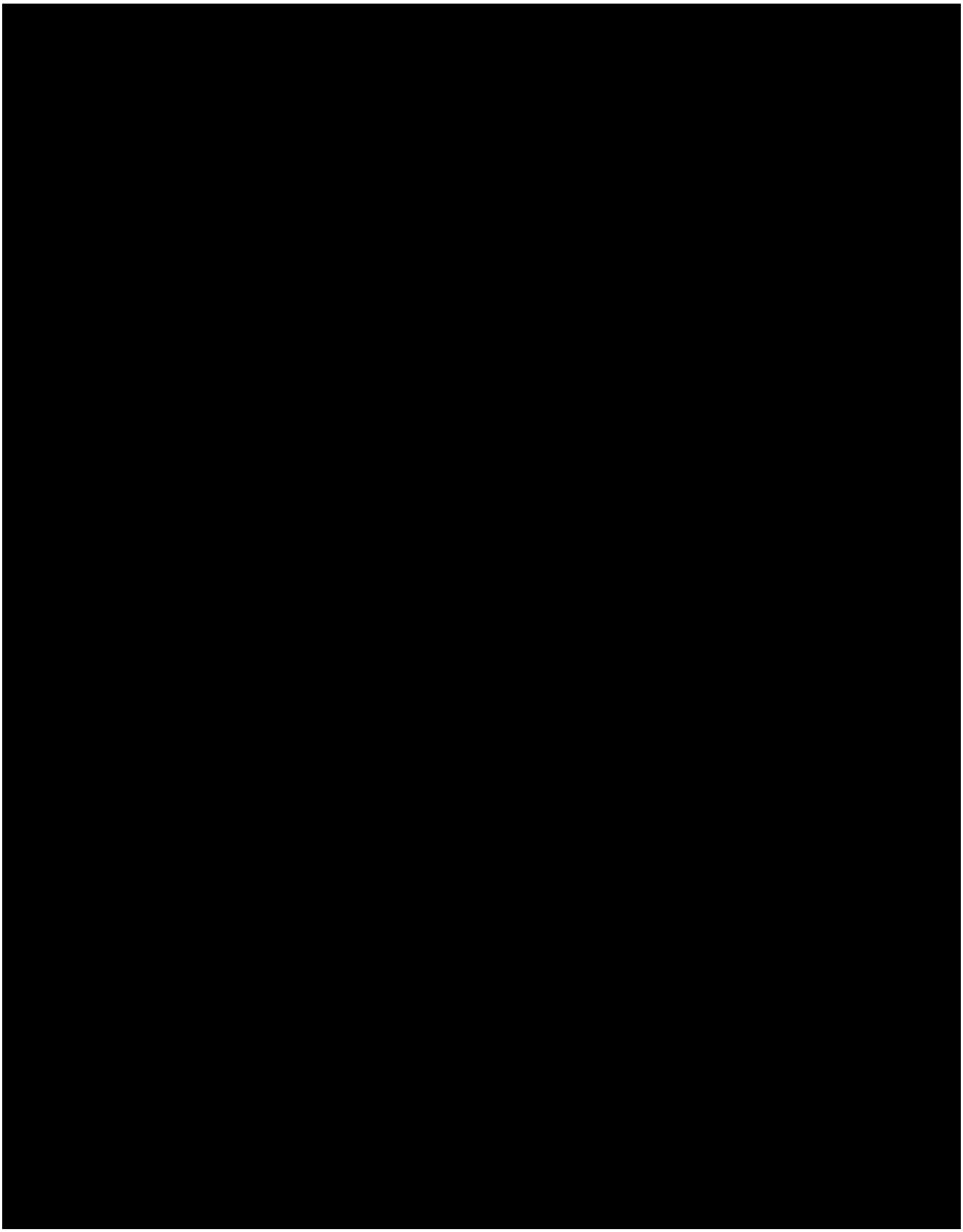
[redacted]

6.2.2 *Secondary Safety Variables*

Not applicable.

[redacted]





[redacted]

7 STATISTICAL ANALYSIS METHODS

In general for the final analysis, the dose groups of 20 U from stage 1 and stage 2 will be pooled in all efficacy and safety analyses if not described differently below. For the interim analyses of stage 1 at Month 6 and at End of MP (see section 7.4.2), subjects randomized into the 20 U group stage 1 (FAS analysis) or treated with 20 U in stage 1 (SES analysis) will be presented as dose group 20 U.

7.1 Efficacy Variables

The efficacy analyses will be based on the FAS and will be done by randomized treatment group (20 U [for final analysis pooled over stages 1 and 2 if not stated otherwise below], 50 U, 75 U, 100 U). Statistical tests will be two-sided hypothesis tests for between-treatment differences in general. Continuous variables (values and changes from baseline) will be summarized by number of values analyzed (n), mean, standard deviation, median, quartiles, minimum, and maximum. [redacted]

For qualitative variables, absolute and percent frequencies (n, %) and, if applicable, shift tables will be displayed. Confidence limits [redacted] will be given, where appropriate.

[redacted]

Mean, first quartile (Q1), median, and third quartile (Q3) 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, an adequate number of decimal places will be chosen. Percentages will be calculated using the denominator of all subjects in a specified population and treatment group. The denominator will be specified in a footnote to the tables for clarification if necessary. Percentages will be reported to one decimal place.

Efficacy analysis will be based on the MP solely, ending with the End of MP visit. (As specified in the CSP, the End of MP visit is performed between Day 180±7 and 360±7, depending on when the individual subject relapsed to baseline according to the investigator's assessment on the FWS at maximum frown.) Generally, only data from visits until regular End of MP will be used for

efficacy analyses, i.e. data from erroneously performed visits after occurrence of relapse[6] (V10 to V14 and/or from End of MP visit) will be listed only.

[redacted]

[6] based on investigator's FWS scores at maximum frown

[redacted]

[redacted]

7.1.1 *Primary Efficacy Variable(s)*

The primary variable is the duration of effect defined in section 6.1.1, with effect being defined as improvement at maximum frown as assessed by the investigator on the FWS. Duration of effect will be described by Kaplan-Meier Curves per group and the respective medians of times with associated 2-sided 95% confidence interval.

[redacted]pairwise log-rank tests will be applied to compare differences between high dose groups (50 U, 75 U and 100 U) and the 20 U group to explore statistical significance. The 20 U group data will be pooled from stage 1 and 2 for the final analysis.

For the primary endpoint, a Cox proportional hazards regression model will be applied to the data with covariates dose group (20, 50, 75, 100 U), study site[8], and baseline FWS score at maximum frown as assessed by the investigator.

The timing and censoring variables for the Kaplan-Meier analysis and the Cox regression model will be calculated[redacted]

[redacted]

[8] The term “center” as used in the description of this model in the CSP was replaced by the “study site” to increase consistency with other parts of the CSP and throughout this SAP. The meaning remains the same.

For the Cox regression analysis in the final analysis, the pooled 20 U group will be used as reference group. Hazard ratios will be calculated for each dose group to the reference group including Wald 95% confidence intervals and two-sided p-values.

For the Cox regression analysis in the interim analysis, the 20 U group will be used as reference group. Hazard ratios will be calculated for each dose group to the reference group including Wald 95% confidence intervals and two-sided p-values.

To explore potential stage effects in the final analysis, a sensitivity analysis will use dose groups 20 (stage 1), 20 (stage 2), 50, 75 and 100 U with 20 (stage 1) as reference group. Depending on the outcome further exploration might be needed.

A further sensitivity analysis will be performed to assess the impact of the COVID-19 pandemic on the primary efficacy analysis. To this end, for each subject with missing values due to COVID-19 pandemic (i.e., in case visits not performed due to COVID-19 or performed virtually due to COVID-19 so that live assessments by investigator could not be performed or discontinuation due to COVID-19 before relapse to baseline status), it will be checked whether values are missing from one or more visits directly preceding the visit when relapse to baseline was observed. While intermittently missing values due to COVID-19 (as well as continuous missingness resulting from premature discontinuation due to COVID-19) are generally assumed to be missing at random, this specific pattern of intermittent missingness might lead to bias in assessment of duration of effect. Specifically, the time to relapse to baseline status might be overestimated. Therefore, a sensitivity analysis will be conducted with time to relapse to baseline status from affected cases being censored at the first visit of those subsequent visits with missing values that directly precede the visit when relapse to baseline was observed. By this means, any overestimation of duration of effect resulting of a possibly delayed observation of relapse to baseline will be ruled out. If the first visit of those subsequent visits with missing values that directly precede relapse was performed virtually due to COVID-19 pandemic, the actual visit date will be used for censoring. In case the respective visit was not performed due to COVID-19 pandemic, time to relapse will be censored at the nominal visit date (calculated as injection date + number of days according to visit schedule in the study protocol - 1). The censoring variable CSNR will be set to 3.

The need for further sensitivity analyses was discussed during BDRMs. During the BDRM for the final analysis performed on 7-JAN-2021, it was confirmed that all randomized subjects fulfill the

requirements to be included in the FAS. Furthermore, details on premature study discontinuations, amount, patterns and reasons for missing data on the FWS at maximum frown as assessed by the investigator and relatedness to COVID-19 pandemic as well as any identified intercurrent events were closely reviewed and discussed. It was concluded that the assumption of random missingness underlying the primary duration of effect analysis appears reasonable and that the above-specified analysis to account for a specific pattern of missingness due to COVID-19 pandemic was deemed adequate and sufficient to assess the impact of the COVID-19 pandemic on the primary efficacy analysis (see also section 7.4.7 below). No intercurrent events with relevant impact on interpretation of analysis results were identified. Thus, no additional sensitivity analyses were considered necessary.

[redacted]

7.1.2 *Secondary Efficacy Variables*

The secondary duration of effect variables (see section 6.1.2) will be analyzed in analogy to the primary efficacy variable. However, for the final analysis, the Cox regression analysis will be limited to pooling the 20 U groups of both stages.

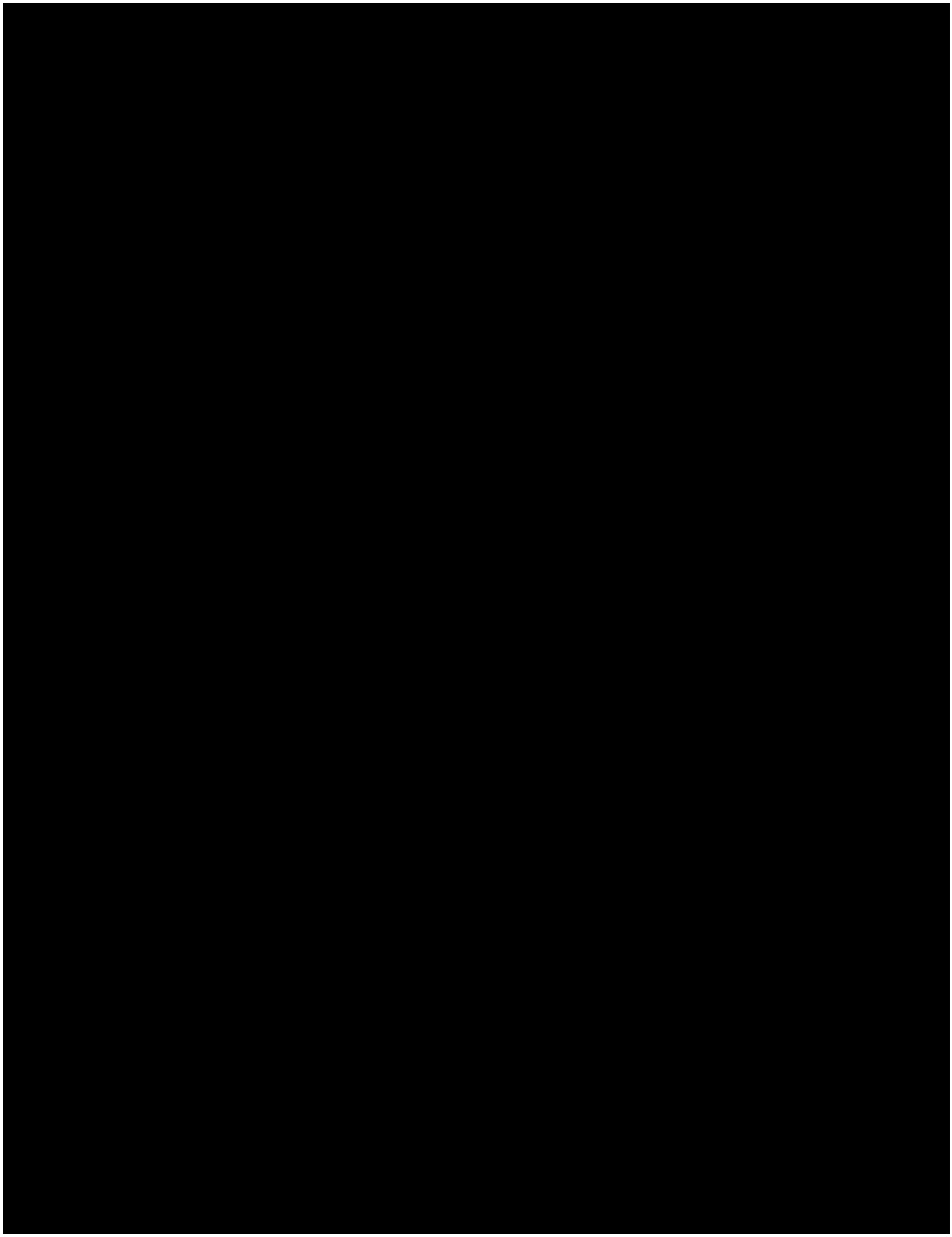
Duration of effects for each definition will be calculated as described in section 7.1.1, timing and censoring variables will be defined in analogy to the primary efficacy variable.

Kaplan Meier and Cox regression analyses will be performed as described above, except for the sensitivity analysis with separate 20 U groups from stage 1 and 2, which will be only done for the primary variable.

For the percentages of subjects fulfilling the criteria as described in the secondary efficacy variables (see section 6.1.2), two-sided 95% Pearson-Clopper confidence intervals will be computed for percentages of subjects showing an effect in each dose group.

[redacted]

[redacted]



7.2 Safety Variables

All safety analyses will be performed on the SES and will be done overall and by actual dose group 20 U [for final analysis pooled over stages 1 and 2 if not stated otherwise below], 50 U, 75 U, 100 U). Analyses of AEs will be based on MP as well as the optional OLEX period, if applicable.

AEs will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA, version 23.1) using the version and the update strategy as defined in the data management plan. Only TEAEs will be summarized, which are defined as AEs with onset or worsening after the first administration of IP. Incidences will be calculated for TEAEs on the system organ class (SOC) level and on the preferred term (PT) level (i.e., total and percentage, by intensity, and by relationship). Listings and (as applicable) tables displaying incidences for TEAEs leading to discontinuation, serious TEAEs, TEAEs of special interest, and deaths will also be provided. Additionally the number and percentage of subjects with at least one non-serious TEAE with incidence ≥ 5% will be displayed by SOC and PT (final analysis only). Time to onset and duration of TEAEs will be listed, and a by subject listing of TEAEs related to COVID-19 pandemic will be provided.

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 for AE starting after start of treatment]
- The duration will be calculated as stop date - onset date + 1.

If a subject has more than one outcome within a preferred term (PT) only the worst outcome will be used in the frequency tables. Also on subject level only the worst outcome category per subject will be counted in the frequency table. The worst outcome is defined in the following order:

- recovered/resolved

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

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".

Laboratory evaluations and vital signs (values and changes from baseline) will be analyzed descriptively by treatment and study visit. Visits as documented in the eCRF will be used, i.e. visit "End of MP" and not "V15" will be analyzed. The same applies to analysis of TEAE incidence by time interval.

7.2.1 *Primary Safety Variable(s)*

Incidences (absolute and percentage) of TEAEs, TEAESIs, TESAEs, and related TEAEs will be calculated on the level of system organ class (SOC) and preferred term (PT). In addition, incidences for TEAE will also be presented by worst intensity and by worst causal relationship. In general, incidences for TEAEs will be reported by the following actual dose groups: 20 U [pooled over stage 1 and stage 2], 50 U, 75 U, 100 U. To explore potential stage effects, incidences for TEAE by worst causal relationship will additionally be provided separately for actual 20 U dose groups from stage 1 and stage 2.[11]

The following AE summary tables will be created by actual dose group and overall (entire study period), whereby in general the dose groups 20 U pooled, 50 U, 75 U, 100 U will be used:

- TEAEs, subjects with TEAEs and number of TEAEs by SOC and PT.

[11] This analysis was added after finalization of the CSP.

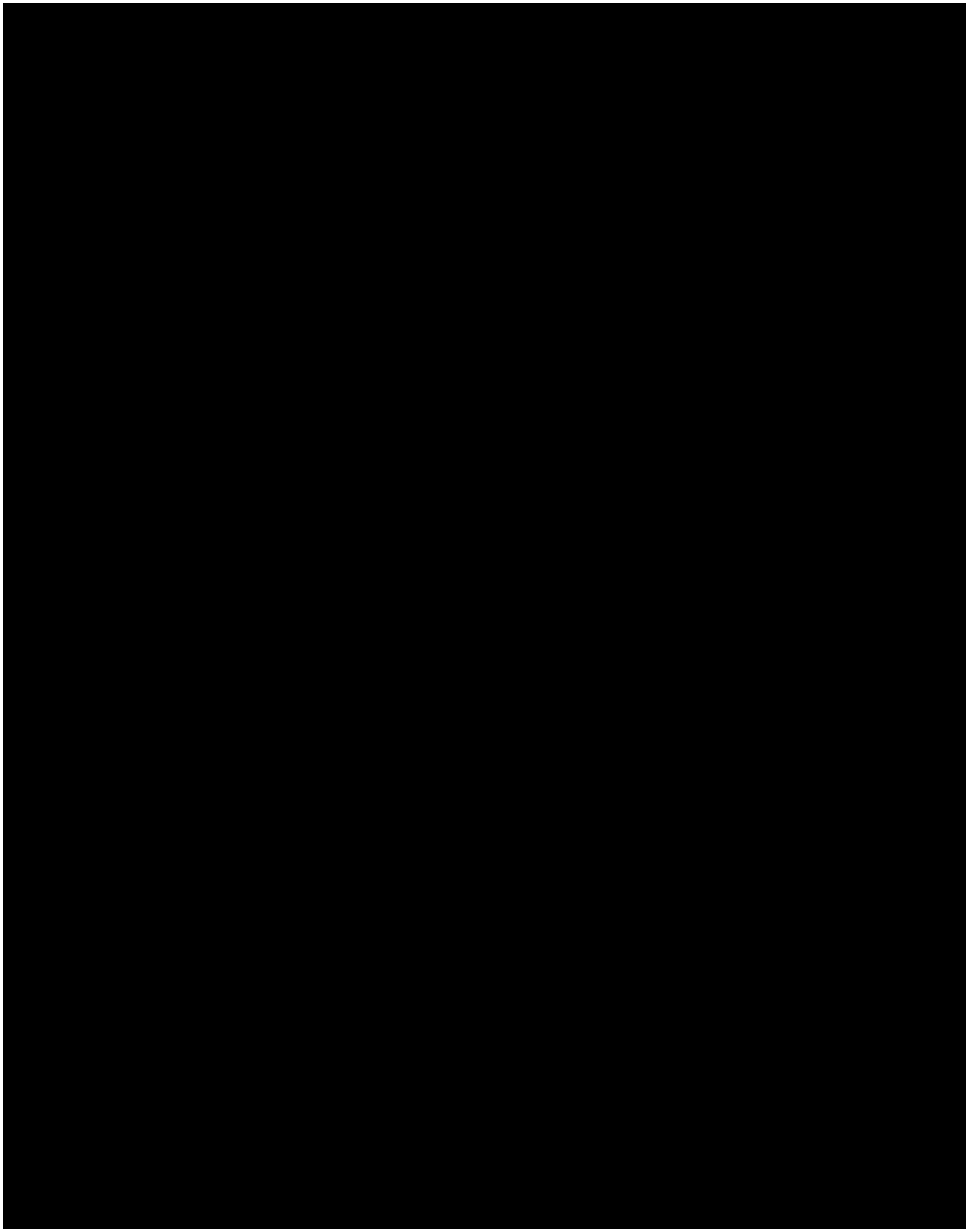
- TEAEs, subjects with TEAEs by PT.
- TEAEs, subjects with TEAEs and number of TEAEs by SOC and PT [redacted]
- Non-serious TEAEs, [redacted] by SOC and PT.
- TEAEs by worst intensity, subjects with TEAEs.
- TEAEs by worst causal relationship, subjects with TEAEs (20 U stage 1, 20 U stage 2, 20 U pooled, 50 U, 75 U, 100 U).
- Treatment related TEAEs by SOC and PT.
- Treatment related TEAEs occurring between baseline and Day 180 (V9 or End of MP), by SOC and PT.
- Treatment related TEAEs occurring between baseline and End of MP, by SOC and PT.
- TEAEs by worst outcome, subjects with TEAEs.
- TESAEs, subjects with TESAEs by SOC and PT.
- TESAEs, listing of subjects.
- Related serious TEAEs, listing of subjects.
- Deaths, listing of all serious TEAEs leading to death.
- TEAEs leading to discontinuation, listing of subjects.
- TEAESIs, subjects by SOC and PT.
- TEAESIs, listing of subjects

Related TESAEs, deaths, TEAEs leading to premature discontinuation and TEAEs related to COVID-19 will be listed only.

7.2.2 *Secondary Safety Variables*

Not applicable

[redacted]



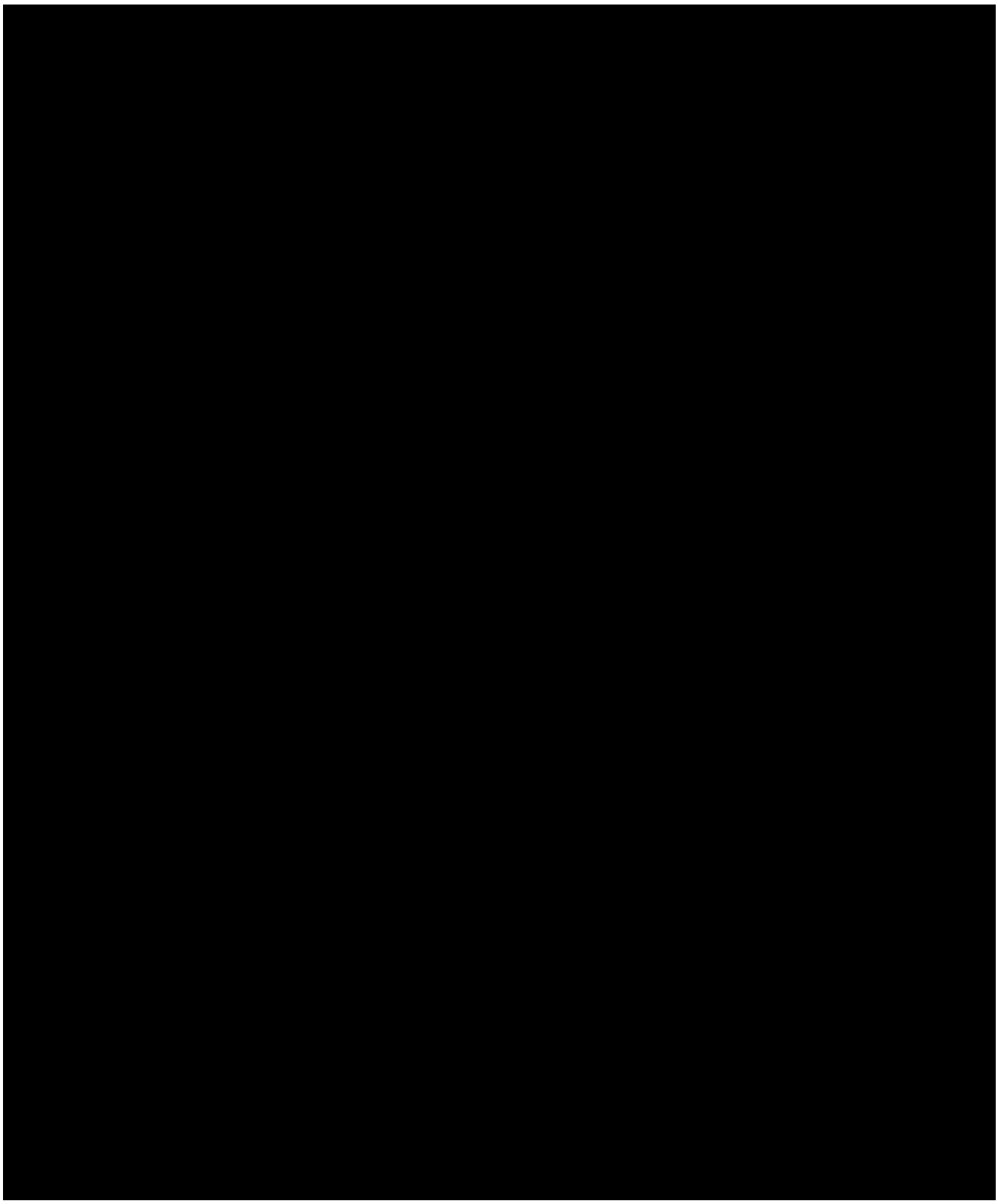
7.4 Special Statistical/Analytical Issues

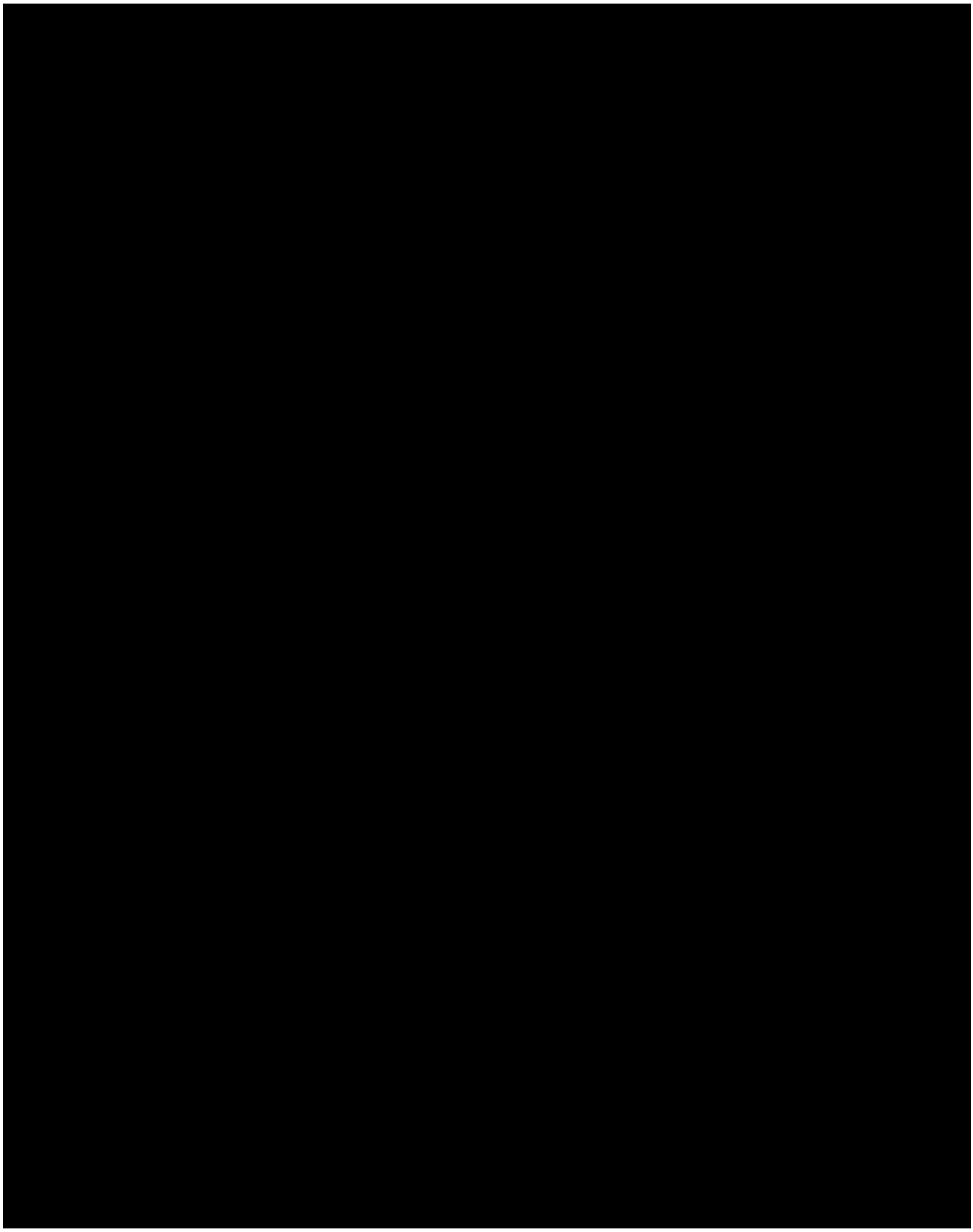
7.4.1 *Discontinuations and Missing Data*

With relevance to efficacy objectives, the treatment will be applied only once and hence cannot be discontinued. Therefore, no premature discontinuation from treatment can occur. Missing values likewise occur at random or because the subject completes the study early due to relapse to baseline. The latter will not impact the primary analysis of duration of effect since this was completely captured. Hence, for any Kaplan-Meier or Cox regression analysis and inferential statistics, intermittent missing data will be ignored and monotone missingness be considered as censored.

Intermittently missing values due to COVID-19 pandemic (resulting from missing visits not performed due to COVID-19 pandemic or visits performed virtually due to COVID-19 pandemic, thus precluding live assessments by the investigator as needed for the primary efficacy variable) as well as monotonously missing values resulting from premature discontinuation due to COVID-19 pandemic are as well considered missing at random. However, a specific pattern of missing values might introduce bias to assessment of the primary variable of duration of effect: If data are missing from visits directly preceding the visit when relapse to baseline status is observed, time to relapse to baseline status might be overestimated. A sensitivity to assess such impact on missing

values due to COVID-19 pandemic on the primary duration of effect variable is described in section 7.1.1.





Handling of missing values due to the COVID-19 pandemic

Due to the COVID-19 pandemic, a considerable amount of missing data is expected for subjects from stage 2. Restrictions related to COVID-19 in both the US and in Germany (temporarily closed sites, stay-at-home orders, quarantines imposed to subjects suspected to be infected) may make it impossible for many subjects to attend all planned on-site visits. Since live assessments by investigators cannot be performed remotely, this will inevitably lead to missing data.

Handling of missing data due to the COVID-19 pandemic in the final statistical analysis will depend on the number and pattern of missing values, which will be thoroughly described in the clinical study report.

Also, reasons for missing data will be described. As specified in CSP amendment 2, for each planned study visit that was not performed at the study site, the site staff should document whether it was performed as virtual visit due to COVID-19 pandemic, not performed due to COVID-19 pandemic (with specific reason) or not performed (due to other reasons). Also, it should be documented if a subject is discontinued from the study due to COVID-19 pandemic, again with specific reason.

During the BDRM for the final analysis performed on 7-JAN-2021 (see also section 7.1.1 above and section 7.4.7 below), the sensitivity analysis described in section 7.1.1 to account for a specific pattern of missingness due to COVID-19 pandemic was considered adequate and sufficient to assess the impact of COVID-19 pandemic on primary efficacy analysis. No additional sensitivity analyses were deemed necessary by BDRM participants.

7.4.2 *Interim Analyses*

Dose Escalation Meeting

A blinded evaluation of safety data was performed on 24th July 2019 in a DEM when all enrolled subjects for stage 1 had their Day 30±7 Follow-up Observational Visit V4 completed. Based on the stage 1 data, a decision to proceed with stage 2 of the study with a dose group of 100 U NT 201 was made. The evaluation included blinded safety data of stage 1 subjects of the 20 U, 50 U, and 75 U NT 201 dose group up to Follow-up Observational Visit V4, Day 30±7.

For this DEM the following data was presented overall for the pooled blinded data:

- Subject disposition and demographic data
- Listings of medical history, concomitant diseases, previous and concomitant treatments
- Incidences of TEAEs, TEAESIs, TESAEs, related TESAEs and related TEAEs by SOC and PT
- Listings of TEAE, TEAESIs, TESAEs, Deaths and AEs leading to premature discontinuation

The DEM received complete blinded safety data of stage 1 subjects of the 20 U, 50 U, and 75 U NT 201 dose group up to Follow-up Observational Visit V4, Day 30±7. Based on the stage 1 data, it was decided to proceed with stage 2 of the study with a dose group of 100 U NT 201.

Interim Analysis after Day 180 of stage 1

An interim analysis of complete 180-day data (V9) of stage 1 was performed. This interim analysis served to get early insights regarding the benefit-risk assessment and duration of effect of the dose groups investigated in the stage 1. However, it was not the aim of this analysis to decide on premature termination of the study.

Data was unblinded for sponsor staff and vendor data management and biostatistics staff only as required for the sake of the interim analysis.

For data of stage 1 until Day 180, the blind was not be broken until a BDRM focusing on the interim analysis had convened, the first version of the SAP had been finalized, and a database snapshot had been saved. After the blind was broken, the statistical analysis of stage 1 Day 180 results proceeded.

The study continued and investigators, subjects, site staff and monitors (except the unblinded monitor responsible for interactive web response system [IWRS] compliance) were kept blinded with regard to stage 1 subject's assignment to treatment groups. Stage 2 subjects were not included in the 180-day interim analysis. The blind for subjects in stage 2 is maintained. Moreover, the blindness of investigators and subjects, site staff and blinded monitor will be maintained throughout for both stages until final unblinding.

For the remaining data, the blind will not be broken until the final BDRM has convened, the final version of the SAP has been finalized, and the database has been closed. After the blind is broken, the overall statistical analysis of results will proceed. The SAP might be amended to stage 2 specific aspects before unblinding of stage 2.

As the blind of the subjects of stage 2 will be maintained and procedures for stage 2 will be very advanced when results are disseminated, there is no impact on validity of stage 2 data and the credibility of study results.

For this interim analysis the following data were be presented based on the data up to Day 180 of stage 1:

- Subject disposition and demographic data
- Primary and secondary duration of effect variables
  - Kaplan-Meier estimates
  - Pairwise log-rank test vs. 20 U group
  - Cox regression results with dose group, FWS score at baseline, and study site[12] as fixed effects
- Investigator's FWS and subject's FWS at maximum frown [redacted]: Summary statistics and frequency tables for scores and percentages of subjects fulfilling criteria as defined for secondary [redacted] variables including 95% confidence intervals

[redacted]

- Listings of medical history, concomitant diseases, previous and concomitant therapies [redacted]
- Incidences of TEAE, TEAESIs, TESAEs, and related TEAEs by SOC and PT
- Listings of SAEs, Deaths and AEs leading to premature discontinuation

[12] The term "center" as used in the CSP to refer to this fixed effect was replaced by "study site" to increase consistency with other parts of the CSP and throughout this SAP. The meaning remains the same.

[redacted]

For presentation of Kaplan-Meier and Cox regression analyses, the 20 U group of stage 1 served as reference group.

The interim analysis was performed including all data [redacted] up to V9 (Day 180±7) from stage 1 whereby End of MP visit was included for subjects who have relapsed to baseline status at V9 (Day 180±7) according to the investigator's assessment on the Facial Wrinkle Scale at maximum frown and completed End of MP at Day 180±7. Information on the main reason for end of study was included only for subjects who terminated the study before V9 (Day 180±7), i.e. if the End of MP Visit was performed at Day 180±7.

Interim Analysis after End of Main Period of stage 1

An interim analysis of complete stage 1 MP data was performed as soon as data until and including the End of MP visit V15 from stage 1 subjects were completely available (except for missingness due to premature study discontinuation or intermittently missing visits). The analysis of stage 1 MP data served to get further insights regarding duration of effect and occurrence of TEAEs over the entire Main Period of up to 360 days. As outlined and substantiated in more detail in the newly added section 7.4.7, this allowed for better-informed decision making during the further conduct of the study in stage 2 under the specific circumstances of the COVID-19 pandemic and related restrictions that might make it difficult or impossible for subjects to further attend on-site visits for a certain period of time.

For the interim analysis of stage 1 at Day 180 (see previous section), the blind had already been broken for sponsor staff and vendor data management and biostatistics staff with regards to all subjects randomized in stage 1. Thus, no new subjects needed to be unblinded for this additional interim analysis at End of MP of stage 1.

Data as needed for the analyses below have been cleaned. After a database snapshot had been saved, the statistical analysis of the entire MP of stage 1 were carried out.

The study continued and investigators, subjects, site staff and monitors (except the unblinded monitor responsible for IWRS compliance) were kept blinded with regards to stage 1 subject's assignment to treatment groups. Stage 2 subjects were not included in the interim analysis of stage 1 MP data. The blind for subjects in stage 2 is maintained. Moreover, the blindness of investigators and subjects, site staff and blinded monitor will be maintained throughout both stages until final unblinding.

For stage 2 subjects, the blind will not be broken until the final BDRM has convened, the final version of the SAP has been finalized, and the database has been closed. After the blind is broken, the overall statistical analysis of results will proceed. The SAP might be amended to stage 2 specific aspects before unblinding of stage 2.

As the blind of the subjects of stage 2 will be maintained and procedures for stage 2 will be very advanced when results of stage 1 MP are disseminated, there is no impact on validity of stage 2 data and the credibility of study results.

For this interim analysis the statistical evaluation were presented based on at least the following data up to End of MP of stage 1:

- Subject disposition and demographic data
- Primary duration of effect variable
  - Kaplan-Meier estimates
  - Pairwise log-rank test vs. 20 U group
  - Cox proportional hazards regression with dose group, center and baseline FWS score at maximum frown as assessed by the investigator as fixed effects
- Incidences of TEAEs, TEAESIs, TESAEs, and related TEAEs by SOC and PT

7.4.3 *Data Monitoring Committee*

Except for the internal DEM committee (see section 7.4.2), no further committee is implemented for this study.

7.4.4 *Multiple Comparisons/Multiplicity*

[redacted]no multiplicity adjustment is foreseen. [redacted]

[redacted]

7.4.6 *Pooling of sites*

Site is planned to be included as fixed factor in the Cox regression model which will be applied for primary and secondary efficacy analyses.

Any decision to pool sites (e.g. due to small or unbalanced sample sizes between sites) was taken during the first BDRM performed on 16-JAN-2020. Sample sizes were found to be well balanced across sites per country. Therefore, no pooling of sites was considered necessary for the analyses.

7.4.7 *Additional analyses due to outbreak of COVID-19 pandemic*

Due to the COVID-19 pandemic, a considerable amount of missing data was expected for subjects from stage 2 (see also section 7.4.1). This especially applies to efficacy variables that are based on live assessments by investigators, which cannot be performed remotely. The primary as well as several secondary [redacted] parameters would be affected. It should be noted that under specific circumstances, subsequently missing live ratings on the FWS by the investigator are likely to introduce bias to the assessment of the primary and secondary duration of effect variables. If relapse to baseline status was first assessed only after several intermittently missing follow-up visits due to canceled on-site visits, the time to relapse might be overestimated.

Data from Main Period of stage 1 are in contrast not affected by this or any other kind of bias caused by COVID-19 pandemic because stage 1 subjects did not miss any MP visits due to COVID-19 outbreak. Therefore, conduct of separate analyses of data from stage 1 and stage 2 subjects might be necessary in addition to pooled analyses as planned for the final analysis of the study; this decision was taken during the BDRM for the final analysis performed on 7-JAN-2021 (see below). However, for reasons outlined in the following, it appeared reasonable to perform an interim analysis of Main Period data from stage 1 as soon as MP data from all stage 1 subjects were available and cleaned.

As described in section 7.4.2 above, the additional interim analysis of complete MP data from stage 1 focused on the assessment of duration of effect of NT 201 according to the primary efficacy parameter and on occurrence of treatment emergent adverse events (TEAEs), especially treatment-related TEAEs, over the 360 Day follow-up period. Efficacy and safety data from stage 1, where doses up to 75 U were applied, can give further insights into the efficacy and safety profile of NT 201. Depending on the future development of the COVID-19 pandemic and related restrictions in both countries this was regarded as very valuable to decide whether resumption of on-site visits for stage 2 subjects with intermittently missing on-site visits still appeared meaningful and might outweigh the specific efforts and risks associated with travelling during the COVID-19 pandemic. Insights from the interim analysis at End of MP of stage 1 might be valuable to decide whether affected stage 2 subjects should be motivated to re-assume on-site visits after a certain time period with missing efficacy data according to investigator's assessments.

Any further additional statistical analyses planned to analyze the impact of missingness (see also section 7.4.1) and/or potential bias on data assessment due to COVID-19 pandemic (e.g. heterogeneity of stage 1 and stage 2 subjects due to COVID-19 related restrictions) were to be decided upon at the latest during the BDRM for the final analysis. During this BDRM, conducted on 7-JAN-2021 (see also section 7.1.1 above), details on premature study discontinuations, amount, patterns and reasons for missing data and relatedness to COVID-19 pandemic as well as any identified intercurrent events were reviewed and assessed. It was concluded that the assumption of random missingness underlying the primary duration of effect analysis still appears reasonable and that the planned analysis to account for a specific pattern of missingness due to COVID-19 pandemic (see section 7.1.1 above) is adequate and sufficient to assess the impact of the COVID-19 pandemic on the primary efficacy analysis. A separate analysis of U 20 Stage 1 and 20 U Stage 2 for the primary efficacy variable is already included in section 7.1.1. No

intercurrent events with relevant impact on interpretation of analysis results were identified. For these reasons, no additional sensitivity analyses were deemed necessary by BDRM participants.

All study subjects had their Screening visit V1 as well as their Baseline visit V2 when they received study treatment in MP before the first COVID-19 infections occurred in both Germany and the United States (US). The last subject included into this study was screened on 01-OCT-2019 and randomized on 10-OCT-2019. In Germany and in the US, the first COVID-19 cases were notified in January 2020 (Robert Koch Institute, 2020; Holshue et al. 2020). On March 11, 2020, the World Health Organization made the assessment that COVID-19 could be characterized as a pandemic. Due to this temporal sequence of events, it is evident that COVID-19 pandemic could not affect study treatment at Baseline Visit 2. Likewise, an impact of COVID-19 pandemic in terms of heterogeneity of baseline characteristics of subjects included prior to and during the pandemic can be ruled out.

9 CHANGES TO FORMER VERSIONS

This update of the SAP became necessary due to amendment 2 of the CSP, dated 30-APR-2020. The main objective of CSP amendment 2 is to assure subjects safety during the COVID-19 pandemic and to mitigate its impact on assessments of adverse events and on assessments of secondary [redacted] efficacy parameters which based on subjects' self-assessment in case subjects cannot attend on-site visits during the current situation.[14]

The following changes to version 1.0 of the SAP were made:

- Section 2: CSP amendment 2, dated 30-APR-2020, added.
- Section 3.1: New subsection on interim analysis of complete stage 1 MP data added to the subsection on "MP: Stage 1 of the study".

[14] The primary efficacy parameter is also expected to be affected by restrictions due to the COVID-19 pandemic as mitigation measures such as, e.g., remote assessments, are not possible for live assessments by the investigator as needed for this parameter. However, subjects will be asked to come for an unplanned on-site visit at the earliest possible date that the subject is able to safely come to the site again.

- Section 6.1: New paragraph added to explain which efficacy assessments can be performed remotely and which ones will be missing if on-site visits due to COVID-19.
- Section 6.2: New paragraph added to specify that and which kind of safety data will be collected via phone in case due to COVID-19 pandemic, planned on-site visits cannot be performed.
- Section 7, second sentence: “For the interim analysis” changed to “For the interim analyses of stage 1 at Month 6 and at End of MP”.
- Section 7.4.1: New subsection “Handling of missing values due to the COVID-19 pandemic” added.
- Section 7.4.2: New subsection “Interim Analysis after End of Main Period of stage 1” added.
- Section 7.4.7 Additional analyses due to outbreak of COVID-19 pandemic added.

[redacted]

The following changes to version 2.0 of the SAP were made:

[redacted]

- Sections 6.2.1, 7.2 and 7.3: MedDRA version updated.
- Section 6.2.3: Weight and body mass index added to vital sign parameters.
- Section 6.2.3 and 7.2.3: Definition of baseline for analysis of clinical biochemistry and hematology.
- Section 7.2.3: Definition of baseline for analysis of clinical biochemistry and hematology.
- Section [redacted] 7.2[redacted] New variables related to impact of COVID-19 pandemic on study conduct added.

[redacted]

- Section 7.1: More detailed specification of how visits not foreseen in the CSP, which were erroneously performed after V9 and relapse to baseline status, will be in handled efficacy analyses, for several special cases that did not yet occur when SAP version 1.0 was finalized after BDRM for the interim analysis at Month 6 of stage 1.

[redacted]

- Section 7.2: Specification that visits as documented in the eCRF will be used for TEAE incidence by time interval.
- Section 7.2.1: [redacted] Specification that TEAEs related to COVID-19 will be listed only.

- Section 7.4.1: Assumptions regarding and handling of missing values due to COVID-19 pandemic described (including reference to sensitivity analysis in section 7.1.1).
- Section 7.4.7: Statement that and explanation why COVID-19 pandemic could not affect study treatment at Baseline Visit 2 and that an impact of COVID-19 pandemic in terms of

heterogeneity of baseline characteristics of subjects included prior to and during the pandemic can be ruled out.

- Correction of minor typing errors.
- Section 7.1/ 7.1.1/ 7.4.1: Decisions made during the BDRM for the final analysis were added [redacted].
- Section 7.4.6: Conclusion from first BDRM added that no pooling of sites was necessary.

[redacted]

10 REFERENCES

Robert Koch Institute (2020). Coronavirus Disease 2019 (COVID-19) Daily Situation Report of the Robert Koch Institute 08/09/2020 - UPDATED STATUS FOR GERMANY. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Sept_2020/2020-09-08-en.pdf?__blob=publicationFile (last access: 08-SEP-2020.

Holshue, M.L. et al. (2020). First Case of 2019 Novel Coronavirus in the United States. *The New England Journal of Medicine*, 382;10; p.929-936.

Appendix [redacted]

[redacted]