NCT04647721, redacted
version v1.0, 01Nov2022

Statistical Analysis Plan

Title:	Evaluation of the Effectiveness and Safety of Radiesse for the Correction of Moderate to Severe Nasolabial Folds
Merz Study Number:	M900311009
SAP for	Final Analysis
Sponsor:	Merz North America, Inc.
Version:	1.0
Date:	21-JUL-2022
Author:	[redacted]

Document References (For Internal Use Only)	
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[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 database close.

[redacted]		
Author [redacted] (print name)	Date (dd-mmm-yyyy)	Signature

[redacted]		
Peer Reviewer [redacted] (print name)	Date (dd-mmm-yyyy)	Signature

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

[redacted]		
[redacted] Merz (print name)	Date (dd-mmm-yyyy)	Signature

[redacted]		
[redacted] Merz (print name)	Date (dd-mmm-yyyy)	Signature

Table of Contents

List of Abbreviations

Abbreviation	Definition
ADaM	Analysis Data Model
AE	Adverse event
ANCOVA	Analysis of covariance
ATC	Anatomical Therapeutic Chemical classification system of the World Health Organization
BMI	Body mass index
CaHA	Calcium hydroxylapatite
CI	Confidence interval
CSP	Clinical study protocol
[redacted]	
eCRF	Electronic case report form
EOT	End of Trial
FAS	Full analysis set
[redacted]	
GAIS	Global Aesthetic Improvement Scale
HA	Hyaluronic acid
ICH	International Council for Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use
iGAIS	Investigator Global Aesthetic Improvement Scale
IMD	Investigational medical device
[redacted]	
max	maximum
MedDRA	Medical Dictionary for Regulatory Activities
min	minimum
n	Number of values analyzed
NLF	Nasolabial fold
OC	Observed cases
PD	Protocol deviation
PEV	Primary endpoint visit
PPS	Per protocol set
[redacted]	
PT	Preferred term
Q1	Lower quartiles
Q3	Upper quartiles

SAE	Serious adverse event
SAP	Statistical analysis plan
SAS	Statistical Analysis Software®
SD	Standard deviation
SDTM	Study Data Tabulation Model
SES	Safety evaluation set
sGAIS	Subject Global Aesthetic Improvement Scale
SOC	System organ class
SOP	Standard operating procedure
TEAE	Treatment-emergent adverse event
TESAE	Treatment-emergent serious adverse event
VAR	SAS variable
WSRS	Wrinkle Severity Rating Scale

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) M900311009, version 2.0, dated 27 Nov 2020.

All programs will be written using SAS version 9.4. A preferred font size of 10 points for the whole document will be used for the tables and figures in Section 14, corresponding to a linesize of 111 digits and a pagesize of 42 lines for an output in A4 landscape format. For listings, a standard font size of 10 points with the linesize and pagesize 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. All outputs will be transferred into PDF files using the Merz internal SAS macro LST2PDF. These PDF files will be generated as needed to populate the subsections of Section 14 and Section 16.2 for the clinical trial report. Each output file will include the corresponding table of contents, preceding the content of the file.

[redacted]Table, Figures, and Listings (TFLs) for medical devices, version 2.0 will be applied and adapted to trial specific requirements as laid down in the CSP M900311009, version 2.0 and any amendments. These mock TFLs will serve as study-specific output specifications for statistical programming.

Log review, output review, and double programming are planned to be performed as quality control measures for statistical programming of SDTMs, ADaMs, and TFLs.

2 Clinical Trial Design and Objectives

2.1 Clinical Trial Design

This is a 48-week, prospective, randomized, multicenter, split-face, active-comparator, blinded [redacted] trial designed to evaluate the effectiveness and safety of Radiesse compared to Restylane for the correction of moderate to severe NLFs in healthy adults. Subjects will be enrolled from participating investigative sites in China.

[redacted]

After a maximum 10-day screening period, the planned trial duration for individual subjects is 48 weeks (± 7 days) after last injection.

[redacted]

Please also refer to the CSP for details on subject randomization (CSP Section 13.2) and blinding procedures including planned and emergency unblinding (CSP Section 10.4).

2.2 Trial Objectives

Effectiveness

The primary objective is to demonstrate non-inferiority of Radiesse (CaHA) to Restylane (HA) following subdermal implantation for the correction of moderate to severe facial wrinkles and folds, such as NLFs.

Safety

The safety objective includes an evaluation of the incidence and type of adverse events (AEs) and serious adverse events (SAEs).

3 Determination of Sample Size

The sample size was calculated separately for safety and for the primary effectiveness endpoint as follows:

3.1 Safety

The safety aim of the trial is to generate a database with short-term and sufficient long-term safety data. Based on the International Council on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) E1 guideline, 100 subjects treated and observed over 48 weeks is deemed sufficient. [redacted]

3.2 Effectiveness

The sample size considerations for the primary effectiveness endpoint (i.e., change from baseline to Week 24 after last injection on the WSRS [redacted]) are based on the following:

- Type I error, $\alpha = 0.05$ (two-sided) and
- Non-inferiority margin of 0.5.

[redacted]

When using the 118 subjects randomized for safety and applying a drop-out rate of 15% from randomization to the Per Protocol Set (PPS), [redacted] 100 subjects would be available for analysis of the primary effectiveness endpoint on PPS.

For analysis on the Full Analysis Set (FAS) even more subjects are expected. [redacted]

[redacted]

To ensure a reasonable number of subjects is enrolled per investigative site, the number of subjects treated at each site should not exceed 20 in case of seven recruiting sites, 24 in case of six recruiting sites, and 28 in case of five recruiting sites.

4 Analysis Sets

The following analysis sets will be defined:

- The randomized set is defined as all subjects randomized into the trial.
- The Safety Evaluation Set (SES) is defined as all subjects who are randomized and receive investigational product at least once.
- The Full Analysis Set (FAS) is defined as all subjects who are randomized, receive investigational product, and have at least one after-baseline effectiveness assessment.
- The Per Protocol Set (PPS) is the subset of subjects in the FAS who have no major protocol deviations or other events that impact analysis of the primary effectiveness endpoint. Final determination of which events lead to exclusion from PPS will be made prior to database close.

Analyses on the randomized subjects set, the FAS and the PPS will be performed *by randomized treatment*. For subjects randomized to "Group 1: left NLF: Radiesse, right NLF: Restylane", the left NLF will be analyzed for Radiesse and the right NLF will be analyzed for Restylane, and for subjects randomized to "Group 2: left NLF: Restylane, right NLF: Radiesse", the left NLF will be analyzed for Restylane and the right NLF will be analyzed for Radiesse.

[redacted]

5 Endpoints for Analysis

5.1 Effectiveness Endpoints

For details on effectiveness assessments, see CSP Section 11.2.1. For definition of baseline see Section 6.4.2.

5.1.1 Primary Effectiveness Endpoint

- Change from baseline to Week 24 after last injection on the Wrinkle Severity Rating Scale (WSRS)[redacted].

  The primary endpoint will be calculated as follows based on WSRS assessments[redacted] at Week 24 for the respective NLF: WSRS score at Week 24 after last injection –WSRS score at baseline.

  [redacted]

5.1.2 Secondary Effectiveness Endpoints

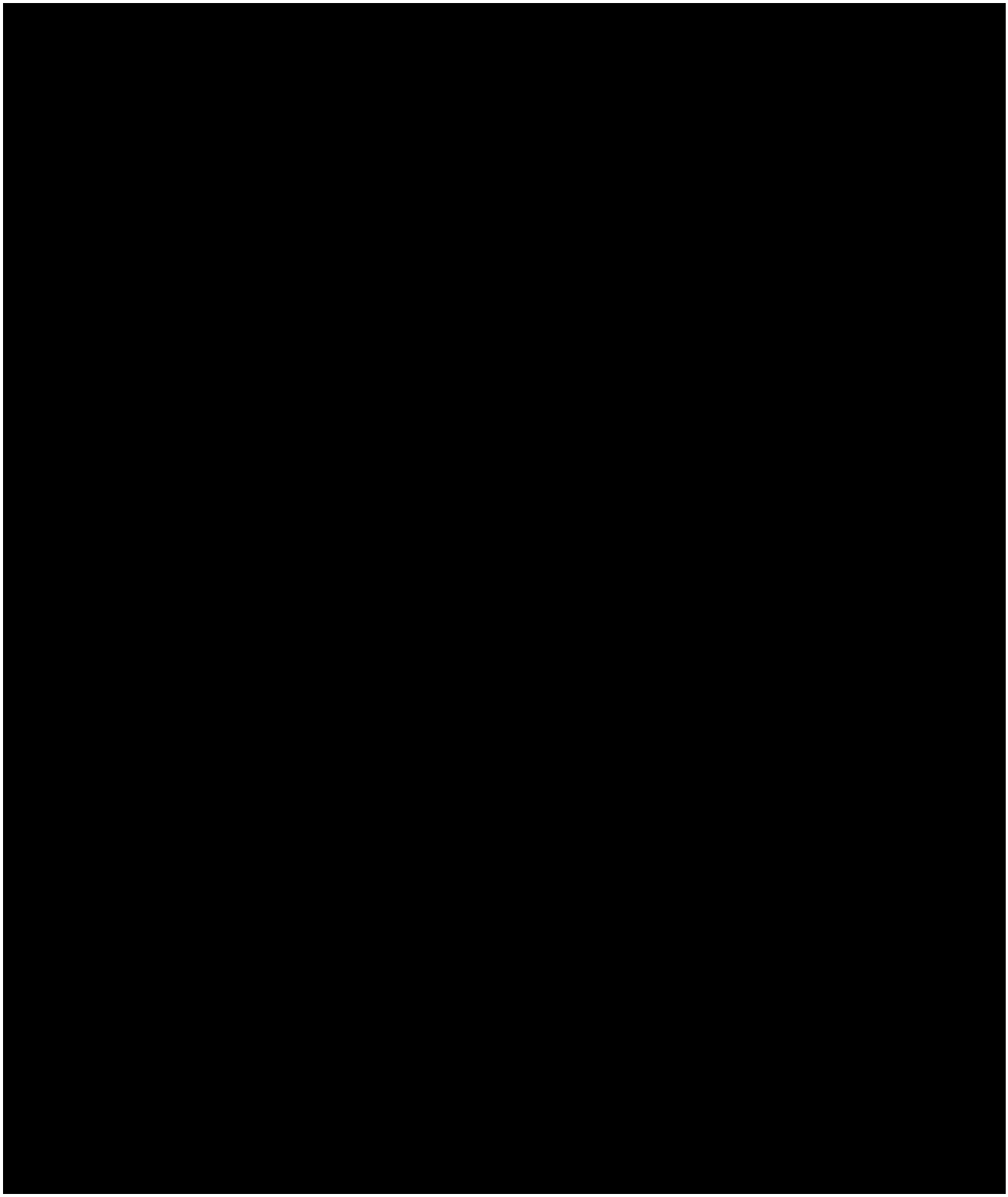
- Proportion of subjects with any improvement, defined as a rating of + 1, + 2, or + 3, on the Investigator Global Aesthetic Improvement Scale (iGAIS) for each NLF at Week 24 after last injection, as assessed by the treating investigator.

  This dichotomous endpoint will be determined based on iGAIS ratings at Week 24 for the respective NLF. It will be set to "Y" for ratings ≥ + 1 and to "N" for ratings < + 1.

- Proportion of subjects with any improvement, defined as a rating of + 1, + 2, or + 3, on the Subject Global Aesthetic Improvement Scale (sGAIS) for each NLF at Week 24 after last injection, as assessed by the masked subject.

  This dichotomous endpoint will be derived from sGAIS ratings at Week 24 for the respective NLF. It will be set to "Y" for ratings ≥ + 1 and to "N" for ratings < + 1.

Calculation of proportions (relative frequencies) is described in Section 6.



5.2 Safety Endpoints

For details on safety assessments see CSP Section 11.2.2. For safety definitions and requirements, including definition of an adverse event (AE), severity greading for an AE, causal relationship of an AE with an Investigational medical device (IMD), outcome categories for an AE, definition of a serious adverse event (SAE), see CSP Section 12.

Treatment-emergent AEs (TEAEs) are defined as AEs with onset at or after the first administration of trial treatment. If an AE starts prior to treatment but worsens at or after treatment, the investigator records this observation as a new AE with onset = time of worsening.

AEs not constituting TEAEs will be referred to as non-TEAEs.

For details on separation of TEAEs and non-TEAEs see Appendix, Section 9.3.

[redacted]

End date or end time or AEs will not be imputed.

[redacted]

A device deficiency is defined as an inadequacy of a medical device with respect to its identity, quality, durability, reliability, safety, or effectiveness. Note: Device deficiencies include malfunctions, use errors, and inadequate labeling. For more information on device deficiencies, see CSP Section 12.7.

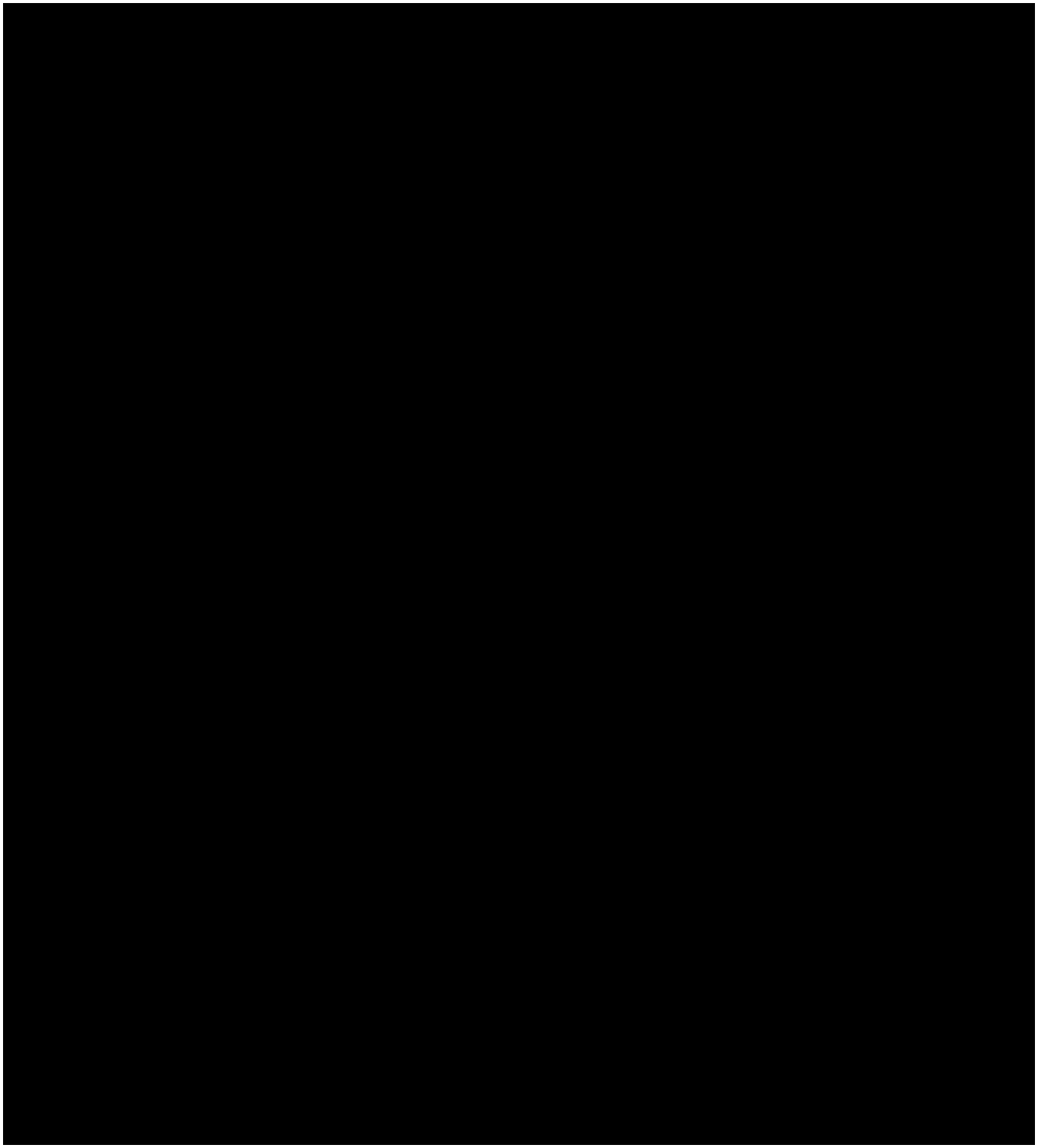
5.2.1 Primary Safety Endpoint

Not applicable.

5.2.2 Secondary Safety Endpoint

- Incidence of TEAEs related to Radiesse, as reported by the treating investigator throughout the trial.

[redacted]



6 Statistical Analysis Methods

Metric statistics for continuous endpoints will be number of values analyzed (n), number of observed values (Nobs), mean, standard deviation (SD), median, lower quartile (Q1), upper quartile (Q3), minimum (min), and maximum (max), if not stated otherwise. 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 dispayed; for derived data, the number of decimal places will explicitely be given in the sections below.

[1] Injection depth is included in list of additional data collected (Section 11.2.3 of CSP version 2.0) but is not analyzed because these data are not collected in the case report form.

Frequency tables for qualitative endpoints will be absolute and percent frequencies (n, %). Cross tables will include number of subjects with non-missing data at the corresponding visit.Percentages will be calculated using the denominator of all subjects in a specified population or random group (only needed for selected disposition tables in 14.1). For TEAE analyses on the SES, which will be conducted by actual treatment (see Section 4), the denominator for calculation of percentages per treatment will be adapted to the number of subjects in SES treated with the respective treatment. For analyses on observed cases (OC), percentages will be calculated using the number of subjects with non-missing data (Nobs) at the corresponding visit as denominator. The denominator will be specified in a footnote to the tables for clarification if not otherwise obvious.

The total number of subjects in a given analysis set (as defined in Section 4) will be referred to with N.

Percentages will be reported to one decimal place. If not otherwise specified, statistical tests will be conducted two-sided at type I error rate 5%. *P* values will be reported to four decimal places (e.g., p=0.0375). *P* values below 0.0001 will be presented as '<0.0001'. *P* values above 0.9999 will be presented as '>0.9999'. Confidence intervals (CIs) will be two-sided 95% CIs.

Detailed information on effectiveness, safety, and other data will be provided in listings.

All statistical analyses will be performed using Statistical Analysis System (SAS) statistical analysis software (Version 9.4).

6.1 Effectiveness Endpoints

Effectiveness endpoints are defined in Section 5.1 [redacted]. Effectiveness data assessed at Week 4 post first injection will be listed only. There will be one confirmatory analysis of the primary effectiveness endpoint. [redacted]

6.1.1 Primary Effectiveness Endpoint

6.1.1.1 Primary analysis

If the expected mean intra-individual difference between Radiesse and Restylane in WSRS change from baseline to Week 24 is denoted by δ, the null and alternative hypotheses to be tested are:

$$H_0\text{: } \delta \geq 0.5 \text{ (null hypothesis) versus } H_1\text{: } \delta < 0.5 \text{ (alternative hypothesis)}$$

with 0.5 being the non-inferiority margin. To test the null hypothesis, a two-sided 95% CI will be constructed around the estimate for δ. H_0 will be rejected if the upper bound of the two-sided 95% CI lies below the non-inferiority margin of 0.5. It will then be concluded that Radiesse is non-inferior to Restylane and is effective. The two-sided 95% CI for δ will be constructed by applying a repeated-measures analysis of covariance (ANCOVA) model to the WSRS change from baseline to Week 24 after last injection. This model will include treatment (i.e., Radiesse or Restylane), investigative site, touch-up performed (i.e., yes/no), and random group (i.e., "Group 1: left NLF: Radiesse, right NLF: Restylane" or "Group 2: left NLF: Restylane, right NLF: Radiesse")[2] as fixed class effects and baseline WSRS [redacted] as a fixed covariate. That observations are dependent over treatment (due to the split-face design) will be modelled using treatment as a repeated factor. The SAS procedure "proc mixed" will be used with the covariance

[2] The term "randomized sequence" as used in Section 13.4.1.1 of the protocol was replaced by "random group", as this is the more appropriate term for split-face design studies. Also, the two categories (group 1 and group 2) were relabeled by using the labels from protocol Sections 10.3 and 13.2. These changes in terms/ category lables do not constitute any change in planned analyses since the new terms/labels refer to the same variable/ categories.

matrix for treatments left "unstructured". The latter allows estimation of different variances for the two treatments.

The least squares means (lsmeans) statement will be used to derive a point estimate for δ as well as the associated two-sided 95% CI.

[redacted]

[redacted]

- In addition, the mean treatment difference in WSRS change from baseline to Week 24 after last injection with the associated 2-sided 95% Wald confidence interval will be presented [redacted]. The calculation will be performed as follows. First, the treatment difference on the primary endpoint will be calculated for each subject as WSRS change from baseline to Week 24 after last injection at NLF randomized to Radiesse – WSRS change from baseline to Week 24 after last injection at NLF randomized to Restylane. Second, the mean and further metric statistics statistics for continuous endpoints as specified in first paragraph of Section 6 will be calculated for the treatment difference in WSRS change from baseline to Week 24 after last injection. [redacted]

6.1.1.3 Subgroup analyses

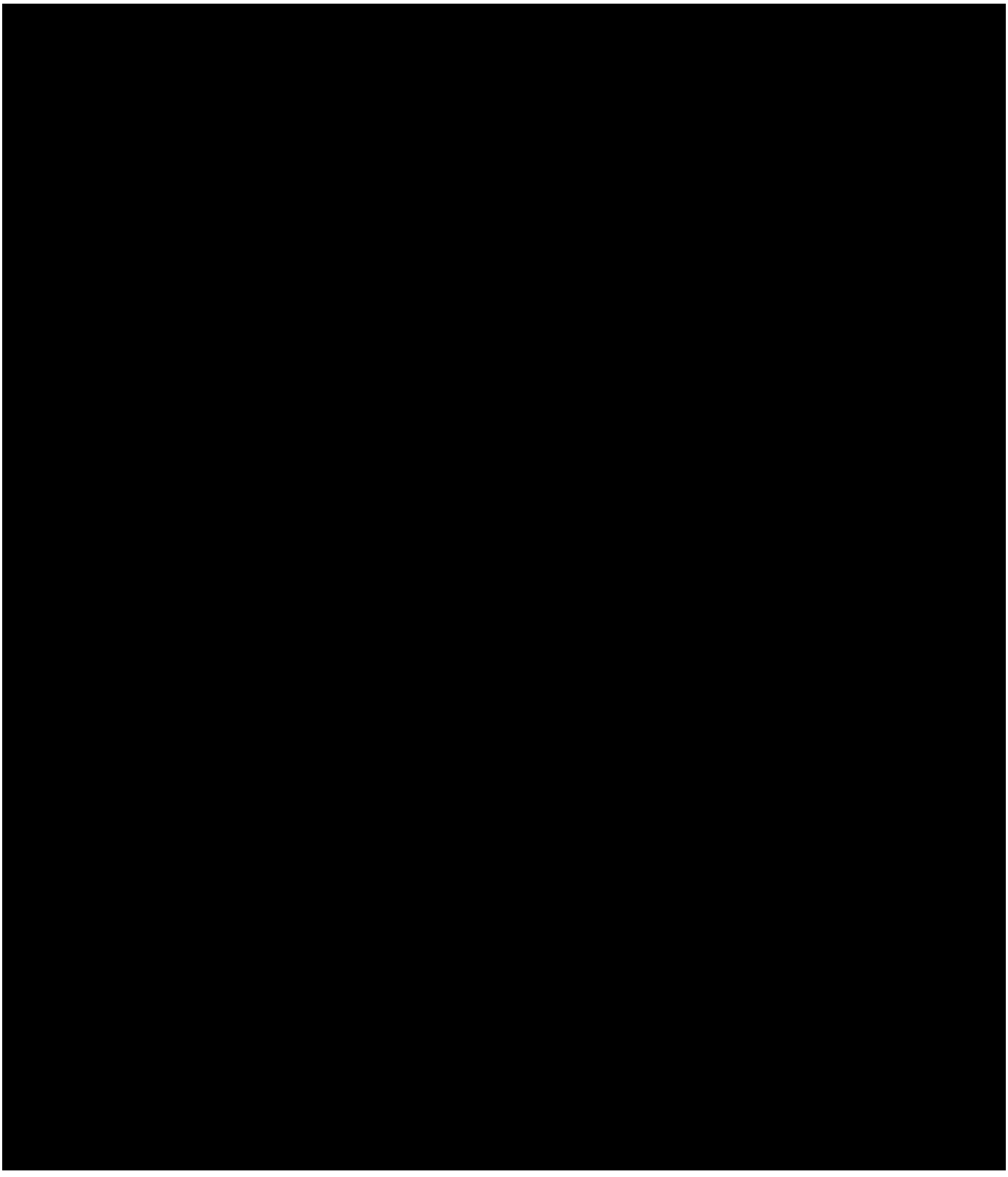
No subgroup analysis is planned for primary effectiveness endpoint.

6.1.2 Secondary Effectiveness Endpoints

Secondary effectiveness endpoints will be analyzed based on observed cases in the PPS and on observed cases in the FAS.

For both GAIS scores (iGAIS and sGAIS), the proportion of subjects with any improvement at Week 24 will be calculated by treatment and given as number of subjects with any

improvement/number of subjects in the corresponding analysis (n/Nobs[3]) and as percentage (%). The 2-sided 95% exact Pearson Clopper CI for the response rate will be reported as well. In addition, the crude difference of the two treatment response proportions will be stated along with the associated two-sided 95% Newcombe-Wilson CIs for paired data [Method No. 10 in Newcombe (1998; [2])]. Cross tables will be provided showing number (and percentage) of subjects with any (or no) GAIS improvement of the Radiesse side versus any (or no) GAIS improvement of the Restylane side. For definition of "any improvement" on iGAIS and sGAIS see Section 5.1.2.



6.2 Safety Endpoints

Safety endpoints are defined in Section 5.2. Safety analyses will be analyzed for the SES if not stated otherwise.

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

Only TEAEs will be analyzed. Non-TEAEs will be listed only. For defnitions of TEAEs and non-TEAEs see Section 5.2.

The investigator will report for each TEAE whether only the right NLF, only the left NLF, or both NLFs were affected or whether the treatment area was not affected.

Incidences of TEAEs will be presented in frequency tables. TEAEs affecting the treatment area (either or both NLF) will be analyzed by actual treatment of the respective NLF. Further, incidences of TEAE not affecting the treatment area (i.e. not affecting any NLF) and incidences of TEAEs in total will be presented. The total column will not only include TEAEs affecting either or both treatment NLFs but also TEAEs not affecting the treatment area. [redacted]

Listings and, if applicable, tables displaying incidences for TEAEs leading to discontinuation, TESAEs, and deaths will also be provided.

[redacted]

[redacted]

Further details on analyses of TEAEs are presented in Section 6.2.1 and Section 6.2.3. [redacted]

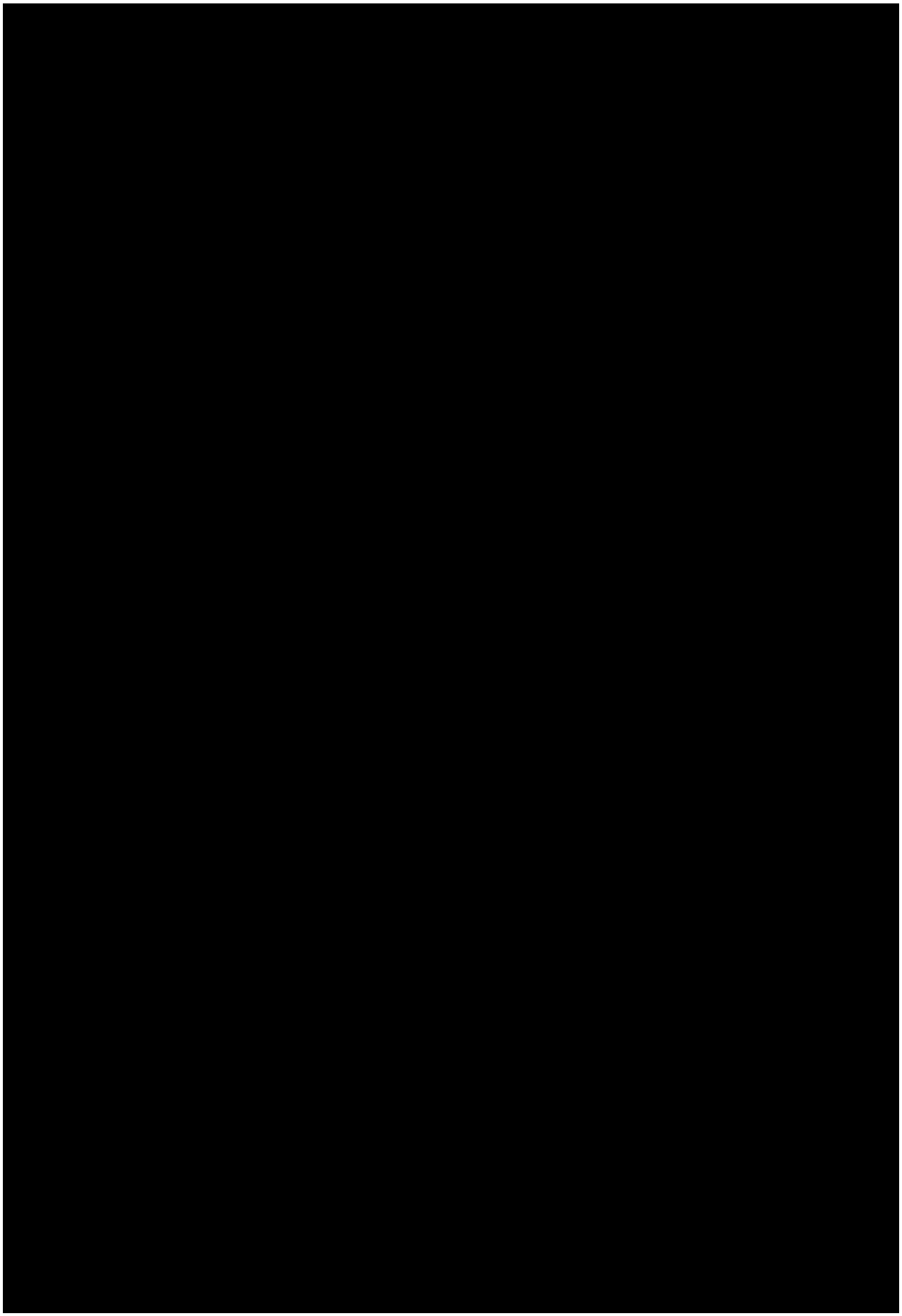
6.2.1 Primary Safety Endpoint

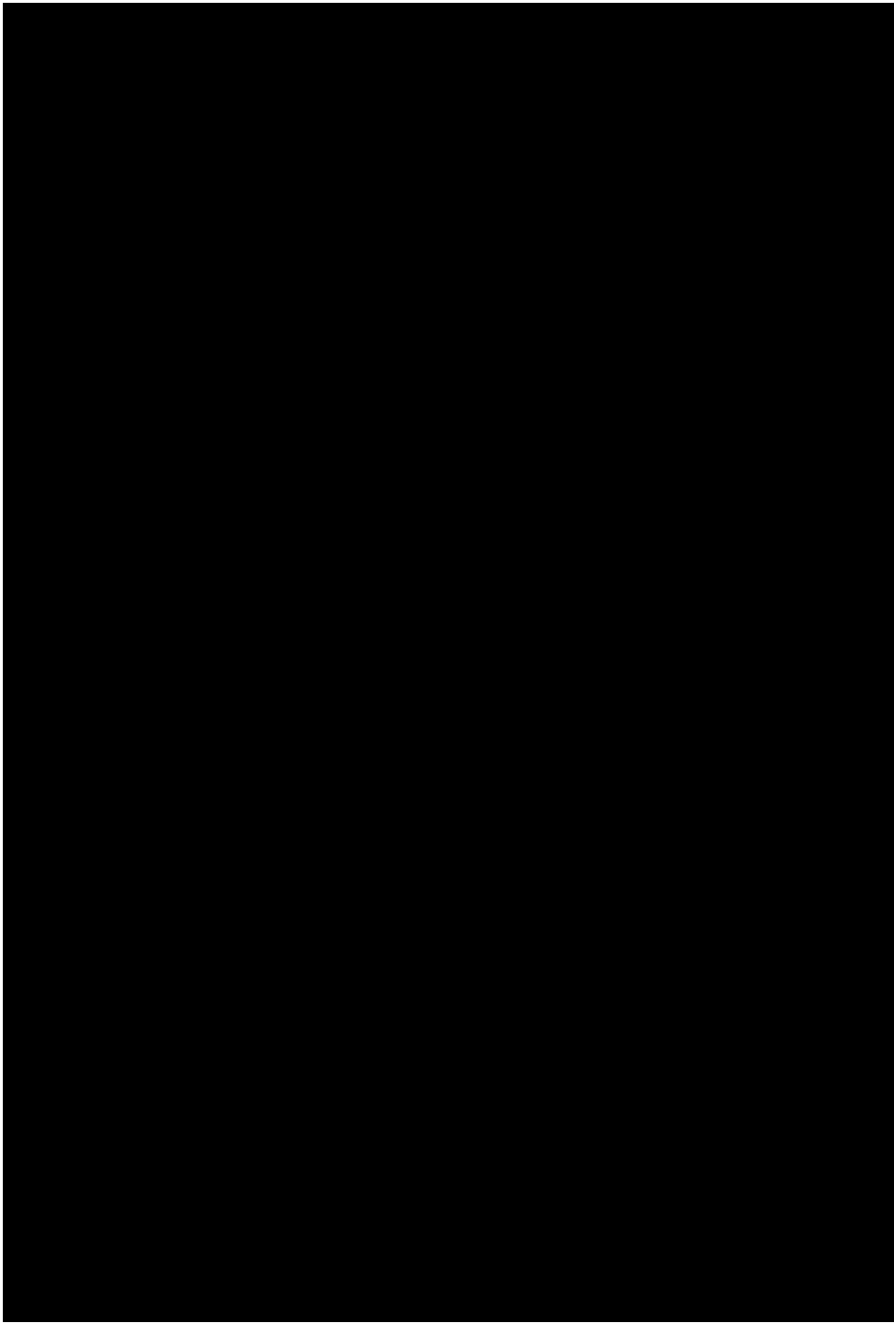
Not applicable.

6.2.2 Secondary Safety Endpoints

Incidences of TEAEs related to Radiesse will be calculated at the SOC level and at the PT level and will be presented in frequency tables (for specification of columns see first part of Section 6.2 above). [redacted]

[redacted]





Protocol deviations and other reasons for exclusion from analysis sets will be summarized in a frequency table, by random group and in total. All protocol deviations will be listed.

Demographic data and other baseline characteristics will be summarized for the SES, FAS, and PPS. If the "randomized subjects set" is not equal to the SES, demographic data and other baseline characteristics will also be summarized for the "randomized subjects set". Detailed information will be provided in listings.

Medical history, concomitant diseases and prior and concomitant non-drug treatment will be coded using the MedDRA dictionary and reported in frequency tables by SOC and PT levels for the SES. Frequencies of prior and concomitant medication will be given based on different Anatomic Therapeutic Chemical (ATC) code levels for the SES.

Indications for prior and concomitant therapies will not be coded and will only be listed.

[redacted]Frequency tables will be used to analyze injection technique. These calculations will be based on the SES. Injection location will not be analyzed but listed by subject and treatment visit. In these listings, it will be displayed as “treatment area and side” (left NLF/ right NLF).

[redacted]

6.4 Special Statistical/Analytical Issues

[redacted]

6.4.2 Baseline

Baseline is defined as the last observed value or test result before the initial injection of IMDs. For efficacy endpoints based on WSRS, this corresponds to the score assessed at Screening visit V1.

6.4.3 Interim Analyses

No interim analysis is planned for this study.

6.4.4 Multiple Comparisons/Multiplicity

No multiplicity adjustments are required. Only one confirmatory analysis of the primary effectiveness endpoint will be performed. All other analyses are for exploratory or sensitivity purposes.

[redacted]

6.4.6 Pooling of Sites

No pooling of sites is planned for this study.

[redacted]

8 References

[1] Li D, Sun J, Wu S. A multi-center comparative efficacy and safety study of two different hyaluronic acid fillers for treatment of nasolabial folds in a Chinese population. J Cosmet Dermatol. 2019; 18 (3): 755-61.

[2] Newcombe, R.G. (1998). Improved confidence intervals for the difference between binomial proportions based on paired data. *Statistics in Medicine,* 17, 2635-2650.

9 Appendix