NCT03777917, redacted
version v1.0, 15Mar2022

Merz North America, Inc.

# Statistical Analysis Plan

A Pilot Study to Assess the Effectiveness and Safety of Belotero Balance® Injection for Volume Augmentation of the Infraorbital Hollow

Device Pre-Market

M930121001

**Version 1.0 Final**

Date: 15-JUL-2019

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.

[redacted]
Senior Biostatistician

15-Jul 2019
Date (dd-mmm-yyyy)

[redacted]
Signature

[redacted]
Director, Head of Biostatistics

15-Jul-2019
Date (dd-mmm-yyyy)

[redacted]
Signature

[redacted]
Study Medical Expert

15-JUL-2019
Date (dd-mmm-yyyy)

[redacted]
Signature

[redacted]
Scientific Expert

15 - Jul - 2019
Date (dd-mmm-yyyy)

[redacted]
Signature

# TABLE OF CONTENTS

# 1 LIST OF ABBREVIATIONS

| Abbreviation/Term | Definition |
|---|---|
| AE | Adverse event |
| ATC | Anatomical Therapeutic Chemical classification system of the World Health Organization |
| BMI | Body mass index |
| CSR | Clinical study report |
| CTR | Common treatment response |
| eCRF | electronic case report form |
| FACE-Q | Set of subject-reported questionnaire modules |
| GAIS | Global Aesthetic Improvement Scale |
| IOH | Infraorbital hollow |
| [redacted] | |
| ITT | Intent-to-Treat |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MIHAS | Merz Infraorbital Hollow Assessment Scale |
| n imp | Number of imputed values |
| n miss | Number of missing values |
| n obs | Number of observed values |
| OC | Observed case |
| PDF | Portable document format |
| PP | Per protocol Population |
| PT | Preferred term |
| SAE | Serious adverse events |
| SAP | Statistical analysis plan |
| SAS® | Statistical Analysis System software |
| SD | Standard deviation |
| SP | Safety Population |
| SOC | System organ class |
| TFLs | Tables / figures / listings |
| TEAE | Treatment emergent 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 M930121001, dated 13-Mar-2017 and the following amendments, dated 09-Aug-2017, 05-Dec-2017, 04-Oct-2018, and 08-Nov-2018.

The Merz standard Table, Figures, and Listings (TFLs) for medical devices, version 1.0 29.06.2018) will be applied and adapted to study specific requirements as laid down in the clinical study protocol and any amendments.

# 3 CLINICAL STUDY DESIGN AND OBJECTIVES

## 3.1 Clinical Study Design

This is a prospective, [redacted] multi-center, randomized-controlled study in subjects with moderate to severe infraorbital hollow (IOH) deficit. Approximately 66 subjects will be enrolled from 3 sites in the United States. Subjects will be randomized to either a treatment group or an untreated control group using a 2:1 (treatment: control) allocation ratio. Approximately 70% of the total sample size will consist of subjects with a Fitzpatrick skin type of I, II, or III and approximately 30% of the subjects will be Fitzpatrick skin type IV, V, or VI. Subjects from the Fitzpatrick skin type IV, V, and VI group will be distributed as follows: ≥ 6 subjects will be enrolled in the IV Fitzpatrick skin type group and ≥ 12 subjects will be enrolled with Fitzpatrick Skin Types V and VI Fitzpatrick skin type group. Each clinical site will enroll a minimum of 6 subjects with Fitzpatrick skin types IV, V, VI. At least 5 males will be enrolled into the study.

For subjects randomized to the treatment group, both right and left IOHs will receive treatment with Belotero Balance. To achieve symmetrical correction a touch-up injection will be given, with the subject's consent, in one or both IOHs if the treating investigator determines a treated subject has asymmetrical IOHs based on a visual assessment.

For the primary effectiveness assessment, IOH deficit will be assessed according to the Merz Infraorbital Hollow Assessment Scale (MIHAS) [redacted].

The treated subjects will have a safety phone call 72 hours after baseline treatment and in-clinic safety visits at Week 2 and Months 2, 3, 6, 9, and 12 post baseline injection. Effectiveness assessments will be performed in clinic at baseline and Month 2 post baseline injection. [redacted]

All treated subjects will be assessed 1 month after baseline injection for asymmetry and safety. If a treated subject receives a touch-up injection for asymmetric correction the visits schedule will be re-calculated relative to the touch-up visit (i.e., 72 hour phone call, 2 weeks, Months 1, 2, 3, 6, 9 and 12). For these subjects effectiveness assessments will occur at baseline and 2 months after the touch-up injection.

If subjects report a safety concern during the 72-hour phone call, an unscheduled visit will be scheduled to bring the subject into the clinic to address safety concerns.

Control-group subjects will be evaluated at enrollment and Month 2 in the clinic. [redacted]

## 3.2 Clinical Study Objectives

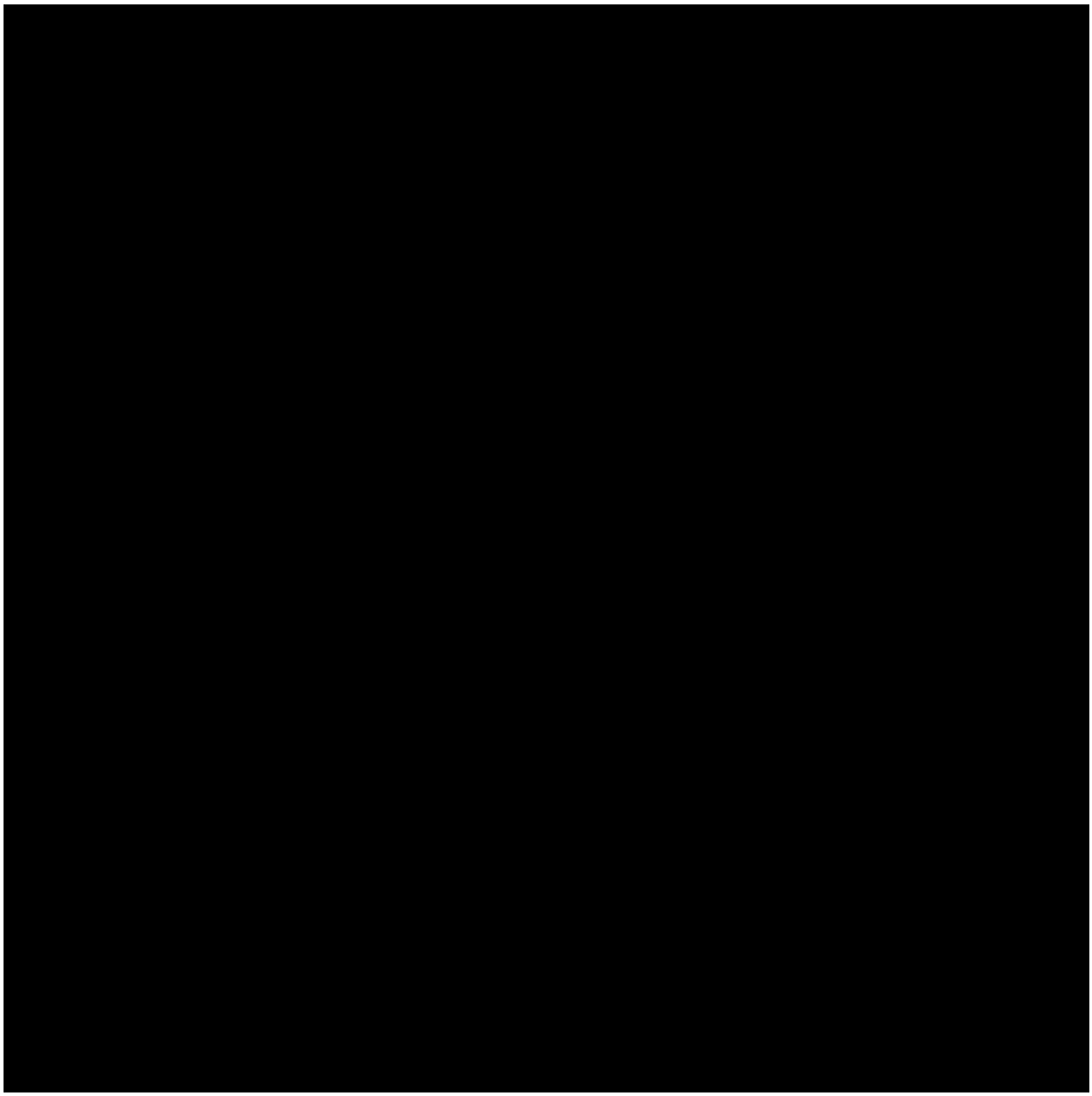
The pilot study aims to define safety, effectiveness, and patient-reported outcomes for Belotero Balance use in the IOH .

### *3.2.1 Effectiveness*

The effectiveness and relevance of aesthetically pleasing outcomes following Belotero Balance injection will be established in the following manner:

- The primary endpoint will establish effectiveness by using the MIHAS to demonstrate that clinically relevant changes of ≥ 1-point on both IOHs can be detected 2-months post-treatment .
- The secondary endpoints including the validated FACE-Q satisfaction with eyes, the investigator GAIS, and the subject GAIS will be utilized to substantiate aesthetically pleasing outcomes.

Additionally, the investigator and subject GAIS scores at Month 2 post last injection visit will be utilized to demonstrate the level of improvement, when compared to baseline photographs, resulting from treatment in the IOHs.



In totality, the results from the MIHAS obtained from the treating investigator and subject perspectives regarding aesthetic improvements, will be used to establish that Belotero Balance injections in the IOH region are effective in producing clinically relevant and aesthetically pleasing outcomes.

### *3.2.2 Safety*

The safety objectives include the identification and description of adverse events (AEs) and serious adverse events (SAEs) during the course of the study. In addition to standard safety assessments, eye assessments (including visual acuity, visual field, ocular motility, and undilated fundoscopic exam) will be evaluated. Common treatment responses (CTRs) will also be assessed.

# 4 DETERMINATION OF SAMPLE SIZE

In order to assess the ability of the planned sample size to provide power for a statistical comparison of the treatment and control groups success proportions (based on a ≥ 1-point improvement on both IOHs using the MIHAS from baseline to Month 2), a power calculation using a Fisher's exact test with a binomial distribution was performed using the following assumptions:

- Type I error rate: 0.05 (two-sided)
- Control group success proportion: 0.20
- Treatment group success proportion: 0.60
- Allocation ratio (control: treatment): 1:2
- Total N (number of evaluable subjects): 60 (40 subjects in the treatment group and 20 subjects in the control group).

Based on the above assumptions, the study power is approximately 80%. To account for possible missing data, attrition, and/or study deviations (up to 10%), a total of 66 subjects will be enrolled (44 subjects in the treatment group and 22 subjects in the no treatment control arm).

Approximately 70% of the total sample size will consist of subjects with a Fitzpatrick skin type of I, II, or III, and approximately 30% of the subjects will consist of subjects with a Fitzpatrick skin type IV, V, or VI. Subjects from the Fitzpatrick skin type IV, V, and VI group will be distributed as follows: ≥ 6 subjects will be enrolled in the Fitzpatrick skin type group IV and ≥ 12 subjects will be enrolled in the Fitzpatrick Skin Type group V and VI. Each clinical site will enroll a minimum of 6 subjects with Fitzpatrick skin types IV, V, VI. At least 5 males will be enrolled into the study.

In the event that the sample size accounting for missing data, attrition, and/or deviations is attained (i.e., 44 treated subjects and 22 untreated/control subjects), using the assumptions above, the study power would be approximately 84%.

nQuery 8, Statistical Solutions Ltd., was used for the power calculation.

## 4.1 Randomization

The randomization procedure for this study was a block randomization stratified per site which was computer-generated independently of the study team. Block randomization controlled for potential deviations from the intended allocation ratio between treated and control subjects. Accordingly, a 2:1 allocation ratio was utilized with approximately 44 subjects randomized to treatment and 22 subjects to control.

# 5 ANALYSIS SETS

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

**Safety Population (SP)**

The SP consists of all randomized subjects who received at least one study treatment in the active treatment group or who were randomized into the control arm.

**Intent-to-Treat Population (ITT)**

The ITT consists of all randomized subjects. This will be the primary population used for the effectiveness analyses. All effectiveness endpoints will be analyzed as randomized.

**Per Protocol Population (PP)**

The PP is a subset of subjects in the ITT population without major protocol deviations. Final determination of what constitutes major or minor protocol deviations will be made prior to database lock.

Subjects which have not been treated according to the randomization list will be excluded from PP. In general, tables performed on the SP will present subjects as treated and tables performed on the ITT will present subjects as randomized. In the subject data listings, the randomized treatment arm will be listed except for the section 16.2.7 (Adverse events) and 16.2.8 (Other safety variables): here the actual treatment arm will be listed.

# 6 ENDPOINTS FOR ANALYSIS

Unless otherwise specified, for all effectiveness endpoints, Month 2 for treated subjects is from the time of last injection (i.e., either baseline treatment or touch-up, if applicable). For randomized and not treated subjects, the effectiveness assessments will be performed at Month 2 from the baseline visit.

## 6.1 Effectiveness Endpoints

### *6.1.1 Primary Endpoint*

The primary endpoint is a comparison of the responder rate between the treatment group and the untreated control group at Month 2, according to the MIHAS [redacted].

Treatment response is defined as ≥ 1-point improvement on the MIHAS scale for both IOHs compared to baseline

### *6.1.2 Secondary Endpoints*

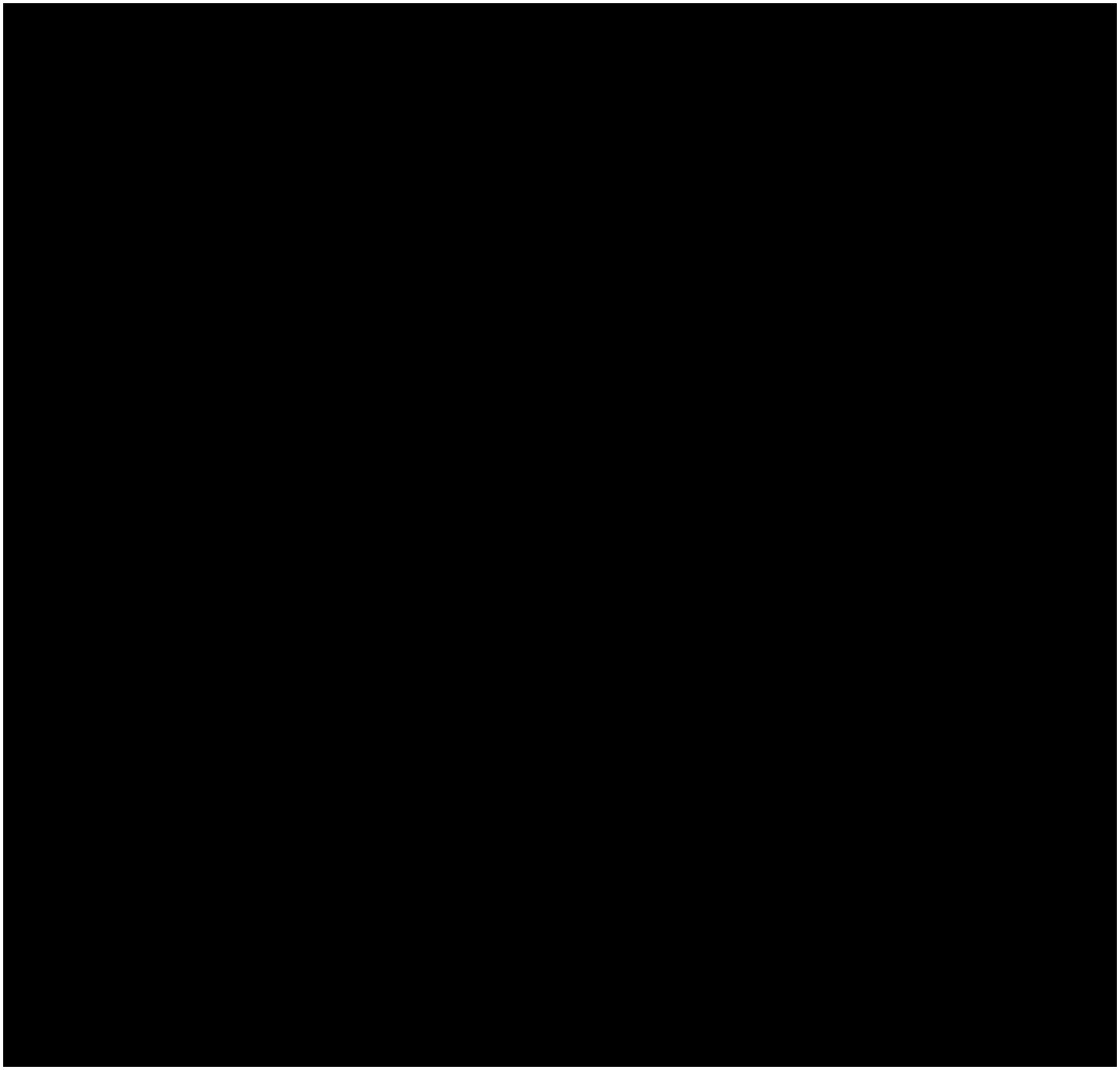
- Summary of the FACE-Q satisfaction with eyes scores for treated and control subjects at baseline and Month 2. The subject's assessment is based on taking into consideration both eyes.

- Descriptive summary of GAIS scores for treated subjects at Month 2, as completed by the treating investigator. [redacted]

- Descriptive summary of GAIS scores for treated subjects at Month 2, as completed by the subject. [redacted]

- Summary of the responder rates in the treatment group and the control group at Month 2, according to the MIHAS [redacted].

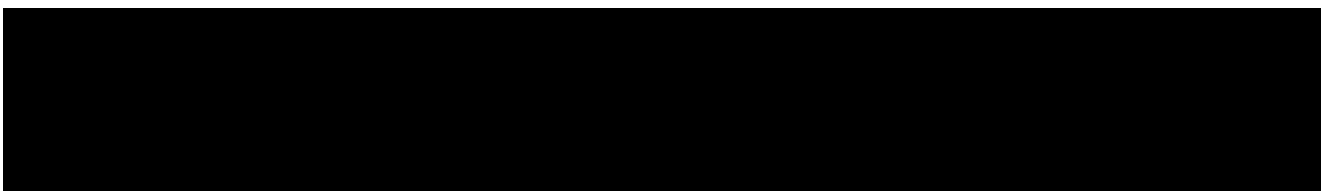
  Treatment response is defined as ≥ 1-point improvement on both IOHs when comparing the change from baseline to Month 2. A subject will be considered a responder if a treatment response of at least 1-point change on both IOHs is determined [redacted].

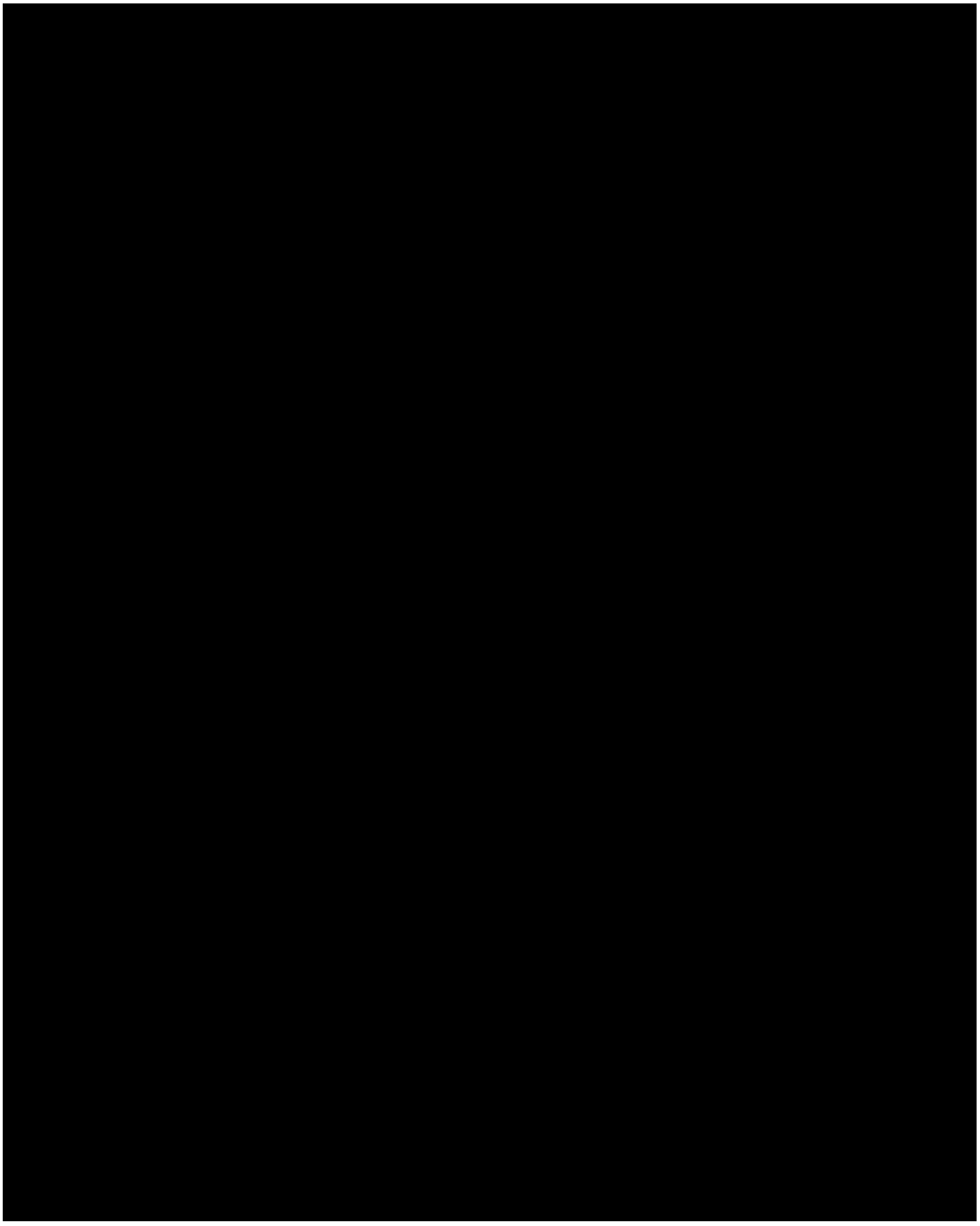


## 6.2 Safety Endpoints

- Incidence and nature of device- and/or injection-related AEs and SAEs observed during the study.

- Incidence, severity, and duration of pre-specified CTRs reported in subject diaries.





# 7 STATISTICAL ANALYSIS METHODS

The sponsor will finalize the formal SAP prior to database lock. Deviations from the analyses outlined in the study protocol are documented in this SAP.

Adequate descriptive statistics will be provided for each endpoint. Continuous variables will be summarized by number of observations, number of missing or number of imputed observations, mean, standard deviation, minimum, median, and maximum. Categorical data will be summarized in frequency tables using counts (n) and percentages (%). Ordered categorical data will be summarized by both.

If not otherwise specified Month 2 data refers to

- Month 2 post last injection visit (i.e., either baseline treatment or touch-up, if applicable) for subjects in the treatment group and
- Month 2 post baseline visit for subjects in the control group.

## 7.1 Effectiveness Endpoints

The primary efficacy analysis and all secondary endpoints will be based primarily on the ITT and additionally, for sensitivity purposes, on the PP. [redacted]

Statistical tests will be two-sided hypothesis tests for between-treatment arm differences in general. Continuous variables (values and changes from baseline) will be summarized by number of observed values (n obs), number of missing values (n miss)/ number of imputed values (n imp), mean, standard deviation (SD), median, minimum, and maximum. For qualitative variables, absolute and percent frequencies (n, %) including the number of missing values and, if applicable, shift tables also including the number of missing values will be displayed. Confidence limits and descriptive p-values will be given, where appropriate.

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'.

Mean, 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. Percentages will be calculated using the denominator of all subjects in a specified population, or treatment arm. The denominator will be specified in a footnote to the tables for clarification if necessary. Percentages will be reported to one decimal place.

### 7.1.1 Primary Effectiveness Endpoint

The primary effectiveness endpoint, the response rate (proportion of subjects with ≥ 1-grade improvement of both IOHs on the MIHAS from baseline to Month 2) will be summarized as percentages for the treatment ($p_{treatment}$) (coded with "TreatmentArm"=0) and control ($p_{control}$) (coded with "TreatmentArm"=1) subjects and inferentially compared according to the following hypothesis:

$H_0$ (null): $p_{treatment} \leq p_{control}$

$H_a$ (alternate): $p_{treatment} > p_{control}$

For this hypothesis testing, the Fisher's exact test will be used to compare both response rates in order to test for the superiority of treatment over control. If the left-sided p-value obtained from this test is < 0.025, $H_0$ will be rejected in favor of $H_a$. The following SAS code will be used:

```
proc freq data = dataset;
 tables TreatmentArm*Response /fisher riskdiff(column=2 cl=newcombe)alpha=
                                0.05;
ods output fishersexact = out pdiffcls=out1;run;
```

Assuming the response is coded with '1' and no response is coded with '0'.

Furthermore, a two-sided 95% Newcombe based on Wilson scores confidence interval for the risk difference will be calculated. The primary analysis of the response rates at Month 2 will be based on the ITT, using the observed case (OC) values. The supportive analyses of primary effectiveness endpoint will be for the ITT using imputed (missing imputed as no change) and for the PP using the OC.

Moreover, the response rates for ITT (OC) at Months 2 will be provided for the following subgroups:

- Fitzpatrick Skin Types (I, II, III versus IV, V, VI)
- Gender

[redacted]

### 7.1.2 Secondary Effectiveness Endpoints

If not otherwise specified, analyses of secondary endpoints will be based on OC for the ITT and the PP subset.

- The FACE-Q satisfaction with eyes sum scores and the Rasch-transformed scores ranging from 0 to 100 will be summarized by treatment arm (i.e. treatment and control) using descriptive statistics for baseline and Month 2 and change from

baseline. The two-sided 95% for the mean (assuming normal distribution) will be provided in addition. The single questions will not be summarized. The Rasch-transformed scores are computed as sum scores from the responses to the separate items and then converting the sum scores to a scale from 0 to 100.

Higher scores reflect a better outcome. If missing data is less than 50% of the scale's items, the rounded mean of the completed items will be inserted for any missing item. If missing data is 50% or more of the scale, the sum score will be set to missing. This will be done for the sum of untransformed scores. [redacted]

- The GAIS at Month 2 completed by investigator and subject will be summarized descriptively for treated subjects. In addition, the investigator and subject GAIS will be categorized in classes [redacted]

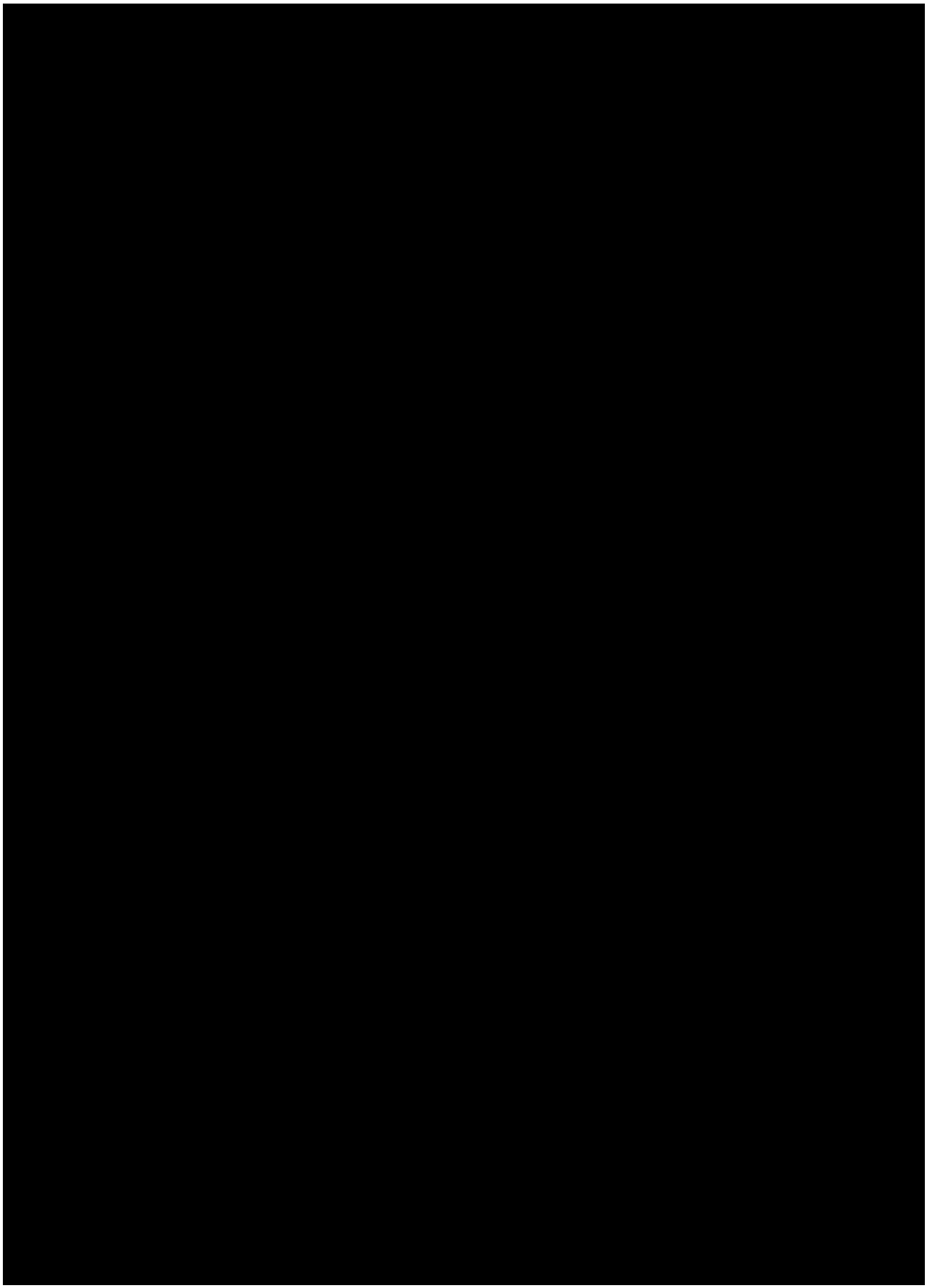
- The responder rates at Month 2 according to the MIHAS [redacted] will be summarized by treatment arm (i.e. treatment and control). Treatment response is defined as ≥ 1-point improvement on both IOHs compared to baseline. A subject will be considered a responder if a treatment response of at least a 1-point change on both IOHs would be determined [redacted]. Additionally, the proportion of responders [redacted] will be computed based on the definition of treatment response detailed above. This will be summarized descriptively and by a two-sided 95% Newcombe confidence interval for the risk differences.

  The original MIHAS results of both IOHs [redacted] will be further presented descriptively in frequency tables for baseline and Month 2 whereby MIHAS [redacted] and the median MIHAS will be tabulated.

Moreover, all secondary effectiveness endpoints will be summarized for the following subgroups:

- Fitzpatrick Skin Types (I, II, III versus IV, V, VI)
- Gender

[redacted]



## 7.2 Safety Endpoints

All safety analyses will be performed on the SP. Summary tables will be stratified by actual treatment received and not by randomized treatment arm.

**Adverse events**

All AEs will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA) version in effect at the time the database is closed.

TEAEs are defined in the following way:

- Treated subjects: all AEs with onset or worsening at or after the first administration of study treatment
- Not treated subjects: all documented AEs.

For treated subjects, TEAEs will be allocated to two different periods:

- Period 1: AEs with onset or worsening on or after date and time of the first administration of study treatment and up to Month 2 visit.
- Period 2: AEs with onset or worsening after Month 2 visit.

All analysis will be performed by period and overall, whereby only analysis for period 1 will be stratified by actual treatment arm. For the interim analysis, only time period 1 will be analyzed. Based on the exact number of TEAEs that will be documented in period 2, some of the summary tables may be omitted. A separate listing that contains only TEAEs of period 2 will be added.

An overall summary of AEs will be provided for the following, any line that is not applicable may be deleted in the final table:

- Any AEs

- Any non-TEAEs (for treated subjects)
- Any TEAEs
- Any treatment-related TEAEs (for treated subjects)
- Any serious TEAEs
- Any serious treatment-related TEAE (for treated subjects)
- Any TEAEs leading to discontinuation
- Any treatment-related TEAEs leading to discontinuation (for treated subjects)
- Any fatal TEAEs
- Any treatment-related fatal TEAEs (for treated subjects)

Incidences will be provided for the following classes of AEs:

- TEAEs, subjects with TEAEs and number of TEAEs by system organ class (SOC) and preferred term (PT)
- TEAEs, subjects with TEAEs by PT
- TEAEs by worst intensity, subjects with TEAEs by PT
- Related TEAEs by worst intensity, subjects with TEAEs by PT
- TEAEs by worst causal relationship, subjects with TEAEs by PT
- TEAEs by worst outcome, subjects with TEAEs by PT
- TEAEs by maximum duration, subjects with TEAEs by PT (1-3 days, 4-7 days, 8-14 days, ≥15 days)
- Treatment related TEAEs, subjects with TEAEs and number of TEAEs by SOC and PT
- TESAEs, listing of subjects (in case of more than 15 TESAEs respective table by SOC and PT will be provided)
- Deaths, listing of all serious TEAEs leading to death
- TEAEs leading to discontinuation, listing of subjects.

The duration of AEs will calculated by the difference of start and stop day plus 1.

If a subject has more than one outcome within a preferred term (PT) only the worst outcome in respective time period will be used in the frequency tables. Also on subject level only the worst outcome category per subject in respective time period 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

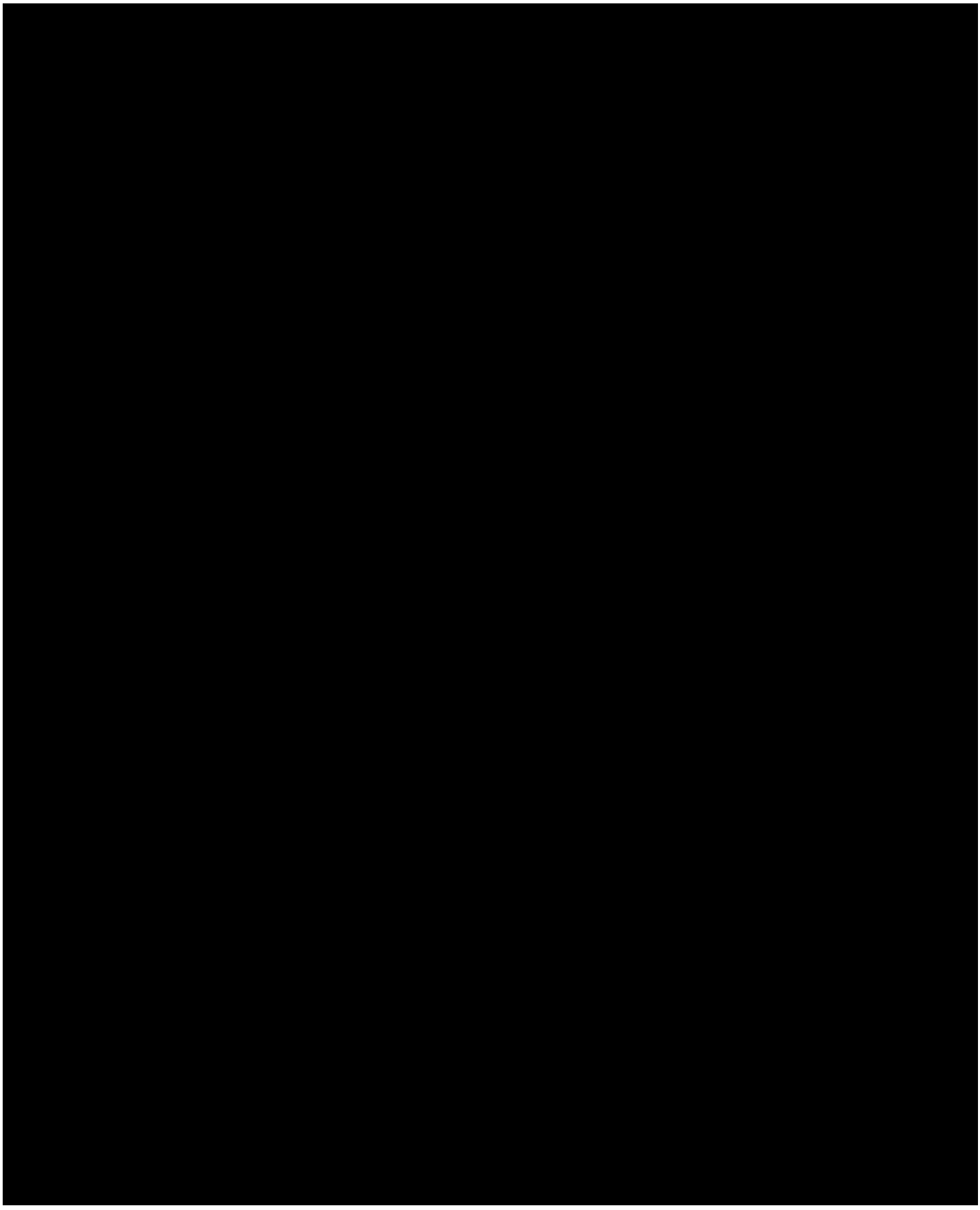
Moreover, incidences of TEAEs (PT and SOC) and of related TEAEs (PT and SOC) will be provided for the following subgroups:

- Fitzpatrick Skin Types (I, II, III versus IV, V, VI)
- Gender

[redacted]

Urine pregnancy tests, visual safety assessments and device deficiencies will only be listed.

[redacted]



## 7.4 Special Statistical/Analytical Issues

### *7.4.1 Discontinuations and Missing Data*

#### *7.4.1.1 Subject discontinuation and missing effectiveness data*

To account for missing effectiveness data for the primary endpoint, missing MIHAS data will be imputed as no change (i.e. subject is considered as non-responder). This approach is conservative because it assumes that missing values are attributable to lack of treatment benefit. Furthermore, given that the primary effectiveness is only performed at baseline and Month 2 post last injection (i.e., either baseline treatment or touch-up, if applicable), the only observation that can be imputed for a missing Month 2 after the last injection time point is the baseline value, which assumes no change. Additionally, analyses will also be performed using the OC method where no missing value imputations will be conducted.

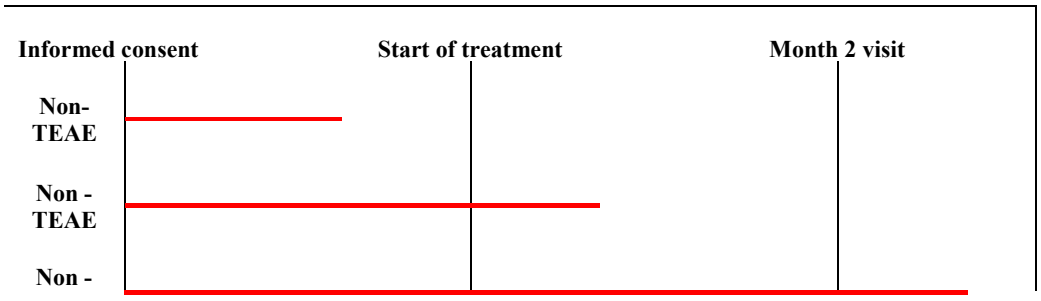
[redacted]

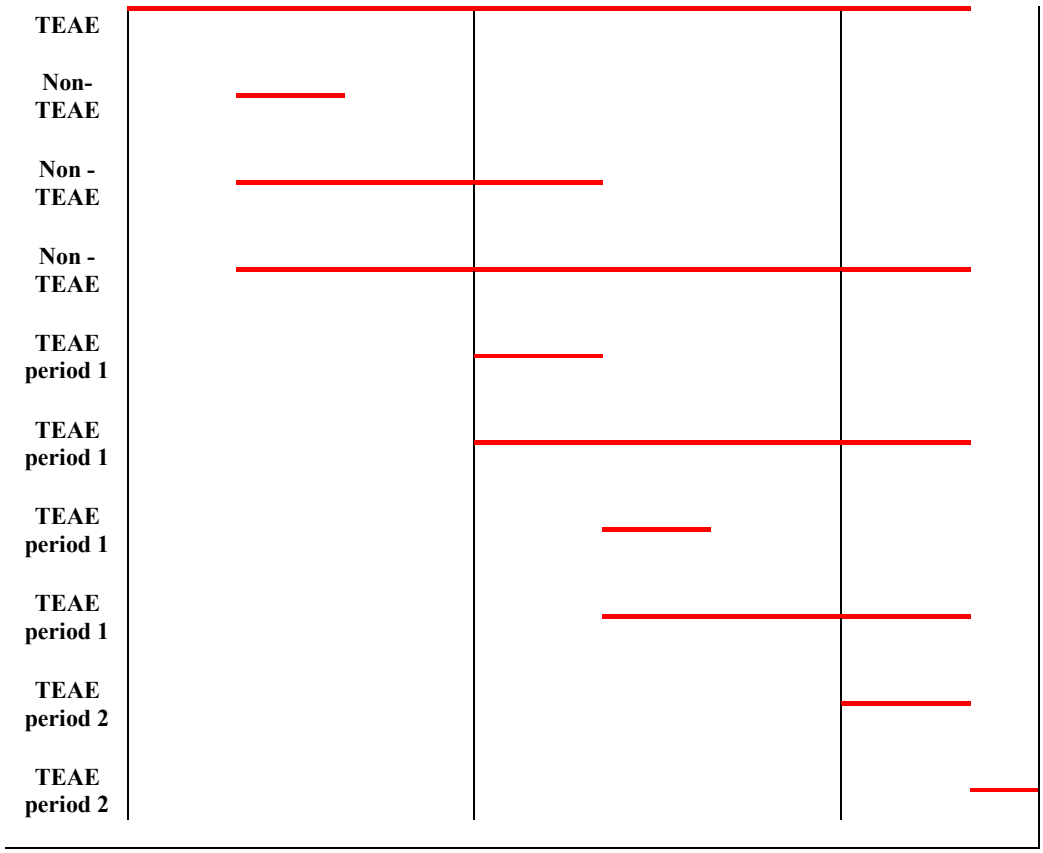
#### *7.4.1.2 Imputation of safety data*

**Adverse events**

For the imputed analysis start date of AEs (variable ASTDT in the ADaM domain ADAE) the following imputation rules related to any treatment period were used if applicable (only relevant for treated subjects):

**Figure 1: Possible start/end dates of AEs**






(Non) TEAE = (Non) Treatment Emergent Adverse Event

1a) If the start date is missing completely and the end date and time* does imply end before first injection then no imputation of start date and time will be performed.

1b) If the start date is partially missing and the partially missing start date and/or the end date and time* imply start and/or end before first injection then the missing start day/month is set to 1, if applicable, and start time (if not available) is set to 23:59, if not after end date/time (else start time is set to end time) (e.g., start date of AE is 2008-11, end date of AE is 2009-02-01, end time is 22:00, and first injection is 2008-12-06 T17:35 → imputed start date and time are 2008-11-01 23:59). Set imputation flags accordingly.

2a) If the start date is missing completely and end date and time* do not imply end before first injection then imputed start date is set to date of first injection and set start time (if not available) to 23:59, if not after end date/time (else start time is set to end time) (e.g., start date of AE is missing, end date of AE is 2008-12-31, end time is 23:00, and first injection date is 2008-12-06T17:35 → imputed start date and time are 2008-12-06 23:59). Set imputation flags accordingly.

2b) If the start date is partially missing and the non-missing part is the same as in first injection date and the end date and time* do not imply end before first injection, then the imputed start date is set to the first injection date and start time (if not available) to 23:59, if not after end date/time (else start time is set to end time) (e.g., start date of AE is 2008-

12, end date of AE is 2008-12-31, end time is 23:00, and first injection date is 2008-12-06T17:35 → imputed start date and time are 2008-12-06 23:59). Set imputation flags accordingly.

3) If the start date is partially missing and imputation rule 1b and 2b do not fit, then missing day/month is set to 1, if applicable and start time (if not available) is set to 23:59, if not after end date/time (else start time is set to end time) (e.g., start date of AE is 2009-02, end date of AE is 2009-02-01, end time is 22:00, and first injection is 2008-12-06 T17:35 → imputed start date and time are 2009-02-01 22:00). Set imputation flags accordingly.

End date or end time will not be imputed.

* The end date and time may be complete, partially missing, or completely missing.

For any missing data of adverse events, a worst case strategy will be applied for all analysis tables. The intensity will be imputed by the worst intensity "severe", the causal relationship will be imputed by the worst relationship "drug related". Missing data of the worst outcome will be imputed by "unknown".

[redacted]

### *7.4.2 Timing of statistical analyses*

The time from the baseline visit to the end of study is thirteen months. The primary effectiveness endpoint is at Month 2 post last injection (i.e. either baseline treatment or touch-up, if applicable); however, treated subjects will continue to be followed for safety via in clinic visits at Months 3, 6, 9, and 12 after the last injection (i.e., either baseline treatment or touch-up, if applicable). For the purposes of the clinical study report (CSR), statistical analyses will be performed after all subjects have completed the Month 2 post last injection visit. At this time, the data will be cleaned and the database temporarily locked for all effectiveness and safety analyses corresponding to the Month 2 primary effectiveness endpoint. A CSR will be written for these data and an addendum will be made to the report based on additional safety analyses associated with the safety monitoring for treated subjects up to 12 months after the last injection (i.e., either baseline treatment of touch-up, if applicable).

### *7.4.3 Examination of Subgroups*

The primary effectiveness (ITT using observed cases) and secondary effectiveness (ITT using observed cases) endpoints will be descriptively summarized after stratification into the following subgroups:

- Fitzpatrick Skin types (I-III versus IV-VI)
- Gender (males versus females)

The safety endpoints for incidences of TEAEs using the SP will be descriptively summarized after stratification in the aforementioned subgroups.

There will be no statistical comparisons made for all subgroup analyses.