Statistical Analysis Plan

Title:	A prospective, randomized, double-blind, placebo-controlled, multicenter study with an open-label extension period to investigate the efficacy and safety of NT 201 in the simultaneous treatment of upper facial lines (horizontal forehead lines, glabellar frown lines, and lateral canthal lines)
Merz Study Number:	M602011070 / NCT04622254
SAP for	Final Analysis
Sponsor:	Merz Pharmaceuticals GmbH
Version:	1.0
Date:	16-SEP-2022
Author: (Name, Function, and CRO)	PPD , Biostatistician, PPD

Document References (For Internal Use Only)
CCI [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.

PPD Author (print name), PPD	Date (dd-mmm-yyyy)	Signature
PPD Biostatistician Merz (print name) PPD	Date (dd-mmm-yyyy)	Signature
PPD Medical Expert Merz (print name)	Date (dd-mmm-yyyy)	Signature

Table of Contents


CCI

List of Abbreviations

ADaM	Analysis Data Model
AESI	Adverse Event of Special Interest
ATC	Anatomical Therapeutic Chemical classification
BMI	Body mass index
BoNT	Botulinum neurotoxin
BP	Blood pressure
CI	Confidence interval
CSP	Clinical Study Protocol
CSR	Clinical Study Report
eCRF	electronic Case Report Form
CRO	Contract Research Organization
FAS	Full Analysis Set
CCI	[redacted]
GFL	Glabellar frown lines
HFL	Horizontal forehead lines
HR	Heart rate
ICE	Intercurrent event
IP	Investigational product
IRP	Independent rater panel
KM	Kaplan-Meier
LCL	Lateral canthal lines
LS	Least squares
MAS	Merz Aesthetic Scales
MedDRA	Medical Dictionary for Regulatory Activities
MNAR	Missing not at random
MP	Main period
n	Values analyzed
n imp	number of imputed values
NT 201	Clostridium botulinum neurotoxin type A (150 KD, free of complexing proteins) powder for solution for injection
OLEX	Open-label extension
PPS	Per Protocol Set
PT	Preferred term
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System®
SDTM	Study Data Tabulation Model
SES	Safety Evaluation Set
SOC	System organ class
TEAE	Treatment-emergent adverse event
TEAESI	Treatment-emergent adverse event of special interest
TESAE	Treatment-emergent serious adverse event
TFL	Table, Figures, and Listings
Total NT 201	Group U and Group L pooled
U	Unit
UFL	Upper facial lines
WHO	World Health Organization

1 General and Technical Aspects

The objective of this statistical analysis plan (SAP) is to specify the statistical analyses with appropriate detail and precision to serve as a guideline for statistical programming and creation of tables, figures, and listings for clinical study protocol (CSP) M602011070 (version 4.0), dated 24Jun2021.

All programs will be written using SAS version 9.4 or higher. A preferred font size of 10 points, minimum font size of 9 points with a unique font size for the whole document required will be used for the tables and figures in Section 14. For listings, a standard font size of 9 points will be used to produce the output in A4 format. Individual SAS programs will be written for all tables, figures, and listings. All outputs will be transferred into PDF files. These PDF files will be generated as needed to populate the subsections of Section 14 and Section 16.2 of the clinical study report (CSR). Each output file will include the corresponding table of contents, preceding the content of the file.

The Merz standard Table, Figures, and Listings (TFLs) for drugs, version 7.0 will be applied and adapted to trial specific requirements as laid down in CSP M602011070 (version 4.0). These mock TFLs will serve as study-specific output specifications for statistical programming of TFLs. TFL headers are listed in CCI

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

2 Clinical Trial Design and Objectives

This is a prospective, randomized, double-blind, placebo-controlled, multicenter study with a placebo-controlled main period (MP) with one NT 201 injection cycle followed by an open-label extension (OLEX) period with two NT 201 injection cycles.

In the MP (Cycle 1), a total of 360 subjects will be randomized to three different treatment groups: upper facial lines (UFL) treatment (Group U), lateral canthal lines (LCL) treatment (Group L), and placebo (Group P) in a randomization ratio of 2:1:1. During the MP, subjects will be treated either with a total dose of 64 U NT 201 as simultaneous injections in all three facial areas (Group U), with a total dose of 24 U NT 201 in the LCL area and placebo in the GFL and horizontal forehead lines (HCL) areas (Group L), or with placebo in all three facial areas (Group P). All randomized subjects will then be followed up for 120 days before they may enter the OLEX period. The duration of the MP (Cycle 1) will thus be 120 days plus the duration of individual screening (up to 14 days).

The OLEX period will comprise two additional treatment cycles, Cycle 2 and Cycle 3 with durations of 120 days each plus up to 30 days for eligibility reassessments per cycle. During each cycle, eligible subjects will receive simultaneous injections of NT 201 at a total dose of 64 U in all three facial areas. Eligibility criteria will be evaluated before reinjection.

The primary objective of this study is to assess the efficacy of simultaneous intramuscular injections of NT 201 in subjects with moderate to severe UFL in comparison to placebo.

The secondary objectives are to assess efficacy and safety of simultaneous single and repeat-dose intramuscular injections of NT 201 in subjects with moderate to severe UFL.

3 Determination of Sample Size

Overall, 360 subjects are planned to be randomized to three treatment groups (UFL, LCL or Placebo) with a randomization of 2:1:1 at baseline visit V2.

The FDA required that the clinical development program includes at least 300 completed subjects who are treated simultaneously in all three anatomical areas (GFL, HFL, LCL) under placebo-controlled conditions to support safety. Since two pivotal studies are planned for demonstration of NT 201 efficacy and safety in the combined GFL, HFL and LCL treatment and assuming that up to 15% of the subjects will not complete the study, 180 subjects will be randomized per study to be treated in all three anatomical areas (Group U). A randomization ratio of 2:1:1 for the treatment groups in this study results in 90 subjects for each of the L and P Groups.

These sample sizes provide sufficient power also to assess efficacy under conservative assumptions. For proportions of 20% LCL-responders in Group U and 5% in Group P, a power of 93% is achieved for one-sided Fisher's exact test at $\alpha = 2.5\%$. Moreover, if proportions of HFL- and GFL-responders of at least 40% for Group U and at most 5% for Group P are assumed, the samples sizes provide power close to 100%.

In conclusion, when randomizing 180 subjects to Group U and 90 subjects each to Groups L and P, the study will have at least 90% power to establish efficacy of NT 201 using a hierarchical procedure for comparisons of proportions of GFL-, HFL-, and LCL-responders between Groups U and P. Assuming that about 15% of subjects will not complete the MP period, the proposed sample size will result in more than 150 completed subjects in Group U in this study. This number of completed patients in this study is considered sufficient to support the overall safety profile of NT 201 for simultaneous treatment of UFL.

4 Analysis Sets

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

Safety Evaluation Set (SES)

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

Note: In the definition of the SES above, "as treated" relates to actual study treatment in MP.

Full Analysis Set (FAS)

The FAS will consist of all randomized subjects. Subjects in FAS analyses will be analyzed as randomized.

Per Protocol Set (PPS)

The PPS is the subset of subjects in the FAS without major protocol deviations. Major protocol deviations will be defined during the blinded data review meeting.

5 Endpoints and Main Estimands for Analysis

5.1 Efficacy Endpoints and Main Efficacy Estimands

5.1.1 Primary Efficacy Endpoints and Main Primary Efficacy Estimands

The following three endpoints will be used for primary efficacy analysis during the MP:

- Proportion of GFL-responders at Day 30 (Visit 4)

  GFL-response is defined as a score of 0 (no) or 1 (mild) and at least two-grade improvement from baseline to Day 30 of MP on Merz Aesthetic Scales (MAS) for GFL at maximum contraction as assessed by both the investigator and the subject.

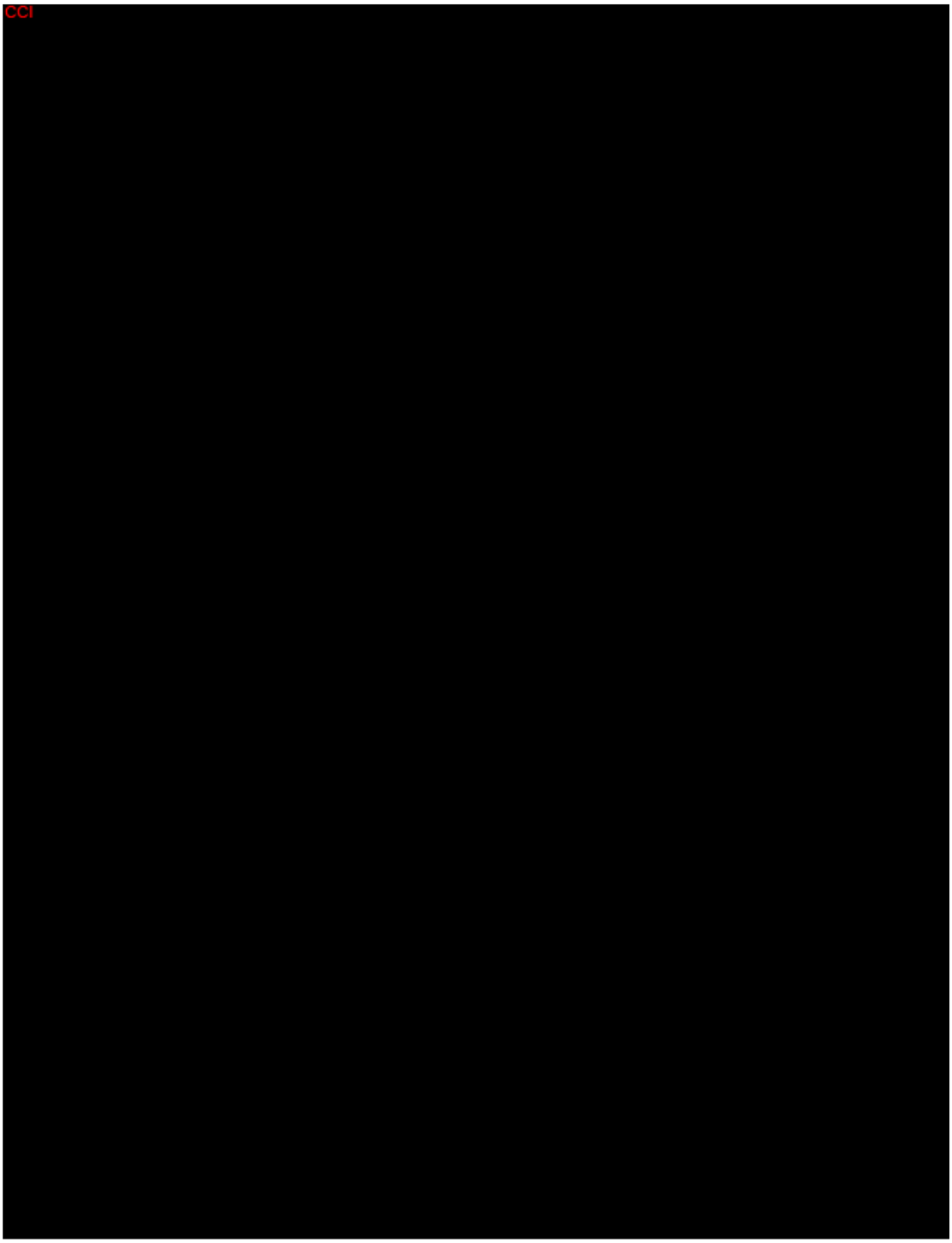
- Proportion of HFL-responders at Day 30 (Visit 4)

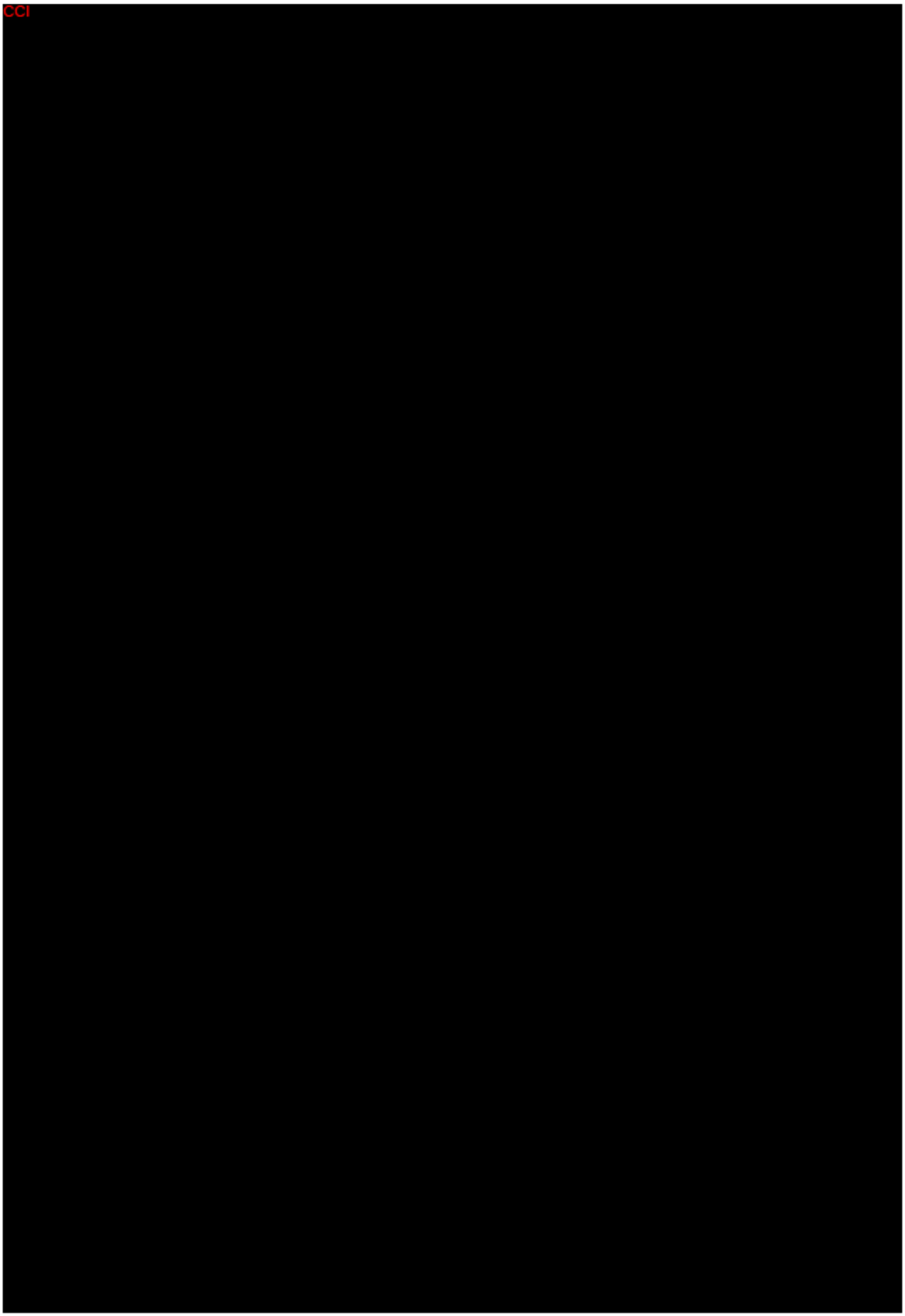
  HFL-response is defined as a score of 0 (no) or 1 (mild) and at least two-grade improvement from baseline to Day 30 of MP on MAS for HFL at maximum contraction as assessed by both the investigator and the subject.

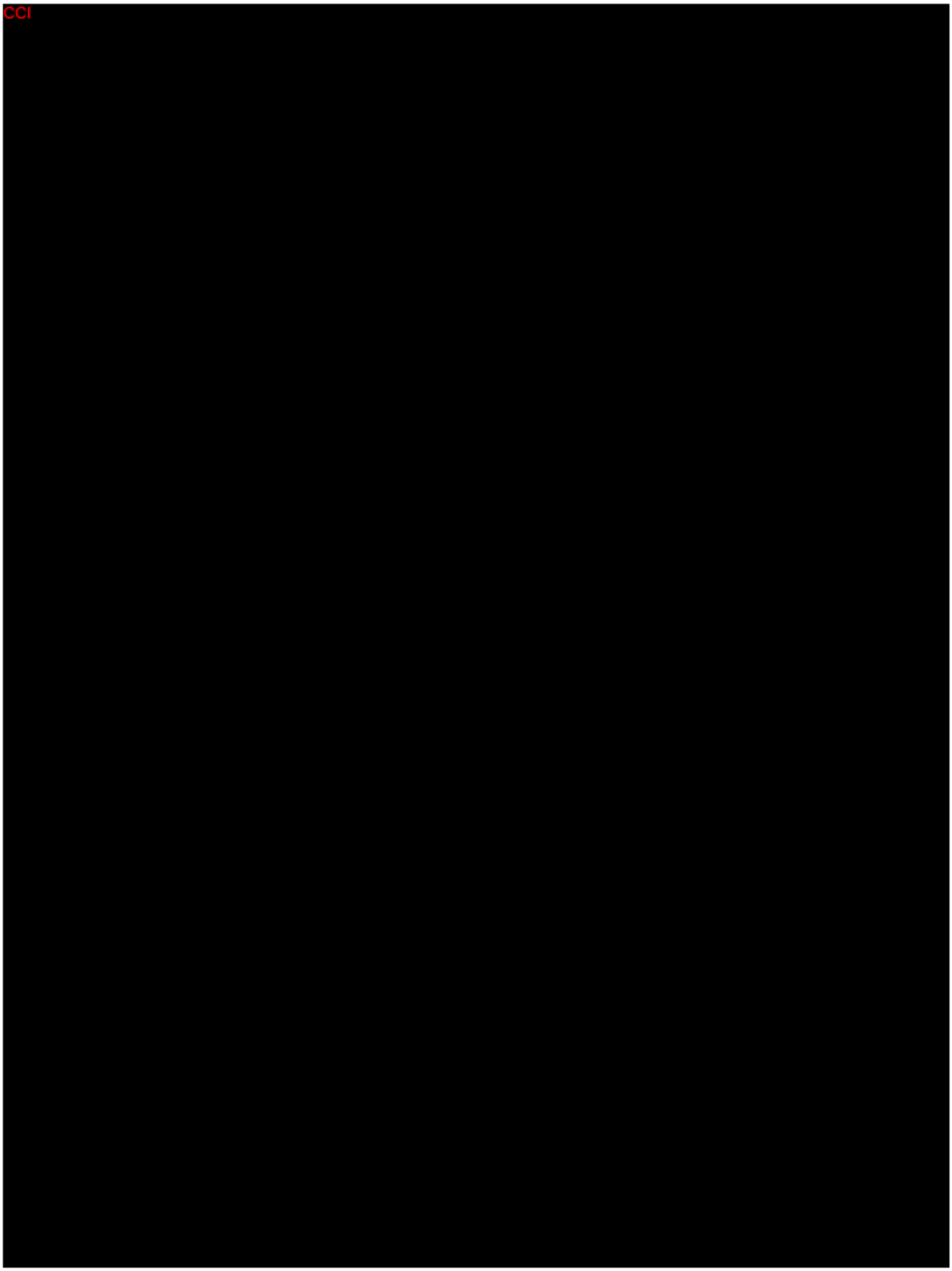
- Proportion of LCL-responders at Day 30 (Visit 4)

  LCL-response is defined as a score of 0 (no) or 1 (mild) and at least two-grade improvement from baseline to Day 30 of MP on MAS for both left and right LCL at maximum contraction as assessed by both the investigator and the subject.

Main estimands for primary efficacy endpoints (see CCI for GFL, CCI for HFL, CCI for LCL) describe the treatment effects of interest for the three treatment areas. For each main primary estimand, the variable attribute corresponds to the respective primary endpoint.


CCI


CCI


CCI

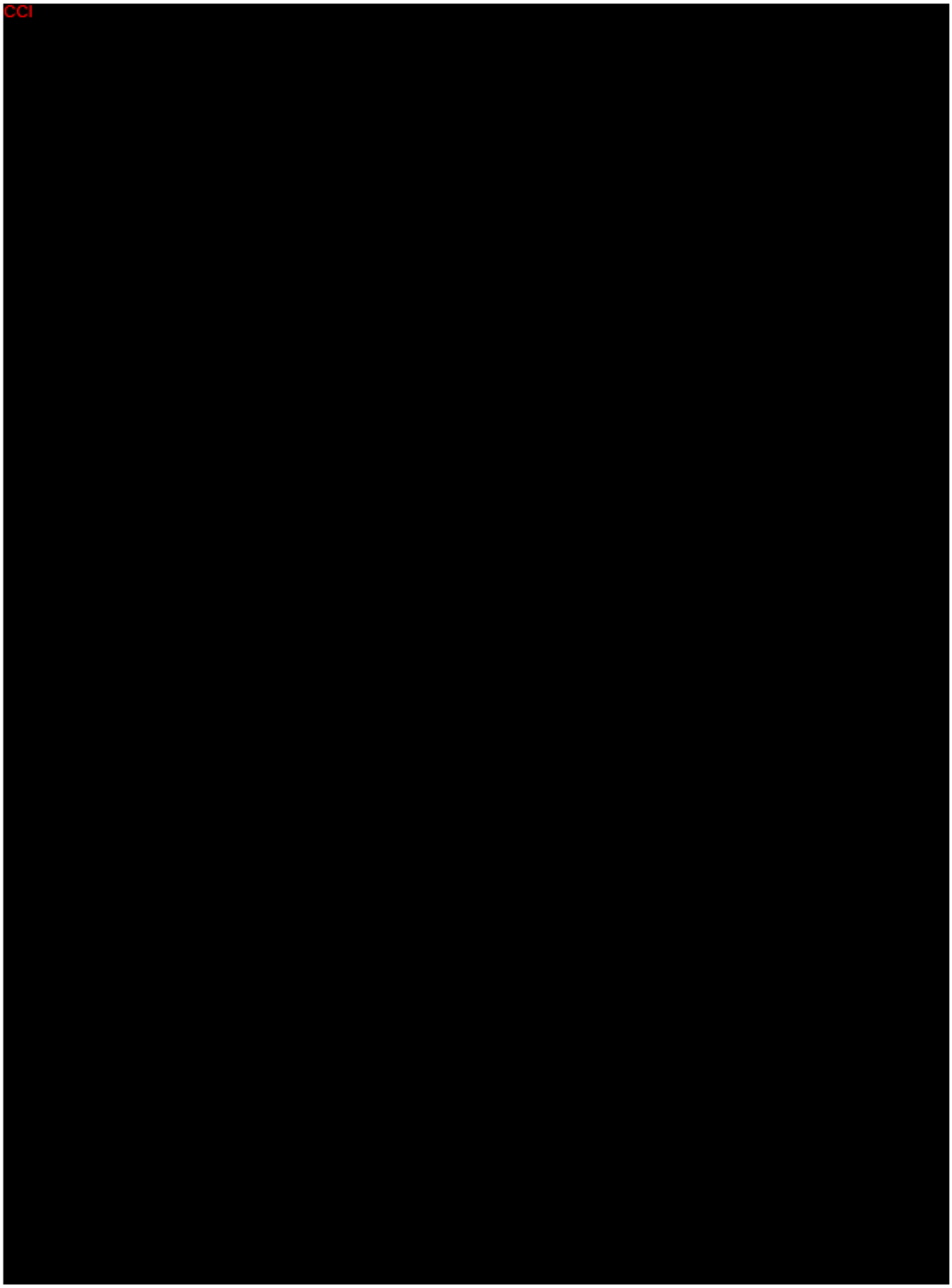
5.1.2 Secondary Efficacy Endpoints and Main Secondary Efficacy Estimands

Key secondary efficacy endpoints:

- Proportion of subjects with a score of 0 (no) or 1 (mild) on MAS for GFL at maximum contraction as assessed by the investigator at Day 30 (Visit 4) of MP.
- Proportion of subjects with a score of 0 (no) or 1 (mild) on MAS for HFL at maximum contraction as assessed by the investigator at Day 30 (Visit 4) of MP.
- Proportion of subjects with a score of 0 (no) or 1 (mild) on MAS for both left and right LCL at maximum contraction as assessed by the investigator at Day 30 (Visit 4) of MP.
- Proportion of subjects with a score of 0 (no) or 1 (mild) on MAS for GFL at maximum contraction as assessed by the subject at Day 30 (Visit 4) of MP.
- Proportion of subjects with a score of 0 (no) or 1 (mild) on MAS for HFL at maximum contraction as assessed by the subject at Day 30 (Visit 4) of MP.
- Proportion of subjects with a score of 0 (no) or 1 (mild) on MAS for both left and right LCL at maximum contraction as assessed by the subject at Day 30 (Visit 4) of MP.
- Global Aesthetic Improvement Scale (GAIS) as assessed by the subject at Day 30 (Visit 4) of MP.

Main estimands for MAS-based key secondary efficacy endpoints are specified in full detail in CCI [redacted]. All key secondary estimands per treatment area were derived from the respective main primary estimand as specified in Section 5.1.1. Definitions of attributes treatment, target population, other intercurrent events (ICEs) and population level summary are identical. The variable attributes of key secondary estimands were adapted according to the above-defined key secondary endpoints for the respective treatment area based on the investigator/ subject assessment on the MAS at maximum contraction. For example, the variable attributes of the main estimands for the two key secondary efficacy endpoints for treatment area GFL are defined as "GFL response defined as a score of 0 (no) or 1 (mild) on MAS for GFL at maximum contraction as assessed by the investigator at Day 30" and "GFL response defined as a score of 0 (no) or 1 (mild) on MAS for GFL at maximum contraction as assessed by the subject at Day 30", respectively.

The main estimand for the key secondary endpoint based on subject's GAIS at Day 30 of MP is presented in CCI [redacted]

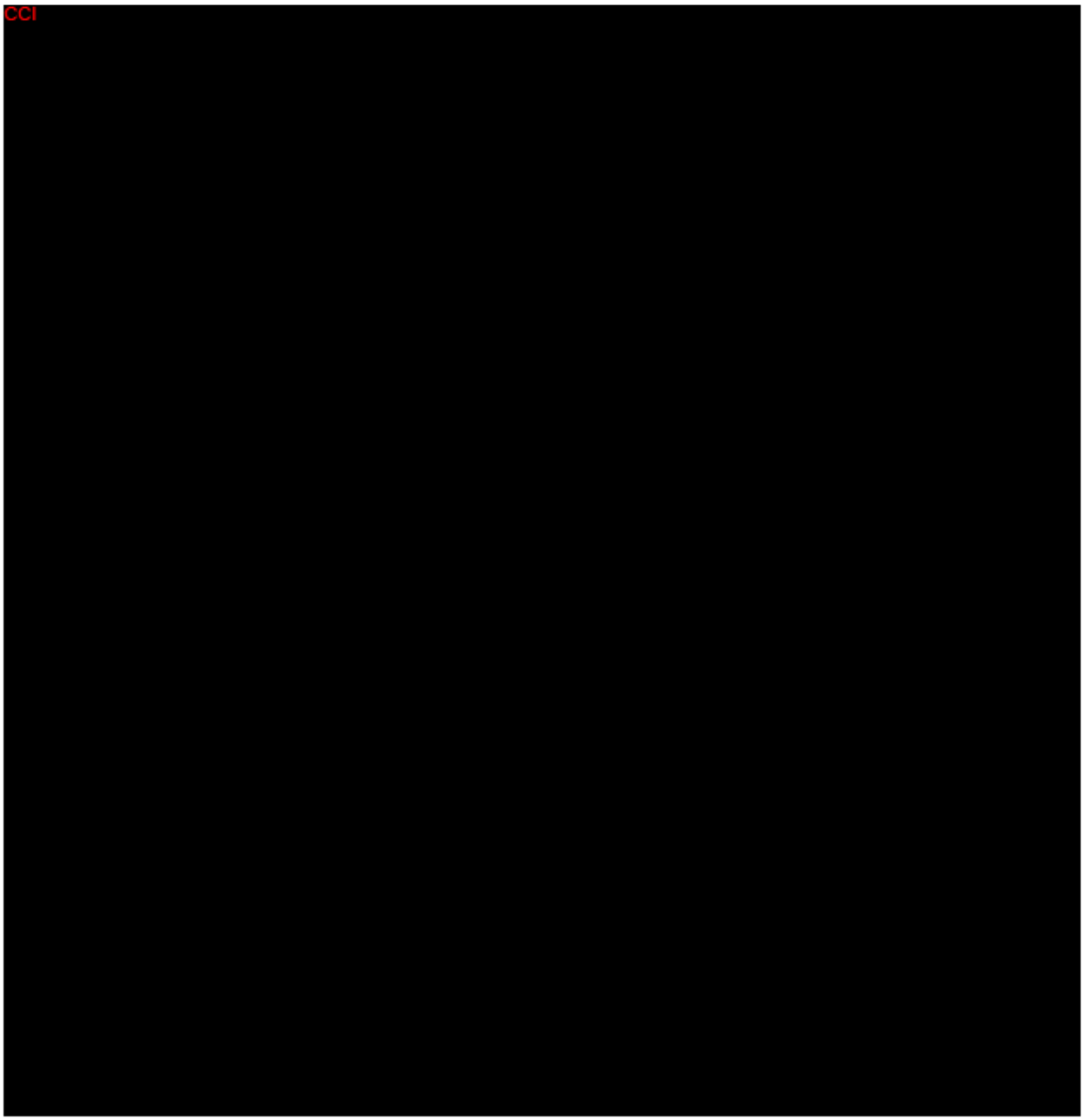

CCI

Further secondary efficacy endpoints:

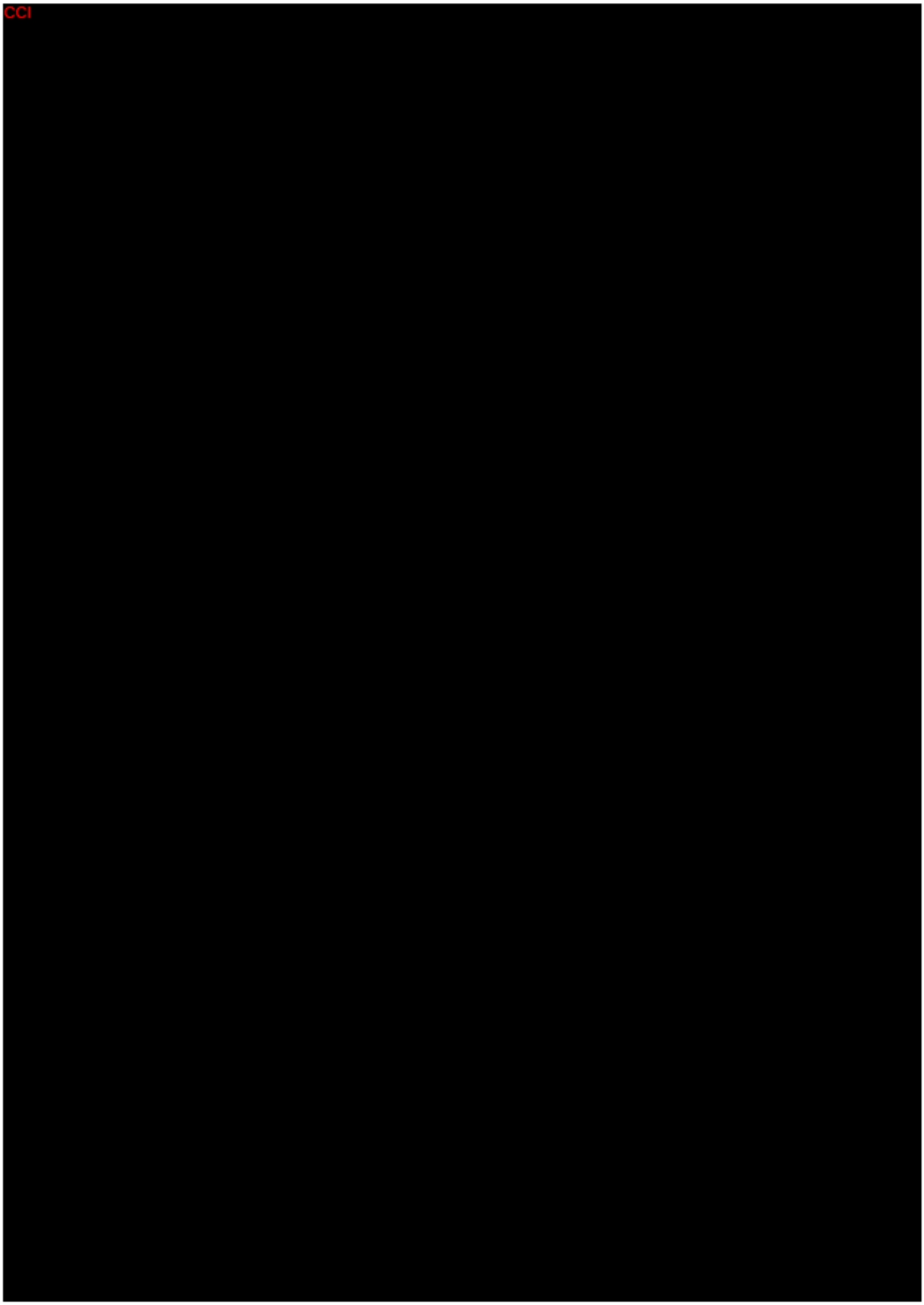
- Proportion of subjects with at least one-grade improvement from baseline to Day 30 (Visit 4) of MP on MAS for GFL at maximum contraction as assessed by the investigator.
- Proportion of subjects with at least one-grade improvement from baseline to Day 30 (Visit 4) of MP on MAS for HFL at maximum contraction as assessed by the investigator.
- Proportion of subjects with at least one-grade improvement from baseline to Day 30 (Visit 4) of MP on MAS for both left and right LCL at maximum contraction as assessed by the investigator.
- GAIS as assessed by the investigator at Day 30 (Visit 4) of MP.

No estimands are defined for further secondary endpoints.

CCI



CCI



5.2 Safety Endpoints/Variables

5.2.1 Primary Safety Endpoint

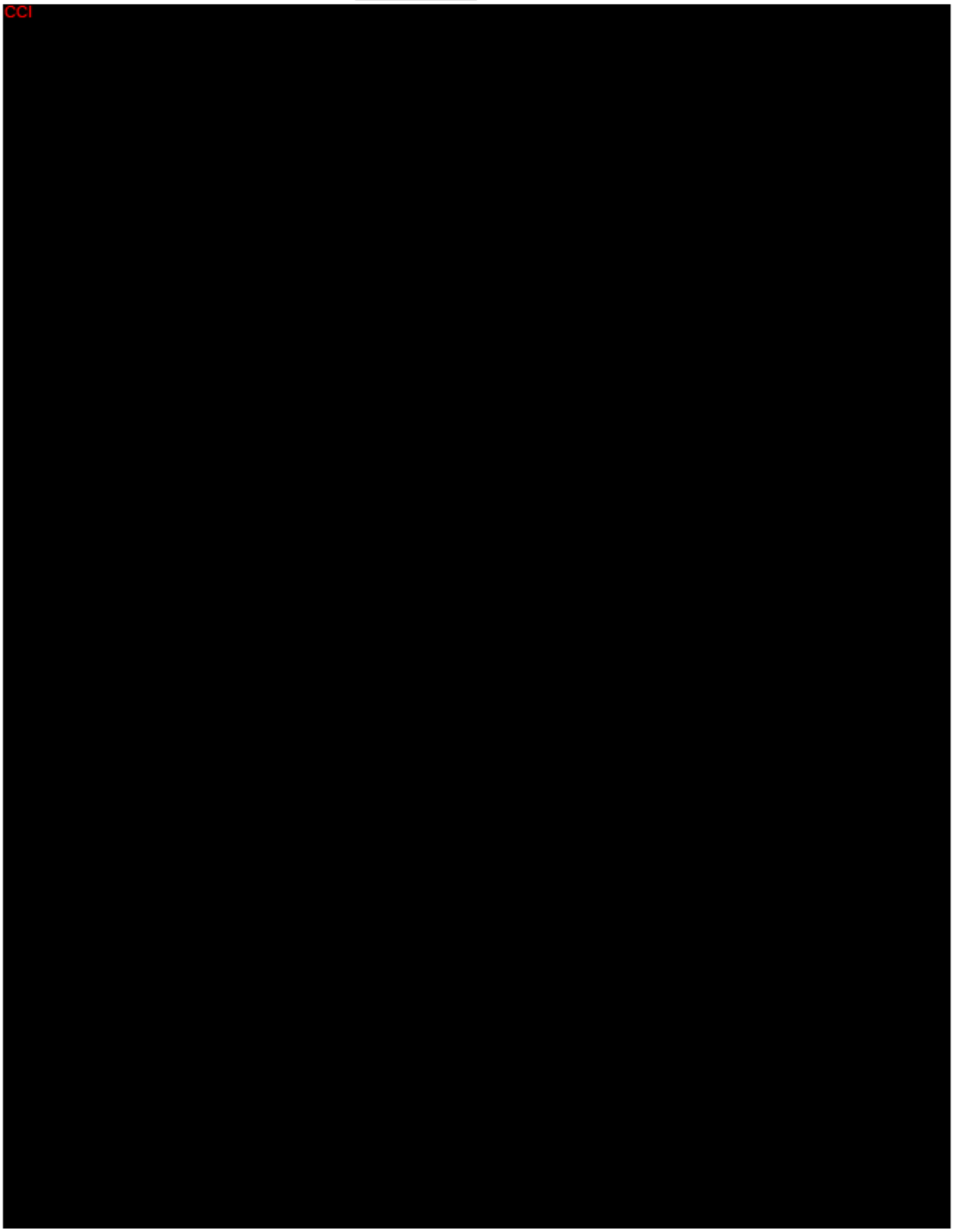
No primary safety endpoint has been defined.

5.2.2 Secondary Safety Endpoints

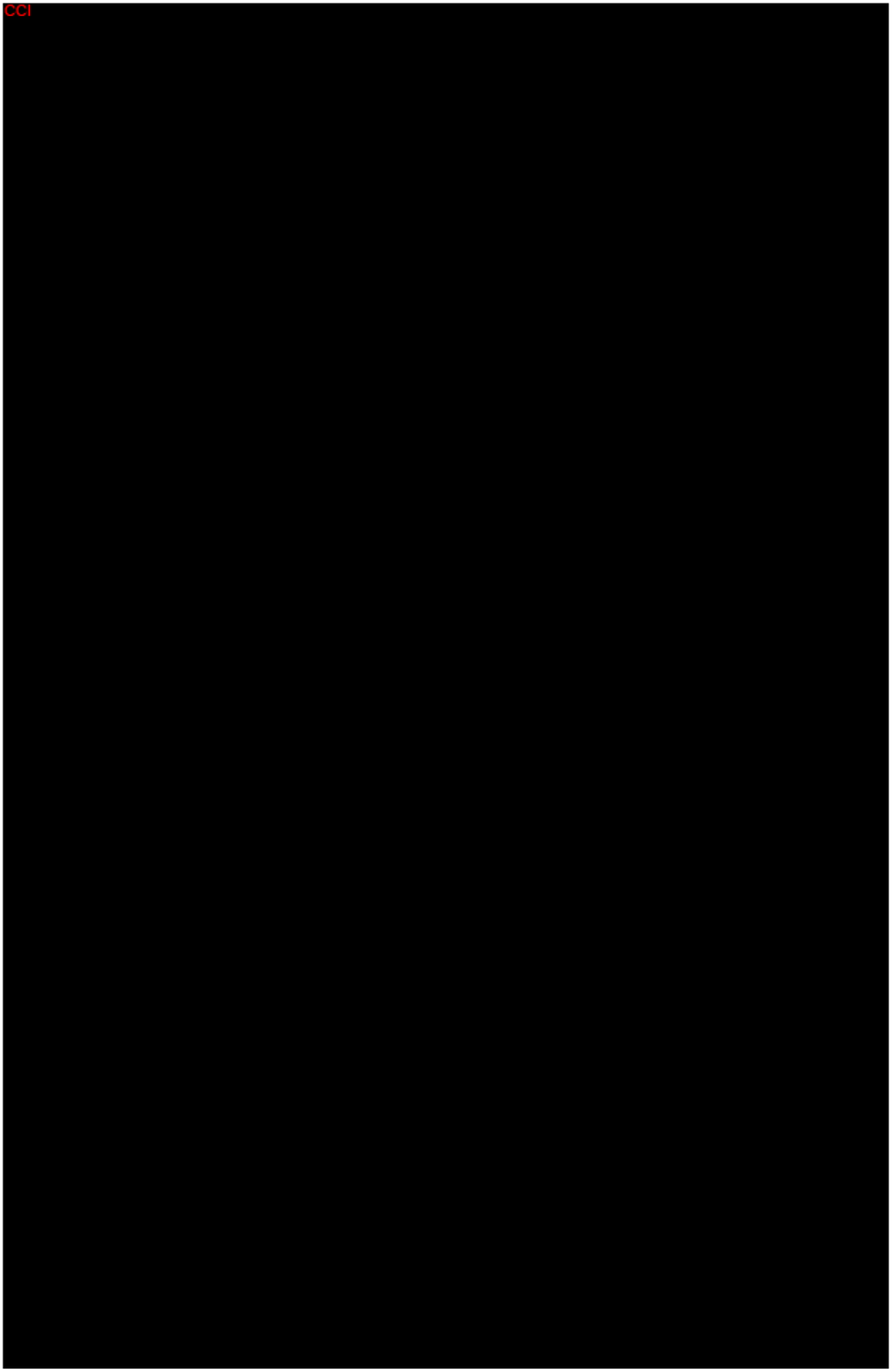
- Incidence of related treatment-emergent adverse events (TEAEs) in MP.
- Incidence of related TEAEs in OLEX period

For the definition of TEAEs, see CCI

CCI



CCI



CCI

[redacted]

6 Statistical Analysis Methods

Continuous endpoints (values and changes from baseline/ cycle baseline, where applicable) will be summarized by number of subjects in the respective analysis set/ population (N), number of values analyzed (n), mean, standard deviation, median, quartiles, minimum, and maximum. If relevant, the number of values imputed due to missingness (nimp) will be reported for efficacy variables (see Section 6.4.1). Mean, quartiles, and median will be reported to one decimal place more than the data were collected, for the standard deviation two decimal places more will be displayed; for derived data, an adequate number of decimal places will be chosen.

For qualitative endpoints, absolute and percent frequencies (n, %) and, if applicable, shift tables will be displayed. Frequency tables will include the number of missing values, if applicable. Percentages will be calculated using the denominator of all subjects in a specified population or treatment group. The denominator will be specified in a footnote to the tables for clarification if not otherwise obvious. Percentages will be reported to one decimal place. In frequency tables based on observed cases, number of missing values will not be presented and percentages will be based on number of observed cases (Nobs).

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

In general, values obtained at Visit 2 are defined as the baseline values. Where data is not recorded at Visit 2, the last measurement before start of treatment is defined as baseline value. These values will be analyzed as baseline (i.e. baseline flag is 'yes' for last record with non-missing value before first treatment in Study Data Tabulation Model [SDTM]). For cycle 2 values obtained at Visit 8 and for cycle 3 values obtained at Visit 13 will be used as cycle baselines.

6.1 Efficacy Endpoints and Efficacy Estimands

Primary and sensitivity analyses of primary efficacy endpoints (see Section 6.1.1.1) and confirmatory analyses of key secondary endpoints will be based on the target population of main primary and key secondary efficacy estimands (see Section 5.1.1 and Section 5.1.2, respectively). It is defined as "Male and female adults (18 years of age or older) with HFL, GFL, and LCL of moderate (score of 2) to severe (score of 3) intensity at maximum contraction as assessed by the investigator and the subject according to MAS, as randomized in this study: Subset of subjects randomized to UFL treatment group or to Placebo treatment group, grouped by randomized treatment assignment.". Further efficacy analyses will be based on the FAS if not otherwise specified. Statistical tests will be two-sided hypothesis tests for between treatment differences in general. Adequate descriptive statistics as specified at the beginning of Section 6 above will be provided for all endpoints.

In case Visit 7/Visit 8 or Visit 12/Visit 13 is performed on same day MAS assessments from Visit 7/Visit 12 will be duplicated to be also analysed and listed at Visit 8/Visit 13.

6.1.1 Primary Efficacy Endpoints and Estimands

The primary endpoints will be proportions of GFL-, HFL- and LCL-responders as defined in Section 5.1.1. These proportions will be compared for treatment groups U and P. It should be noted that Group L is not part of the primary efficacy analysis which has the objective to compare efficacy of simultaneous treatment of UFL with NT 201 versus placebo. Group L and Total NT 201 (Group U and Group L pooled) will be considered in the supplementary analyses of primary endpoints on the FAS, the PPS, observed cases in the FAS, observed cases in the PPS (see below), in supplementary analyses of key secondary endpoints, and in analyses of further secondary and other efficacy endpoints.

The three main primary efficacy estimands are defined in Section 5.1.1 (see CCI for GFL, CCI for HFL, CCI for LCL).

6.1.1.1 Primary analysis

The following main estimators for main primary estimands will serve as primary analysis: For comparison of proportion of GFL-, HFL- and LCL-responders in treatment groups U and P, Mantel-Haenszel stratum weights and the Sato variance estimator will be used to calculate the stratum-adjusted risk difference, with study site serving as stratum variable. For each anatomical region, the stratum-adjusted (common) difference in response rates between treatment Group U and Group P (reference) ("Group U" – "Group P") will be provided with two-sided p-value and 95% Mantel-Haenszel confidence interval (CI). The null hypothesis of the Mantel-Haenszel test is that the common difference in response rates is zero. The family-wise two-sided type I error level $\alpha = 0.05$ will be controlled by a hierarchical test procedure based on two-sided tests with $\alpha = 0.05$ as described in Section 6.4.3.

CCI [redacted]

Strategies for handling of ICEs are described in Section 6.4.6.

Missing data will be imputed so that all subjects in the target population of main primary estimands can be included in the analysis. CCI [redacted]

CCI [redacted]

CCI

Responder rates with two-sided 95% Wilson CIs will be provided as well for both treatment groups included in target population of main primary efficacy estimands, overall and per study site.

A similar SAS code to the following statements will be employed for this analysis (example for comparison of GFL responders):

CCI

[1] In the CSP, it was erroneously stated that the score from V3 (Day 8) is likely lower than the score to be expected at Day 30. This is no change to planned analyses since no change to the imputation rule is implied. Only the justification for imputing monotonously missing data at V4, V5, V6, and V7 with V3 data (conservative approach) was corrected.

CCI

[redacted]

6.1.1.3 Subgroup Analyses

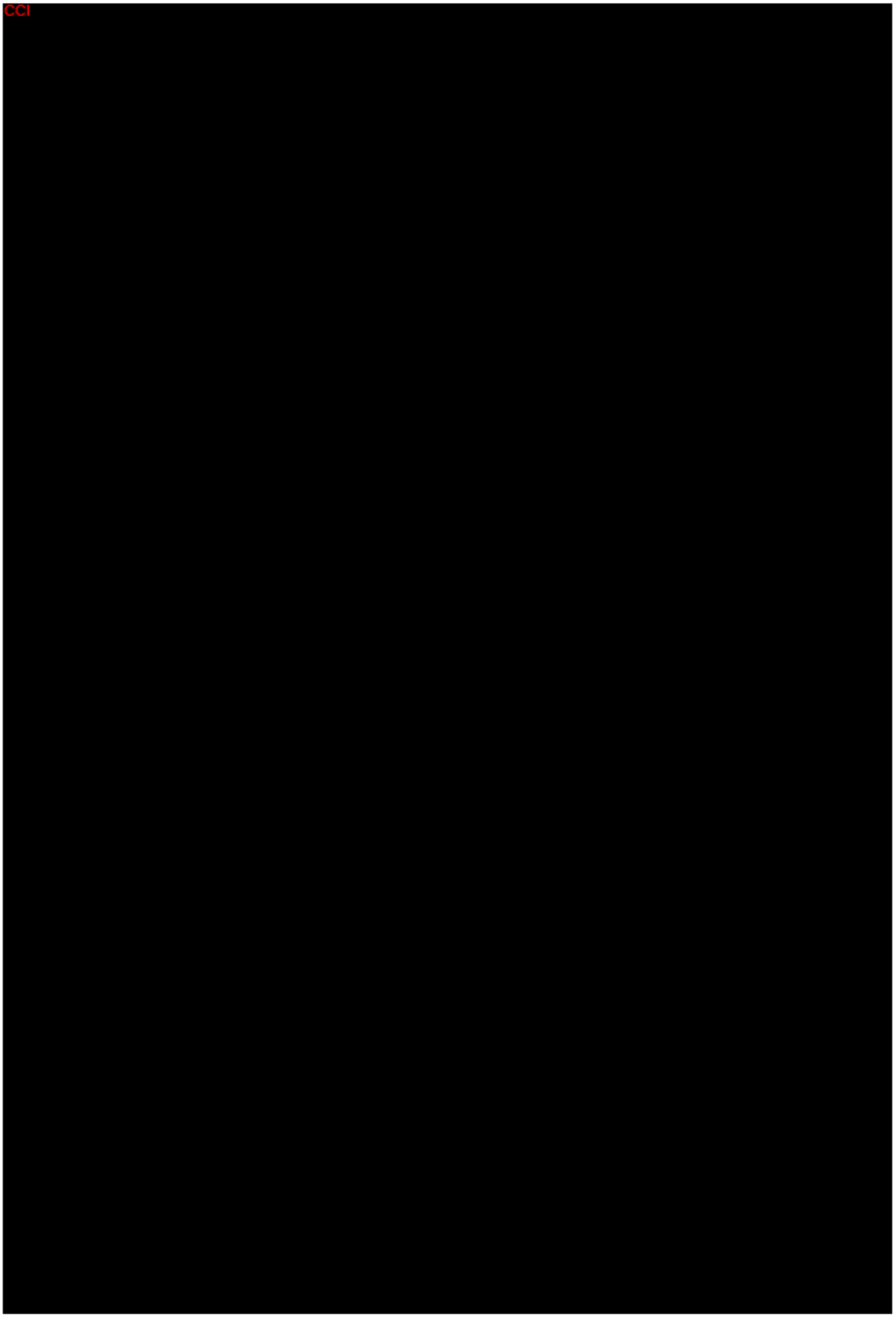
The primary analysis of primary endpoints will be conducted for two subgroups for the target population of main primary estimands. Subgroups are defined by BoNT pre-treatment status. Subjects will be distinguished on whether they have been treated with BoNT of any serotype prior to randomization at any time for any treatment indication or not (treatment naïve). Two-sided 95% Wilson CIs for proportions and 95% Mantel Haenszel CIs for stratum-adjusted differences in proportions with Mantel Haenszel test p-values will be provided for the three regions. ICEs will be handled as described in Section 6.4.6.

The last three supplementary analyses described below in Section 6.1.1.4 also constitute subgroup analyses. They will be performed for the three primary efficacy endpoints on subjects whose in-person attendance of the primary endpoint visit was not affected by COVID-19 pandemic. The target of estimation is specified in full detail in the definitions of three supplementary estimands CCI [redacted] The population attribute is a subset of the population defined in Section 6.1 for main primary estimands. For more details see Section 6.1.1.4.

CCI

[redacted]

CCI



6.1.2 Secondary Efficacy Endpoints and Estimands

Secondary efficacy endpoints and estimands for key secondary endpoints are described in Section 5.1.2. Definitions of MAS-based key secondary estimands are provided in CCI and of the subject's GAIS-based key secondary estimand in CCI No estimands are defined for further secondary endpoints.

6.1.2.1 Key Secondary Efficacy Endpoints and Key Secondary Efficacy Estimands

Main estimators for key secondary estimands will be calculated for the target population of main key secondary estimands and ICEs will be handled as described in Section 6.4.6.

Main estimators for main key secondary dichotomous MAS-based efficacy estimands (Section 5.1.2) correspond to main estimators for main primary estimands (see Section 6.1.1.1). Missing data will be imputed as described in Section 6.1.1.1. For all three anatomical regions and for investigator's as well as subject's assessments on the MAS, stratum-adjusted differences in proportions for overall comparisons of treatment Group U versus Group P (reference) with corresponding 95% Mantel Haenszel CIs and Mantel-Haenszel p-values will be calculated. Responder rates with two-sided 95% Wilson CIs will be provided as well.

Main estimator for the main key secondary estimand based on subject's GAIS CCI will be analysis of covariance with factors treatment group and site, and mean of baseline MAS for HFL, GFL, and LCL at maximum contraction assessed by subject as covariate.

CCI

According to the definition of the key secondary estimand based on subject's GAIS CCI specific ICEs will be handled by applying the composite strategy. In these cases, missing subject's GAIS ratings or ratings of CCI (see Section 6.4.6 on ICEs), regardless of whether the GAIS score from subject assessment at Day 30 (V4) is available or whether it is missing for the respective subject.

Least squares (LS) means, standard error, and two-sided 95% CI per treatment group and LS mean treatment difference ("Group U" – "Group P") with associated two-sided 95% CI and t-Test p-value will be provided. The null hypothesis of this t-Test is that the LS mean treatment difference is zero.

CCI

The resulting value will be rounded to two decimals.

Multiplicity adjustment strategy for comparisons of treatment groups U and P on key secondary endpoints is described in Section 6.4.3.

CCI

Results for Group L from any analyses will only be interpreted descriptively.

6.1.2.2 Further Secondary Endpoints

Similar analysis strategy by region as described for supplementary analysis of primary endpoints on observed cases in FAS/ PPS (see Section 6.1.1.4) will be applied for the

further secondary dichotomous MAS-based efficacy endpoints defined for the MP. In particular, proportions of subjects with at least one-grade improvement in treatment groups U and P will be compared by stratum-adjusted and unadjusted differences in proportions with corresponding 95% Mantel Haenszel CIs and p-values. Moreover, 95% CIs for proportions and unadjusted difference in proportions will be provided.

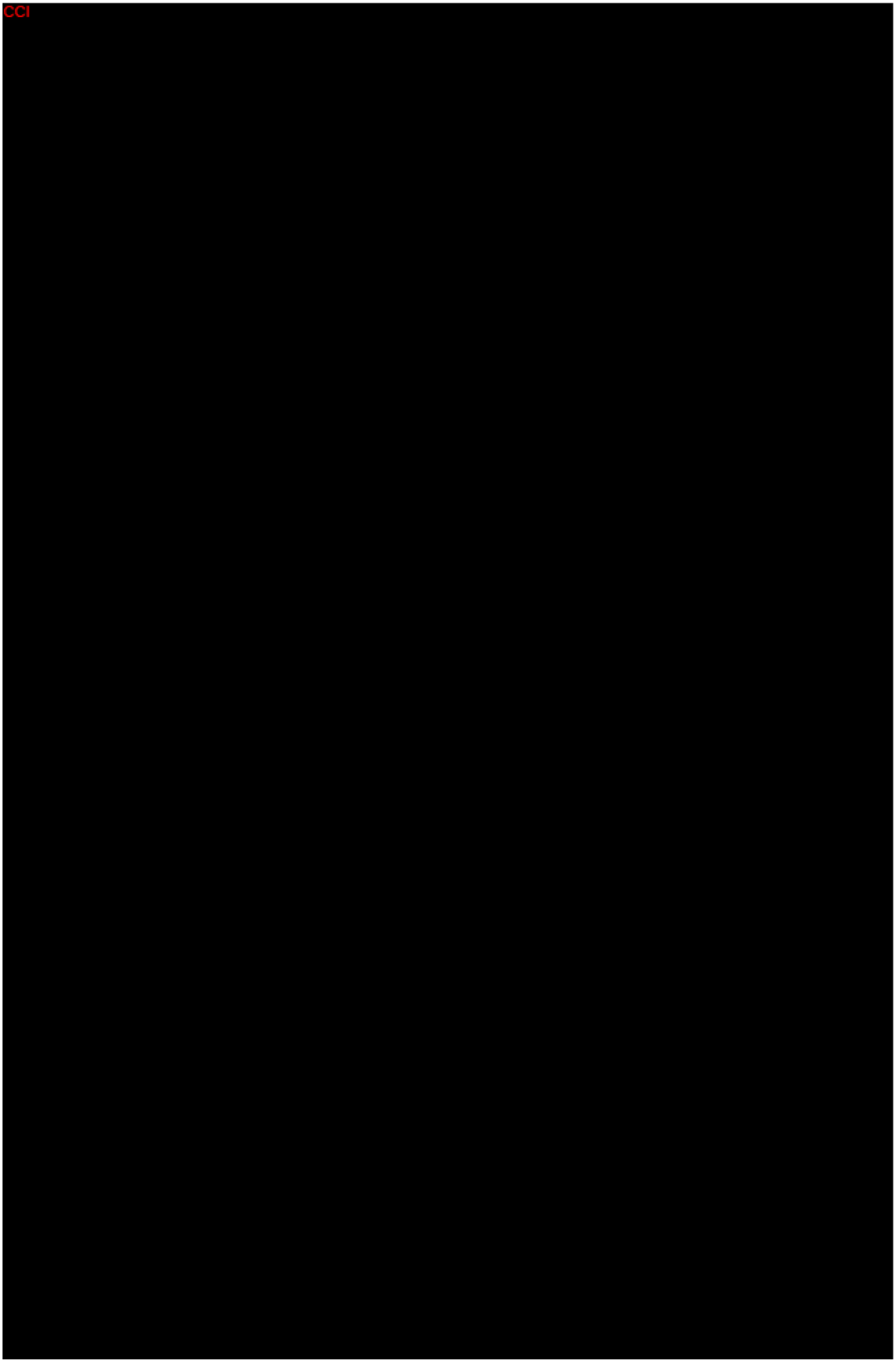
Group L will only be considered for statistical analyses of MAS of the LCL region but not of the GFL and HFL regions. Analysis of stratum-adjusted and unadjusted differences in proportions will be performed separately for Group U vs. Group P, Group L vs. Group P, and Total NT 201 vs. Group P. For Group L, proportions with two-sided 95% Wilson CIs will be provided for all three regions, but difference in proportions compared to Group P only for LCL region. For LCL region, proportions will also be provided for Total NT 201.

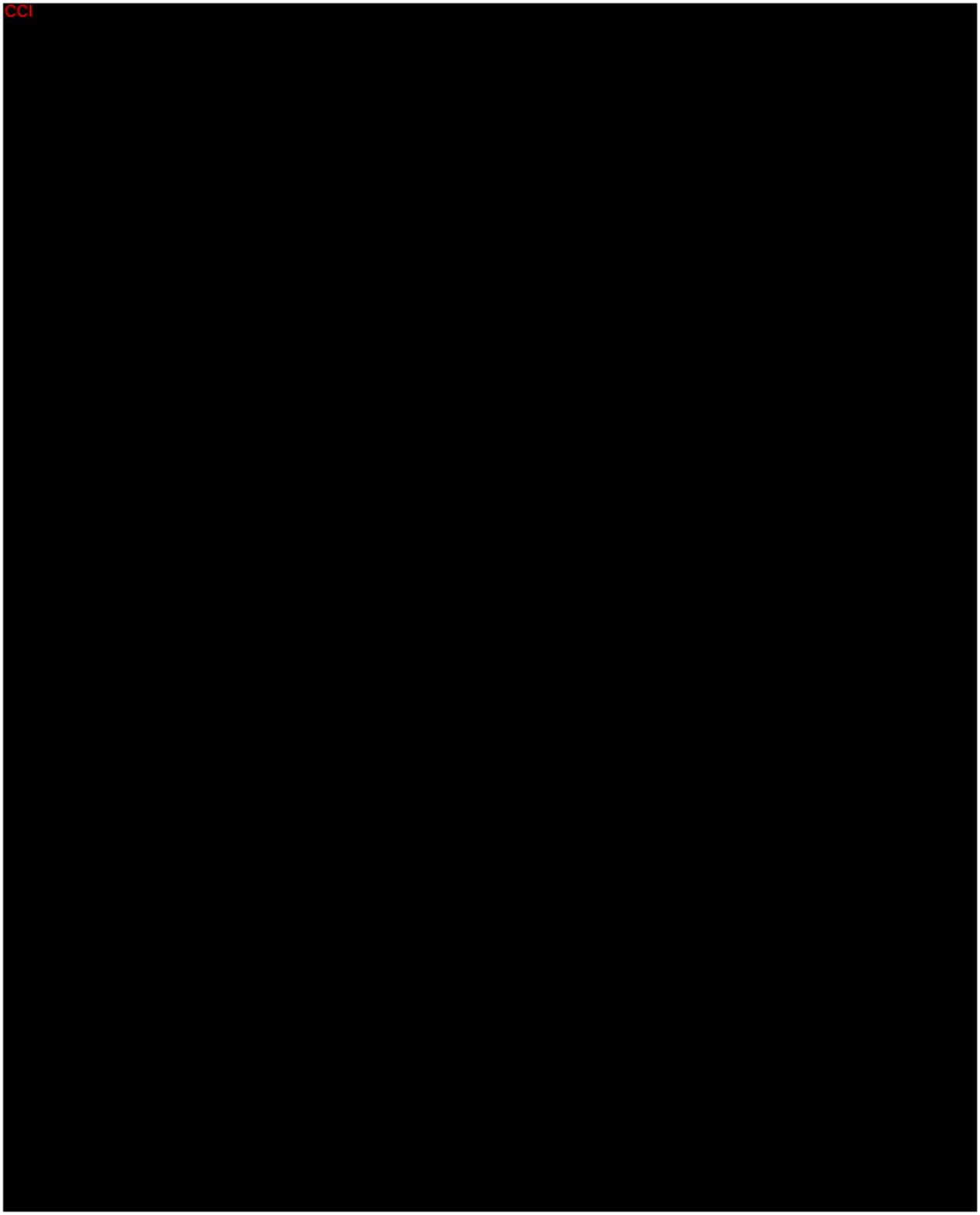
GAIS as assessed by the investigator at Day 30 of the MP will be analyzed using analysis of covariance with factors treatment group and site, and mean of baseline MAS for HFL, GFL, and LCL at maximum contraction assessed by investigator as covariate (see Section 6.1.2.1). Comparisons will be made for Group U to Group P, Group L to Group P, and Total NT 201 to Group P. Two separate ANCOVAs, one for Group P, Group U and Group L and one for Group P and Total NT 201 (Group U and Group L pooled) will be conducted.

Results for all further secondary endpoints including any p-values and 95% CIs will only be interpreted descriptively. No type I error adjustment will be applied.

For further secondary endpoints only observed cases will be analyzed. Analyses will be provided for FAS and PPS.

CCI





6.2 Safety Endpoints

All safety analyses will be performed on the SES. All safety tables will be presented by actual treatment group ("Group P", "Group U", "Group L") at V2 in MP. In TEAE tables for total OLEX period, Cycle 2, and Cycle 3, the column headers for these three treatment groups will be "MP Group P", "MP Group U", "MP Group L". TEAE tables for MP will additionally include a "Total NT 201" column for which active treatment groups (Group U and Group L) will be pooled. TEAE tables for total OLEX period, Cycle 2, and Cycle 3 will present pooled results from all three treatment groups in the fourth column, which will

be labeled "Total". For TEAE analyses over total study period, the columns "MP Group P - MP", "MP Group P - OLEX", "MP Group U", "MP Group L" and "MP Total NT 201" ("MP Group P – OLEX, MP", "Group U" and "Group L" pooled) will be presented. In column "MP Group P - MP" only AEs which occurred during MP are counted, in column "MP Group P - OLEX" only AEs which occurred in OLEX period are counted.

All analyses of laboratory values from any timepoint (Screening V1 and end-of-study visit V17) will be performed by actual treatment group at V2 in MP and in total. All analyses of vital signs from any timepoint will be performed by actual treatment group at V2 in MP, Total NT 201 (labeled "MP Total NT 201" for analysis of vital signs from visits total OLEX period) and in total.

AEs will be coded according to the MedDRA version in effect at the time the database is closed. Only TEAEs will be analyzed, which are defined as AEs with onset or worsening on or after date of the first administration of investigational product (IP) up to and including end of study visit (V17). In case no end of study visit (V17) is performed, all AEs reported during a time period corresponding to the expected cycle length will be regarded as treatment-emergent. Further details for separation of TEAEs and non-TEAEs can be found in CCI [redacted] Non-TEAEs will be listed only.

All analyses of TEAEs will be based on the following time periods:

- Total study period
- MP (Cycle 1)
- Total OLEX period (Cycle 2 and Cycle 3)
- Cycle 2
- Cycle 3.

TEAEs of the MP are defined as AEs with onset or worsening on or after date and time of the first administration of IP at V2 and before date and time of administration of IP in V8 (Cycle 2) or up to and including end of study visit (V17) in case of premature discontinuations prior to V8. For subjects who do not receive study treatment during OLEX period all AEs with onset or worsening on or after date and time of the first administration of IP up to and including end of study visit (V17) will be considered as TEAEs of MP. In case no end of study visit (V17)is performed, all AEs with onset or worsening up to and including Day 120 of MP will be regarded as TEAEs of MP.

TEAEs of the total OLEX period are defined as AEs with onset or worsening on or after date and time of the first administration of IP during OLEX period at V8 and up to and including end of study visit (V17). In case no end of study visit (V17)is performed, all AEs with onset or worsening at or after date and time of the first administration of IP during OLEX period at V8 and up to and including Day 120 of last cycle will be regarded as TEAEs of OLEX. TEAEs of Cycle 2 are defined as AEs with onset or worsening on or after date and time of the first administration of IP at V8 and before date and time of administration of IP at V13 (Cycle 3) or up to and including end of study visit (V17)in case of premature discontinuations prior to V13 or in case no study treatment is administered at V13. In case no end of study visit (V17) is performed, all AEs with onset or worsening at or after date and time of the first administration of IP at V8 and up to and including Day 120 of Cycle 2 will be regarded as TEAEs of Cycle 2. TEAEs of Cycle 3 are defined as AEs with onset or worsening on or after date and time of the first administration of IP at V13 and up to and including end of study visit (V17). In case no end of study visit (V17) is performed, all AEs with onset or worsening at or after date and time of the first

administration of IP at V13 and up to and including Day 120 of Cycle 3 will be regarded as TEAEs of Cycle 3.

For analysis over the total study period, the number of subjects in the SES will serve as the denominator for calculation of incidences. For the analyses by treatment cycle the denominator will be adjusted to all subjects being treated in that treatment cycle. For the analysis over the total OLEX period, the denominator for the incidences will be all subjects being treated at least once in total OLEX period.

If an AE worsens between start and end of this AE it will be considered as new AE starting with the date of worsening. In this case the imputation of end dates within episodes of the same AE will be done by setting the end date/end time to start date/start time of the consecutive worsening record.

For analysis, incompletely recorded or completely missing start of AEs will be estimated according rules laid down in CCI 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.

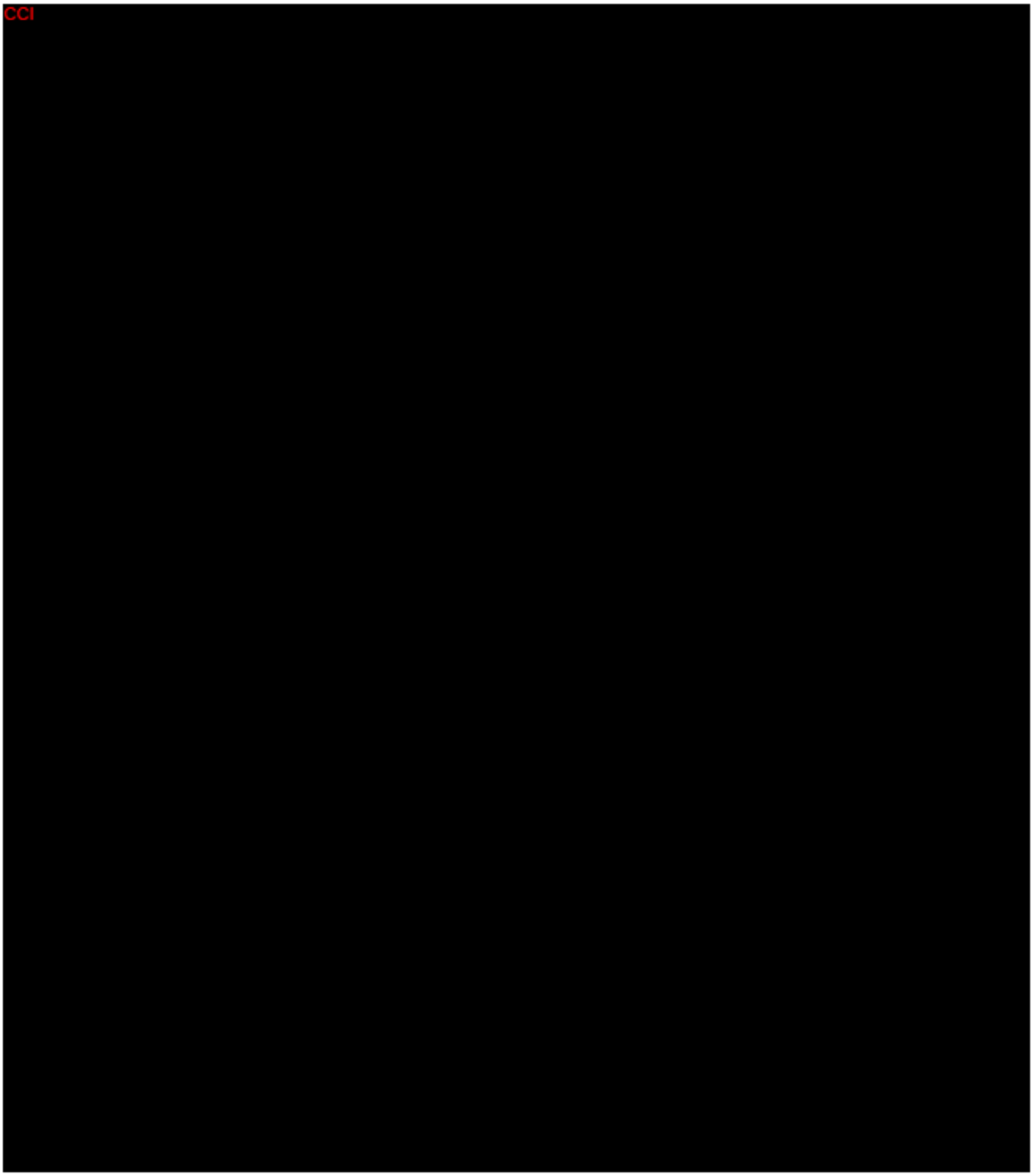
If a subject has more than one intensity/ causal relationship to treatment/ outcome within a preferred term (PT) only the worst intensity/ causal relationship to treatment/ outcome will be considered for calculation of incidence rates in the frequency tables reporting incidences by intensity/ causal relationship to treatment/ outcome. Incidence of TEAEs related to COVID-19 disease and TEAEs related to COVID-19 vaccination will only be calculated overall. Thereby, TEAEs with missing relationship will be ignored but affected subjects will not be removed from the denominator. Also, 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 same applies to calculation of overall incidence of TEAEs related to COVID-19 disease and TEAEs related to COVID-19 vaccination.

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

AEs coded as "Immunisation reaction" (preferred term) are defined as AEs related to COVID-19 vaccination.

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



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". Missing causal relationship to COVID-19 disease will not be imputed. TEAEs with missing relationship to COVID-19 disease will be ignored for calculation of the respective incidence rates but affected subjects will be kept in the denominator and only TEAEs which are documented as related to COVID-19 disease will be included in respective listing. No missingness is expected for relationship of TEAEs to COVID-19 vaccination.

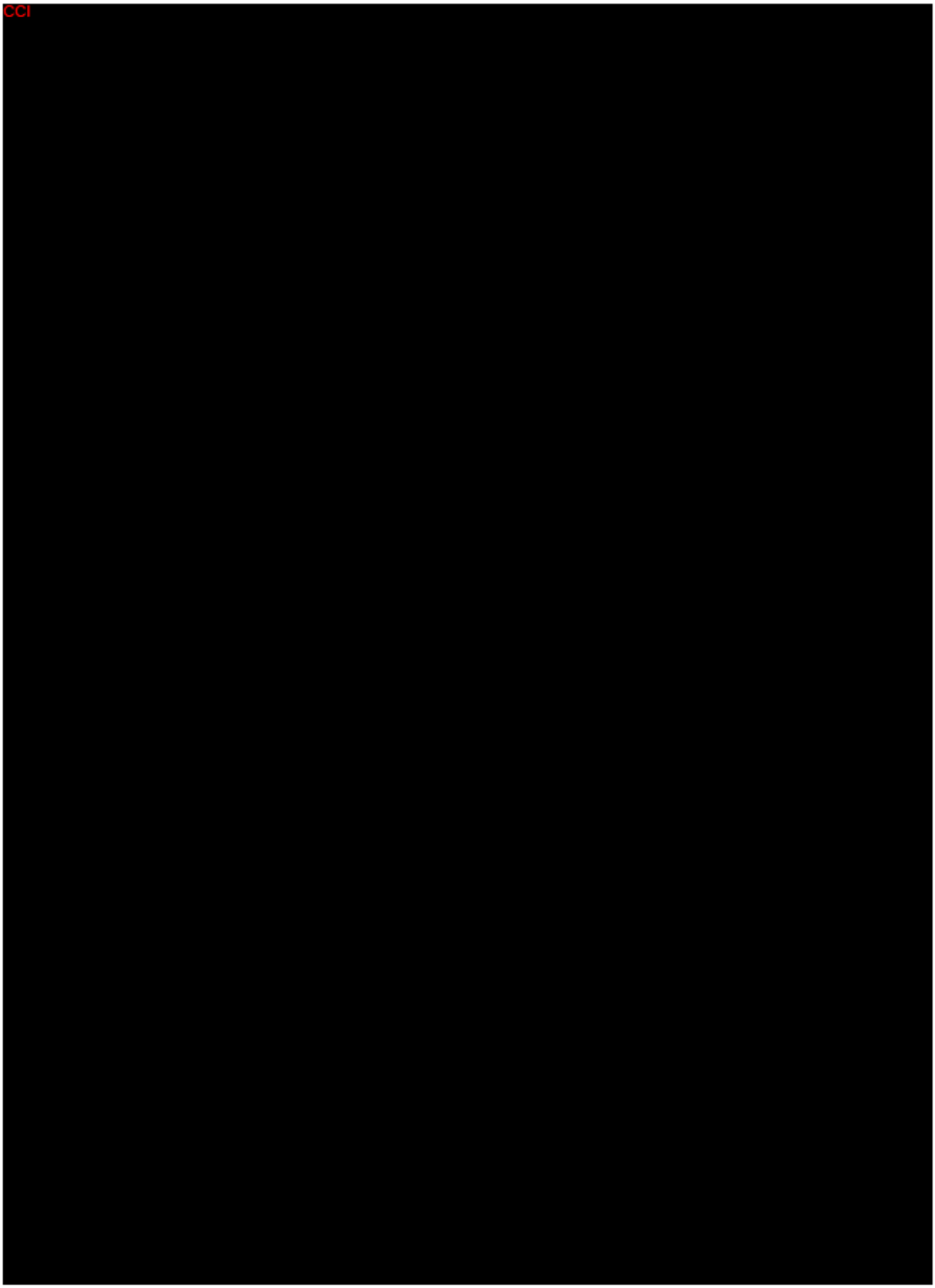
6.2.1 Primary Safety Endpoints

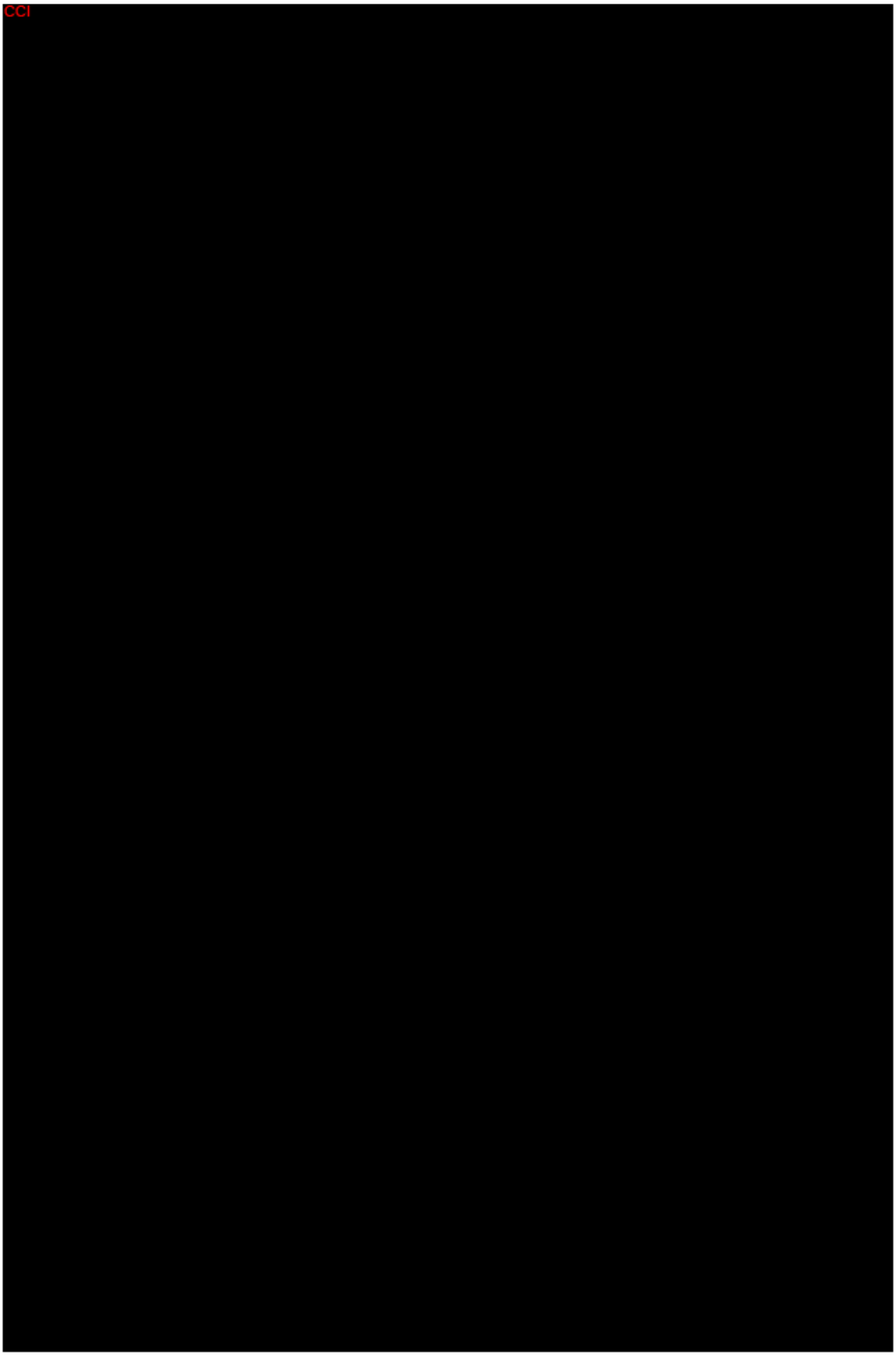
No primary safety endpoint has been defined.

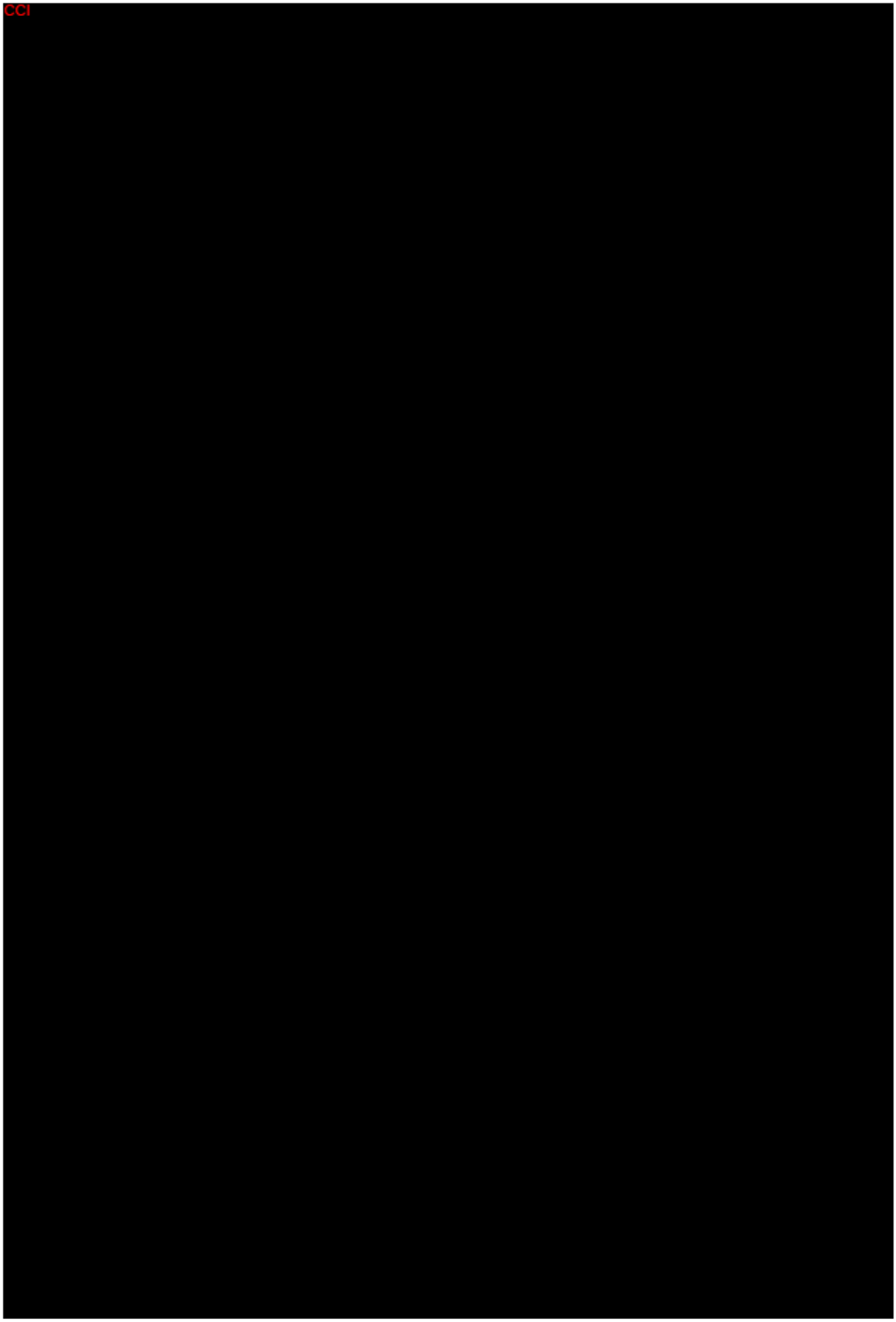
6.2.2 Secondary Safety Endpoints

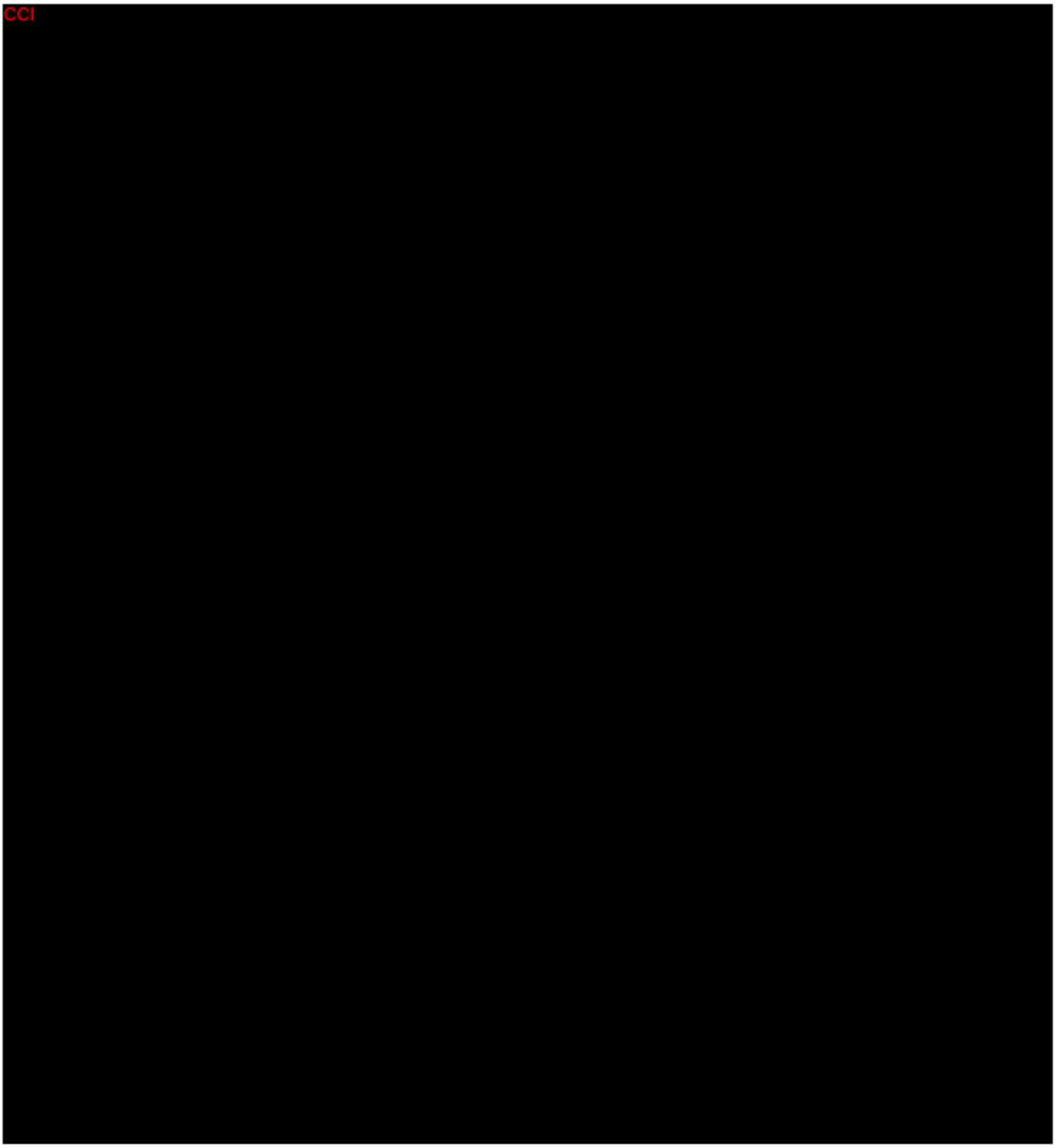
Tables displaying incidences of related TEAEs will be provided by system organ class (SOC) and PT for time periods and with table columns as defined at beginning of Section 6.2 above. Related TEAEs in MP and OLEX will also be listed.

CCI




CCI


CCI

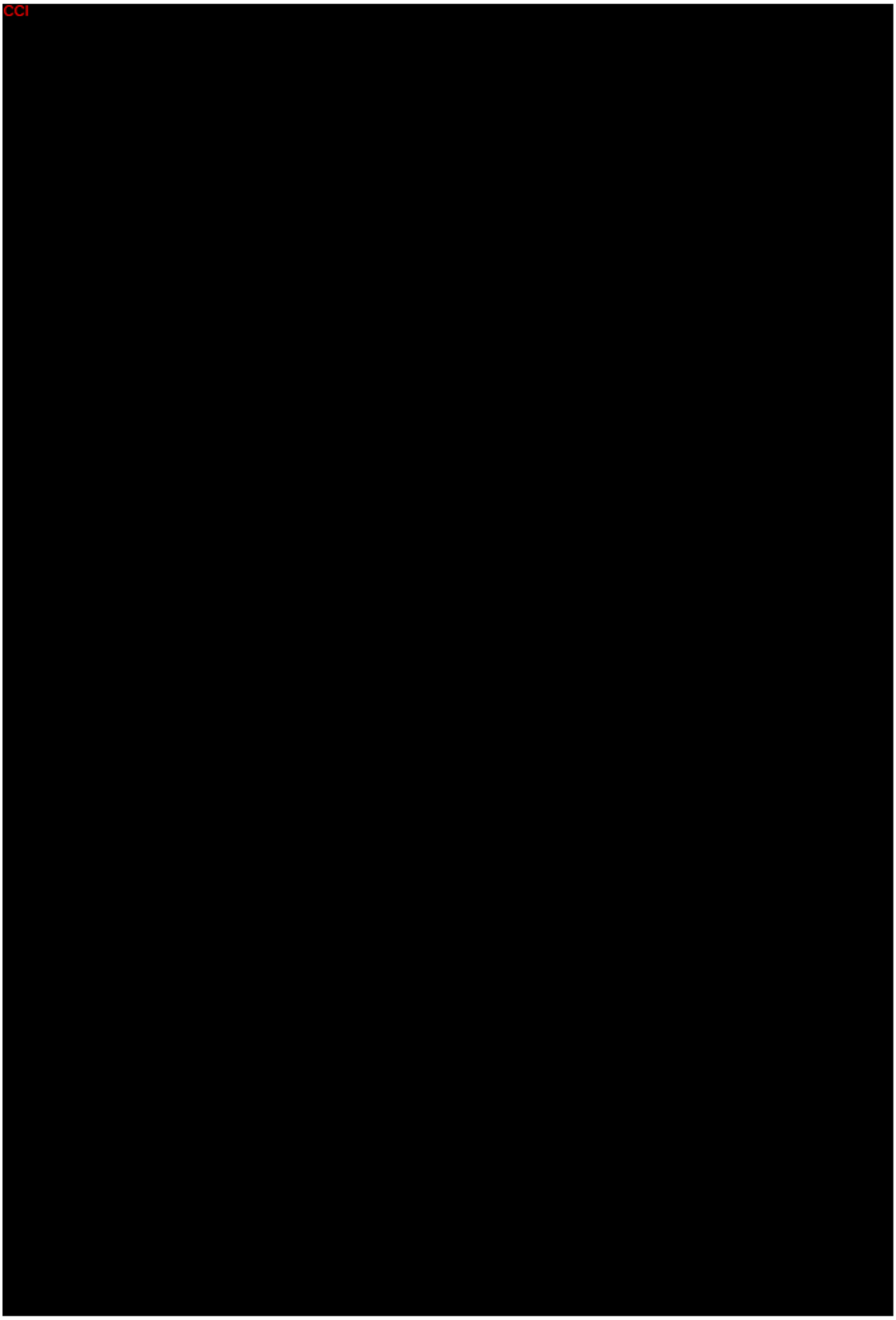


6.4 Special Statistical/Analytical Issues

6.4.1 Discontinuations and Missing Data

Main reasons for premature discontinuation of study will be summarized descriptively. As stated in Section 6.2.3, TEAEs leading to discontinuation will be listed. Discontinued subjects will not be replaced.

For subjects discontinuing the study prematurely and for whom end-of-study visit V17 was documented, efficacy data collected at V17 will be analyzed under the study visit which is closest to the recorded V17 date. This will be based on the assessment date relative to (cycle) baseline, and considering treatment received for each cycle, using the following visit windows.


CCI

CCI [redacted]

For separation of TEAEs from non-TEAEs in case of completely or partly missing start end and time see CCI [redacted] Rules for imputation of start dates and start times of AEs are provided in CCI [redacted] End dates of AEs will not be imputed.

For worst case imputation of missing causal relationship to treatment, intensity or outcome of an AE refer to Section 6.2. Causal relationship to COVID-19 disease will be listed based on observed cases.

If not otherwise specified, all TEAE analyses will be performed on the SES and all further safety analyses on observed cases.

6.4.2 Interim Analyses

Not applicable.

6.4.3 Multiple Comparisons/Multiplicity

The family-wise two-sided type I error level $\alpha = 0.05$ for comparisons of treatment groups U and P on the primary and key secondary endpoints will be controlled by a hierarchical test procedure based on two-sided tests with $\alpha = 0.05$.

The hierarchical procedure for primary analysis (Section 6.1.1.1) will be performed as described below. First, using the Mantel-Haenszel test of the null hypothesis that the common difference in response rates at Day 30 is zero, Group U versus Group P will be tested for superiority for proportion of GFL-responders, second for proportion of HFL-responders, and third for proportion of LCL-responders. The test procedure will stop once statistical significance could not be reached.

In case statistical significance can be established for all three primary efficacy endpoints, the procedure will proceed to key secondary endpoints (Section 6.1.2.1): As next steps, Mantel-Haenszel tests for superiority of Group U versus Group P will be conducted for proportion of subjects with a score of 0 (no) or 1 (mild) on MAS at maximum contraction as assessed by the *investigator* at Day 30 for GFL, then HFL and finally LCL anatomical region. Thereafter, Mantel-Haenszel tests for superiority of Group U versus Group P will be conducted for proportion of subjects with a score of 0 (no) or 1 (mild) on MAS at maximum contraction as assessed by the *subject* at Day 30 for GFL, then HFL and finally LCL anatomical region.

Finally, GAIS as assessed by the subject at Day 30 of MP will be tested as described in Section 6.1.2.1.

The hierarchical test procedure will stop once statistical significance could not be reached.

For any other statistical tests on primary or key secondary endpoints (e.g. comparison of treatment groups L and P), no type I error adjustment will be performed. The same accounts for any statistical analyses of further secondary and other endpoints.

6.4.4 Examination of Subgroups

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

Dichotomized endpoints which are based on MAS (assessed by investigator and subject) for GFL/HFL/LCL at rest will be analyzed in subgroups of subjects with baseline score of at least 2 on the corresponding MAS for GFL/HFL/LCL at rest as assessed by both subject and investigator. Since scores at rest are in general lower than scores at maximum

contraction, some subjects may have a score of 0 (no) or 1 (mild) at cycle baseline on MAS for GFL/HFL/LCL at rest. These subgroup analyses will be more meaningful and sensitive to treatment effects than analyses of the total FAS as they consider effects in 'symptomatic' subjects, i.e. subjects with 0 (no) or 1 (mild) on corresponding MAS will not be analyzed.

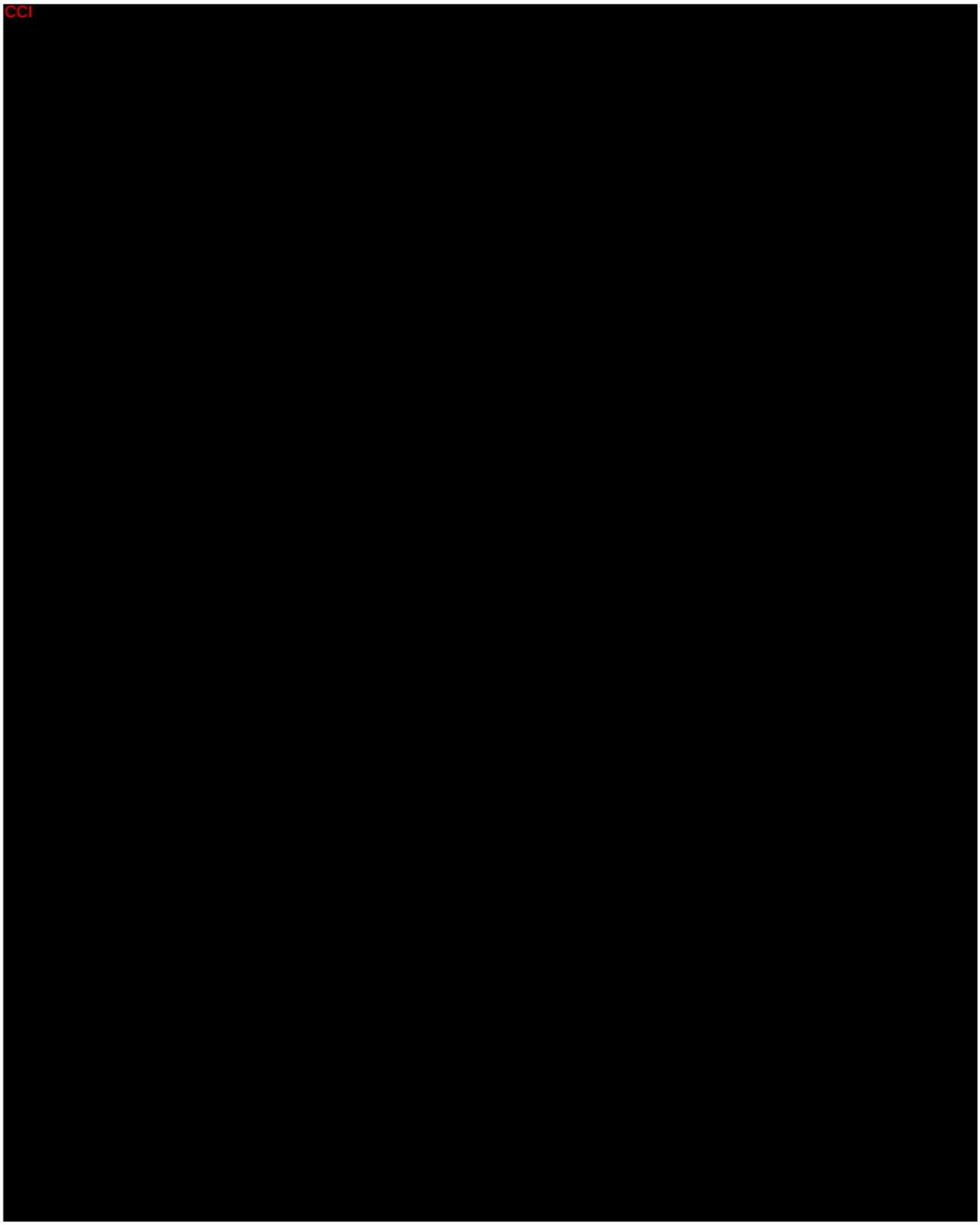
Dichotomized endpoints which are based on IRP assessments on MAS for GFL/HFL/LCL at maximum contraction and at rest will be analyzed in subgroups of subjects with baseline score of at least 2 on the corresponding MAS as assessed by all three raters. These subgroup analyses will be more meaningful and sensitive to treatment effects than analyses of the total FAS as they consider effects in 'symptomatic' subjects, i.e. subjects with 0 (no) or 1 (mild) on corresponding MAS will not be analyzed.

6.4.5 Pooling of Sites

No pooling of sites is planned.

6.4.6 Intercurrent Events

ICEs according to ICH E9 (R1) [1] are prespecified for main primary, supplementary primary, and main key secondary efficacy estimands (see Section 5.1.1, CCI and Section 5.1.2, respectively) and will be addressed by the handling strategies described in CCI

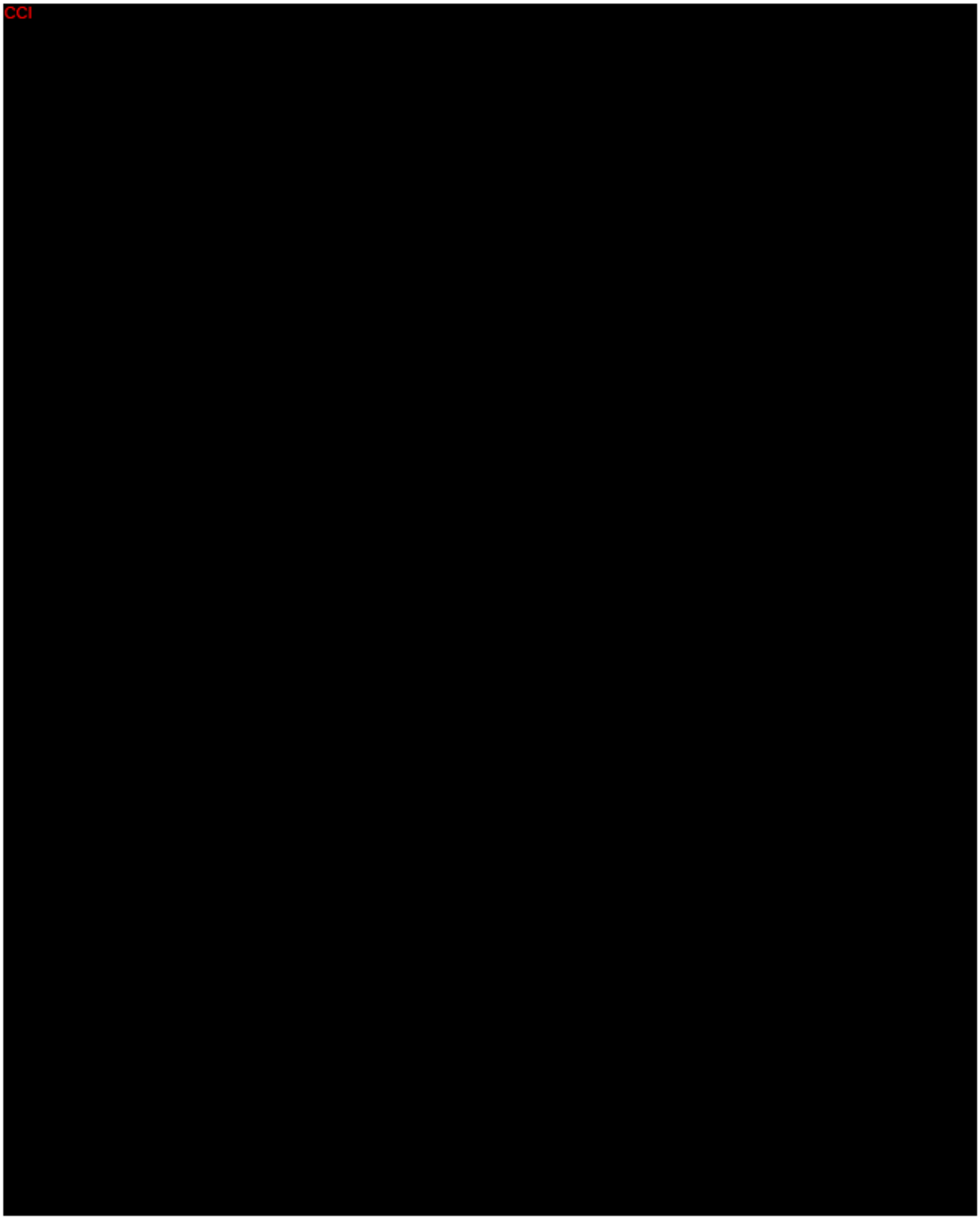


In case multiple ICEs with relevance for the same estimand occur for a specific subject, a conservative approach will be adopted for the respective estimand. For example, a composite strategy will be used if this is invoked by at least one of the ICEs relevant for the respective estimand and a treatment policy is used otherwise.

Generally, the above strategies for handling of ICEs will only be applied to estimators for affected estimands. An estimand can only be affected by an ICE if the subject is included

in the target population of the estimand, and only in this case a handling strategy for an identified ICE will be applied.

Likewise, handling strategies for ICEs will not be applied in analyses that are not defined as estimators for any efficacy estimands. This includes, e.g., analyses of primary and key secondary efficacy endpoints on FAS, PPS, observed cases in FAS/PPS and any analyses of further secondary or other efficacy endpoints or any safety or other endpoints.



8 References

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