Statistical Analysis Plan

Title:	A prospective, multicenter, randomized, comparator-controlled, evaluator-blinded study to evaluate the safety and effectiveness of Belotero® Intense Lidocaine for lip augmentation
Merz Study Number:	M930041001
SAP for	Final Analysis
Sponsor:	Merz North America, Inc.
Version:	1.0
Date:	27-NOV-2024
Author:	PPD Biostatistician, PPD (PPD)

CONFIDENTIAL AND PROPRIETARY

The contents of this document are confidential and proprietary of Merz North America, Inc.

Unauthorized use, disclosure or reproduction is strictly prohibited. This document or parts thereof may not be disclosed to parties not associated with the clinical investigation without the prior written consent of Merz North America, Inc.

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.



Date (dd-mmm-yyyy)	Signature
Date (dd-mmm-yyyy)	Signature
Date (dd-mmm-yyyy)	Signature

Table of Contents

1 General and Technical Aspects 6
2 Clinical Trial Design and Objectives 7
2.1 Clinical Study Design 7
2.2 Clinical Study Objectives 7
3 Determination of Sample Size 8
4 Analysis Sets 9
5 Endpoints for Analysis 10
5.1 Effectiveness Endpoints 10
5.1.1 Primary Effectiveness Endpoint 10
5.1.2 Co-primary Effectiveness Endpoint 10
5.1.3 Secondary Effectiveness Endpoints 10
[redacted] 10
5.2 Safety Endpoints 12
5.2.1 Primary Safety Endpoint 12
5.2.2 Secondary Safety Endpoint 12
[redacted] 12
[redacted] 12
6 Statistical Analysis Methods 13
6.1 Effectiveness Endpoints 13
6.1.1 Primary and Co-primary Effectiveness Analysis 13
6.1.2 Secondary Effectiveness Endpoints 17
[redacted] 18
6.2 Safety Endpoints 19
6.2.1 Primary Safety Endpoints 20
6.2.2 Secondary Safety Endpoints 20
[redacted] 20
[redacted] 22
6.4 Special Statistical/Analytical Issues 26
6.4.1 Discontinuations and Missing Data 26
6.4.2 Interim Analyses 26
6.4.3 Multiple Comparisons/Multiplicity 26
6.4.4 Examination of Subgroups 26
6.4.5 Pooling of Sites 26
[redacted] 27
8 References 28
9 Appendix 29
[redacted] 29
..... 29
..... 29
..... 30
..... 31
..... 32
..... 33

[redacted] 34

List of Abbreviations

AE	Adverse event
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
ATC	Anatomical Therapeutic Chemical
BIL	Belotero Intense Lidocaine
BMI	Body mass index
CI	Confidence interval
CONSORT	Consolidated Standards of Reporting Trials
CSP	Clinical study protocol
CSR	Clinical study report
CTR	Common treatment response
eCRF	Electronic case report form
EOS	End of Study
FCS	Fully conditional specification
FST	Fitzpatrick skin type
GAIS	Global Aesthetic Improvement Scale
HA	Hyaluronic acid
IMD	Investigational medical device
IRP	Independent panel reviewers
ITT	Intent to Treat
LS	Least square
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputations
MLFAS	Merz Lips Fullness Assessment Scale
PPS	Per protocol set
PT	Preferred Term
RS	Restylane Silk
SAP	Statistical analysis plan
SAS	Statistical Analysis System
SES	Safety evaluation set
SOC	System Organ Class
TEAE	Treatment-emergent adverse event
TFL	Tables, Figures and Listings
VAR	Variables
WHO-DD	World Health Organization Drug Dictionary

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) M930041001 (version 2.0), dated 14-NOV-2022.

All programs will be written using Statistical Analysis System® (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 or directly created as PDF files. These PDF files will be generated as needed to populate the subsections of Section 14 and Section 16.2 for the clinical study report (CSR). Separate documents will be created for sections 14.1, 14.2, 14.3 and sections 16.2.1 to 16.2.8. 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 medical devices, version 2.0 will be applied and adapted to study-specific requirements as laid down in the CSP M930041001 (version 2.0) and adapted according to some changes of version 3.0 of Merz TFLs for medical devices. These mock TFLs will serve as study-specific output specifications for statistical programming of TFLs.

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 analysis datasets and TFLs) will be documented in the quality control plan for the creation of statistical output.

2 Clinical Trial Design and Objectives

2.1 Clinical Study Design

This is a prospective, multicenter, randomized, comparator-controlled, evaluator-blinded study to evaluate the safety and effectiveness of Belotero Intense Lidocaine (BIL) for lip augmentation in subjects with upper and/or lower lip deficit on the Merz Lips Fullness Assessment Scale (MLFAS).

A total of 220 subjects with a grade of 0 (very thin), 1 (thin), or 2 (medium) on the MLFAS on each lip will be enrolled at 12 to 14 sites in the US. For this study, randomization will be block-stratified by investigational site. Enrolled subjects will be randomized using a 3:1 ratio, to one of two treatment groups:

- BIL: n=165 or
- Restylane Silk (RS): n=55.

A minimum of 20% of randomized subjects will be Fitzpatrick Skin Type (FST) IV, V, or VI. In this subset, subjects will be distributed as follows: ≥ 22 subjects of FST IV and ≥ 22 subjects of FST V or VI. Approximately 5% of the subjects enrolled will be male.

Subjects eligible for study enrollment were those seeking lip augmentation for correction of volume deficit in at least one lip for FST V or VI subjects and in both lips for FST I - IV subjects. Further, FST I - IV subjects were to have upper **and** lower lip deficit of 0 (very thin), 1 (thin), or 2 (medium) on the MLFAS, while FST V or VI subjects were to have upper **and/or** lower lip deficit 0 (very thin), 1 (thin), or 2 (medium) on the MLFAS.

All subjects will receive an initial treatment at baseline (V1-a) and then will be assessed for a touchup treatment at Week 4 (V2-a), if applicable.

A touch-up injection is required when the subject fails to show a 1-point improvement in one or both lip(s) on the MLFAS, as assessed by the blinded evaluator four weeks post-treatment compared to baseline in at least one of the lips. If a 1-point improvement is reached, but the subject does not achieve optimal aesthetic correction, according to the treating investigator, a touch-up injection may be performed at the discretion of the investigator.

Subjects randomized to BIL will have the option of retreatment at 48-weeks post last injection (i.e., either baseline treatment or touch-up, if applicable). RS subjects will not be offered retreatment. All BIL subjects, irrespective of retreatment status, will be followed for a total study duration of up to 64 weeks. Subjects treated with RS will be followed for 48-weeks post last injection, for a total study duration of up to 52 weeks.

Subjects enrolled will have a screening period of up to 10 days; thereafter, they will participate for up to 64 weeks after baseline depending on the treatment group.

For both BIL and RS, the volume to achieve optimal aesthetic correction is at the treating investigator's discretion. The maximum recommended dose per treatment is 1.5 mL per lip.

2.2 Clinical Study Objectives

The primary objective is to confirm the effectiveness of BIL for lip augmentation by demonstrating its non-inferiority versus an approved hyaluronic acid (HA) dermal filler following submucosal and/or subcutaneous injection.

The secondary objective is to demonstrate safety of BIL for single and repeat use in lip augmentation.

3 Determination of Sample Size

To provide a sufficient exposure in terms of safety, it is intended to treat n = 165 subjects with BIL. In 165 subjects, an adverse event (AE) with a true incidence of 1.5% will be observed at least once with a probability of 91.7%. It has been decided to randomize with a ratio of 3:1 (i.e., 55 subjects to RS) and 220 subjects in total.

The planned study is pivotal. Therefore, at least 90% power for effectiveness is anticipated. It is intended to demonstrate non-inferiority with a margin of δ = -0.5 points and a one-sided type I error rate of 2.5%. Further, up to 10% of randomized subjects might need to be excluded from the per-protocol set (PPS) due to important protocol deviations in statistical sense or other reasons. CCI [redacted] CCI For this calculation, the Statistical Analysis Software (SAS) procedure PROC POWER (SAS 9.4) was used.

Further, it is intended to show that the response rate of the BIL group is statistically significantly larger than 50%. Sample-size estimation for this test is performed with a one-sided exact binomial test at an error level of 2.5%. The binomial test is applied to the treatment Group BIL only. CCI [redacted] CCI [redacted] For testing this hypothesis, 88 evaluable subjects in the BIL Group in the PPS would reach 100% power. For this calculation, the NQuery version 8.5 was used.

A hierarchical-testing procedure will be applied for co-primary and primary effectiveness endpoints to keep the overall one-sided type I error rate of α = 0.025, and no adjustment for multiplicity is necessary.

In summary, the sample size of n=220 randomized subjects, with a randomization ratio of 3:1 leading to at least 198 in the PPS analysis, provides sound balance between power for effectiveness and ability for detection level of safety risks.

4 Analysis Sets

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

Safety Evaluation Set (SES)

The SES is the subset of all subjects who were exposed to the investigational medical device (IMD) or comparator device at least once. Subjects in the SES analyses will be analyzed as treated.

Intent to treat (ITT)

The ITT is the set of all randomized subjects. Subjects in the ITT analyses will be analyzed as randomized.

Per protocol set (PPS)

The PPS is the subset of subjects in the ITT without important deviations in statistical sense or other reasons for exclusion. The following protocol deviations and reasons for exclusion will lead to exclusion from the PPS:

- No treatment
- Deviation from randomized treatment
- Missing primary effectiveness endpoint
- Unblinding of blinded evaluator before or at Visit V3
- Visit window deviation at Visit V3 of more than 14 days
- Treatment with hyaluronidase between Screening and Visit V3
- Cosmetic plastic surgery in the lower third part of the face or oral surgery between Screening and Visit V3
- Facial treatments with permanent and/or semi-permanent dermal fillers in the lower face between Screening and Visit V3

The PPS will be the primary population used for effectiveness analyses, due to the non-inferiority study design. Subjects in the PPS analyses will be analyzed as randomized.

5 Endpoints for Analysis

5.1 Effectiveness Endpoints

5.1.1 Primary Effectiveness Endpoint

Subject's mean change of treated lips from baseline to Visit V3, according to the MLFAS, as assessed live by a blinded evaluator.

- Mean is calculated at the subject level.
- Assessments conducted at screening will be used as baseline values during analysis.
- If a single lip was treated, only its MLFAS values will be used for analysis. For each subject, the primary effectiveness endpoint will be derived as follows:
    - First, the change from baseline to Visit V3 for each individual lip will be calculated.
    - Second, for every subject treated in both lips, the average of change for upper and lower lip will be calculated. For every subject treated in only one lip, the change for the treated lip will be regarded as the primary effectiveness endpoint.

5.1.2 Co-primary Effectiveness Endpoint

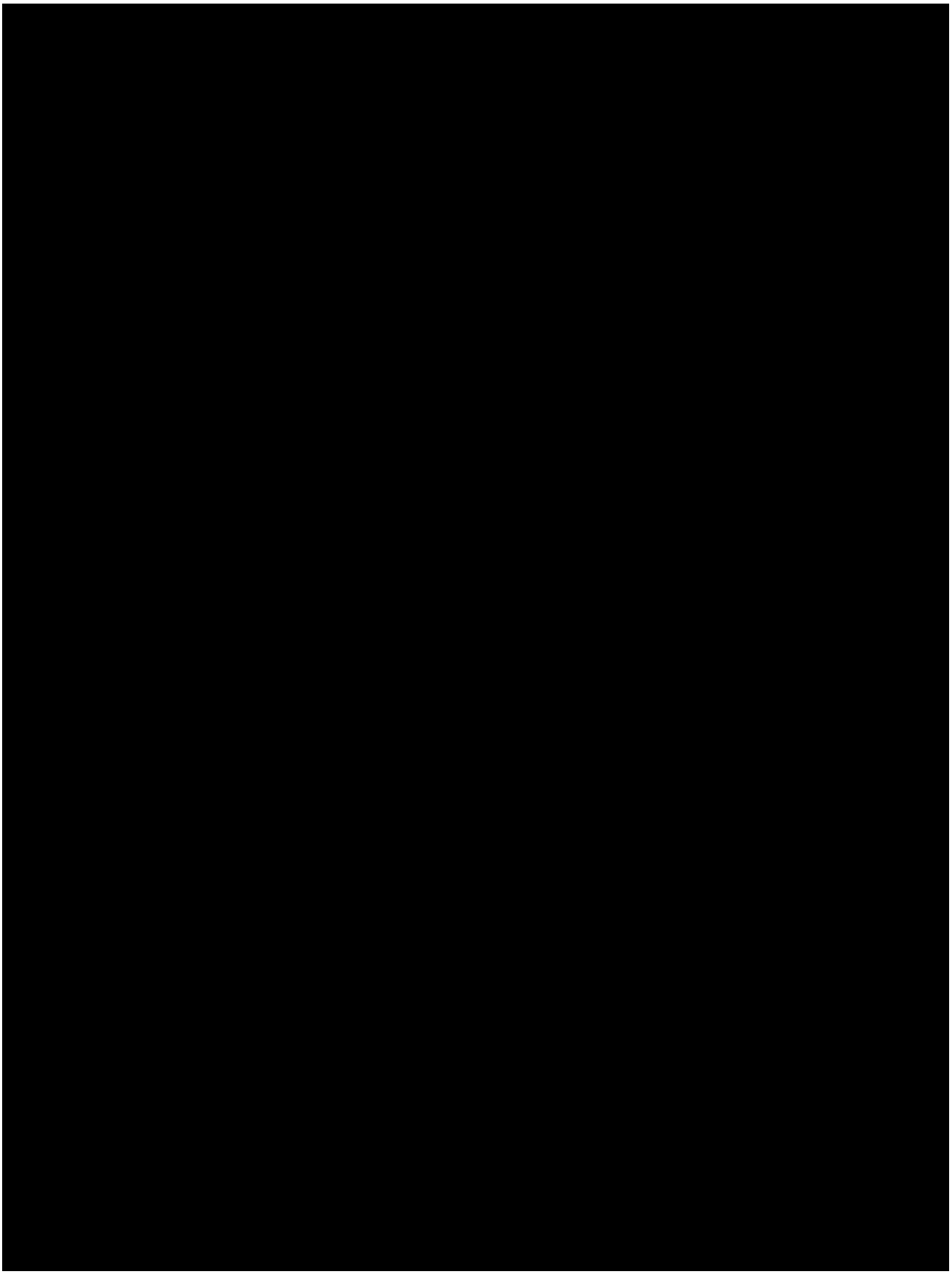
Responder rate at Visit V3, according to the MLFAS, as assessed live by a blinded evaluator, where response is defined as ≥ 1-point improvement on subject's treated lips compared to baseline. Subjects treated in both lips need to have a ≥ 1-point improvement in both lips in order to qualify as responder.

5.1.3 Secondary Effectiveness Endpoints

- Global Aesthetic Improvement Scale (GAIS) scores for treated subjects at Visit V3, as completed by the treating investigator.
- GAIS scores for treated subjects at Visit V3, as completed by the subject.
- Change from baseline to Visit V3 in Rasch-transformed score of the FACE-Q Satisfaction with Lips (absolute scores and change from baseline), as assessed by the subject.

For assessment of GAIS in-clinic, live comparisons will be made relative to the baseline photograph for Visit 3.





[redacted]

5.2 Safety Endpoints

5.2.1 Primary Safety Endpoint

Not applicable

5.2.2 Secondary Safety Endpoint

Incidence of related serious or related delayed-onset treatment-emergent adverse events (TEAEs) > 21 days after last treatment, as reported by the investigators until Week 48 post-last treatment. For subjects without touch-up, related delayed onset AEs are defined as related TEAEs starting after first treatment (V1-a/Day 1) + 21 days and before retreatment injection, if any.

For subjects with touch-up, related delayed onset AEs are defined as related TEAEs starting after first treatment (V1-a/Day 1) + 21 days and before touch-up treatment or after touchup treatment (V2-a) + 21 days and before retreatment, if any.

[redacted]

6 Statistical Analysis Methods

Adequate descriptive statistics will be provided for each endpoint, by treatment group and overall, where appropriate. Continuous endpoints (values and changes from baseline) will be summarized by metric descriptive statistics including the number of observations (n), the number of missing values, mean, standard deviation (SD), median, quartiles, minimum, and maximum. Mean, quartiles, and median will be reported to one decimal place more than the data were collected; for the SD, two decimal places more will be reported.

For qualitative endpoints, absolute and percent frequencies (n, %) will be displayed. Percentages will be calculated using the denominator of all subjects in the respective analysis set or treatment group, unless otherwise specified, and will be rounded to one decimal place. Frequency tables will include the number of missing values as applicable.

Ordered, categorical data will be summarized by metric and categorical statistics. These variables will be analyzed as absolute data and as change from baseline, as applicable.

Confidence intervals (CIs) and p-values will be reported, where appropriate. If not otherwise stated, statistical tests will be conducted two-sided at type I error rate 5%. p-values will be reported up to four decimal places, and p-values below 0.0001 will be presented as '<0.0001'. CIs will be two-sided at 95% level. If not otherwise specified, 95% exact CIs, based on a binomial distribution, will be presented for categorical data.

Unscheduled visits will not be considered for analyses except for baseline calculations.

Baseline will be taken as the latest assessments conducted (including unscheduled and repeat assessments) before the initial treatment.

Change from baseline = Result at visit/timepoint – Baseline result; if baseline result is missing, change from baseline will not be calculated.

All data captured in the electronic case report form (eCRF) and data derived for analyses will be listed.

6.1 Effectiveness Endpoints

The primary effectiveness analysis and the analyses of the co-primary and secondary effectiveness endpoints will be based primarily on the PPS and additionally, for sensitivity purposes, on the ITT. All the remaining effectiveness endpoints will be analyzed based on the PPS.

For applicable analyses, subjects will be classified into two FST groups:

1. "FST Group 1" – skin types I, II, and III
2. "FST Group 2" – skin types IV, V, and VI

6.1.1 Primary and Co-primary Effectiveness Analysis

Two hypothesis tests will be evaluated in a sequential order on the PPS at a one-sided α level of 2.5%. To control for multiplicity, a hierarchical-testing procedure will be performed.

The first hypothesis is to demonstrate that the response rate of the BIL group is statistically significantly larger than 50%. This 50% threshold is commonly used in aesthetic medical device studies to measure a meaningful response.

$H_{10}: \Pi_{BIL} \leq 50\%$ VS. $H_{11}: \Pi_{BIL} > 50\%$

Where, Π_{BIL} denotes the response rate in the BIL group (co-primary effectiveness endpoint)

With the second hypothesis, it is intended to show non-inferiority of BIL compared to an approved HA dermal filler for the subject's mean change from baseline to Visit V3 of treated lips on the MLFAS as assessed by the blinded evaluator at a non-inferiority margin of δ = -0.5 points. Non-inferiority is to be demonstrated in this study at type I error one-sided α level of 2.5%.

H_{20}: $\mu_{BIL} - \mu_{control} \leq -0.5$ VS. H_{21}: $\mu_{BIL} - \mu_{control} > -0.5$

Where, μ_{BIL} *and* $\mu_{control}$ *denote the mean MLFAS change from baseline at Visit V3 for BIL and comparator groups, respectively (primary effectiveness endpoint).*

6.1.1.1 Co-primary Analysis

The response rate of BIL will be assessed using MLFAS change from baseline, as assessed live by a blinded evaluator at Visit V3.

A subject treated in only one lip attaining at least 1-point improvement compared to baseline (i.e., change from baseline ≥ 1) on the treated lip is considered as responder. If a subject had both lips treated, improvement (i.e., change from baseline ≥ 1) must have been attained on both lips for the subject to be categorized as responder.

Absolute and percent frequencies of responders at Visit V3, along with exact two-sided 95% CI based on a binomial distribution, will be provided by treatment group. Risk difference of response rate (i.e., BIL – RS [control]) and its 95% exact CI will also be provided. Analysis will be based on PPS.

Since subjects with missing MLFAS ratings at screening or Visit V3 for lip(s) which were treated will be excluded from the PPS due to important PDs, no imputation of missing data is required.

The response rate of BIL will be deemed statistically significantly larger than 50% if the lower limit of the 95% CI of the BIL response rate exceeds the margin of 50%.

The following SAS code will be run for the co-primary analysis:

[redacted]

6.1.1.2 Primary Analysis

The non-inferiority of BIL versus RS will be tested by comparing the mean change on treated lip(s) from baseline to Visit V3, according to the MLFAS, as assessed live by a blinded evaluator.

A baseline-adjusted analysis of covariance (ANCOVA), with change from baseline to Visit V3 in MLFAS rating as a dependent variable and treatment, site, and FST group as fixed effects, will be used. MLFAS rating(s) of treated lip(s) at baseline, as assessed by a blinded evaluator, will be included as a covariate in the ANCOVA model. The model is specified as follows:

$$y_i = \sum_{j=1}^{4} x_{ij}\,\beta_j + \varepsilon_i\,, i = 1, \ldots., n$$

where y_i is the observation of the dependent variable in subject *i*, x_{11}, , …, x_{14}, x_{21}, ….x_{24}, …., x_{n1}, …, x_{n4} are regression-type variables modelling baseline covariate and dummy variables modelling treatment, site, and FST group, respectively; β_1, …, β_4 are unknown fixed-effects parameters to be estimated; and ε_1, …, ε_n are unknown independent and identically distributed normal (Gaussian) error variables with mean 0 and variance σ^2.

If a subject has both lips treated, the average change from baseline of both lips will be used for analysis. Similarly, the baseline ratings of both lips will be averaged for analysis.

Number of observations, least-square (LS) mean, standard error (SE), and 95% CI estimates from the above model will be presented by treatment group. The LS-mean difference of BIL – RS (control) and corresponding 95% CI will also be provided. The p-value resulting from the significance test for LS-mean difference will be reported.

Analysis will be based on PPS.

The following (or similar) SAS code will be run for the primary analysis:

[redacted]

No missing values will occur for the primary effectiveness endpoint in the PPS by definition; therefore, observed-case analysis on PPS will be performed.

Non-inferiority will be concluded if the lower bound of the two-sided 95% CI for the LS-mean difference between BIL and RS is above -0.5 points.

Further, metric descriptive statistics for absolute values and changes from baseline for MLFAS rating(s) of treated lip(s), as assessed live by a blinded evaluator at Visit V3, will be provided per treatment group for the PPS. For subjects with both lips treated, absolute values and change from baseline ratings of both lips will be averaged and used for analysis.

6.1.1.3 Sensitivity analyses

Robustness of results from the primary and co-primary analyses will be assessed using the following sensitivity analyses:

Co-primary Effectiveness Endpoint:

The co-primary analysis, as described in Section 6.1.1.1, will be repeated for the ITT, in which subjects with missing values will be evaluated as non-responders.

Primary Effectiveness Endpoint:

The primary analysis, as described in Section 6.1.1.2 will be repeated for the ITT using observed cases.

Furthermore, an additional sensitivity analysis was originally planned in the CSP which is not needed due to low number of missing data. If the primary endpoint was missing for more than 10% in any treatment group, the primary analysis would have been repeated for the ITT with handing of missing as follows:

Missing values due to reasons likely related to treatment outcome (adverse events related to treatment or lack of effectiveness) would have been imputed using worst case imputation. Post-baseline MLFAS ratings would have been imputed with the subject's baseline MLFAS rating of the corresponding lip (i.e., change from baseline on the MLFAS would have been 0). When compared to baseline, it is not expected that age-related reductions in lip volume occur from initial/touchup treatment to Visit V3 (8 weeks), which is why the baseline MLFAS rating is appropriate for worst-case imputation of the primary effectiveness endpoint.

Data missing other reasons (if any) would have been imputed by multiple imputations (MI) using the SAS procedure PROC MI with 1000 iterations and seed 930041001. The variables (VAR) statement considers the following variables: treatment group, site, FST group, sex, touchup at Visit V2-a, age at screening, and subject's mean MLFAS ratings at screening, Visit V2-a (Week 4), and Visit V3 (Week 8). Because of the lack of information about the missing-data pattern for MLFAS ratings over time, the fully conditional specification (FCS) option would have been chosen for the model. All missing MLFAS ratings would have been replaced for the post-baseline Visits V2-a and V3. The data would have been sorted by subject identifier. For all variables the format would have been deleted.

The following (or similar) SAS code would have been run for this sensitivity analysis:

[redacted]

The multiple ANCOVA results generated from each imputed dataset would have been combined with SAS PROC MIANALYZE using the following (or similar) SAS code

[redacted]

6.1.1.4 Subgroup analyses

Subgroup analyses will be based on observed cases in the PPS.

For analysis, subjects will be classified into the following "Age group" categories based on age at the time of informed consent: 21 - 30 years, 31 – 40 years, 41 – 50, and 51 – 65 years.

Co-primary Effectiveness Endpoint:

Absolute and percent frequencies of responders at Visit V3 will be summarized by treatment group for subgroups based on sex, race, ethnicity, FST group, and age group.

Primary Effectiveness Endpoint:

Metric statistics for absolute values and changes from baseline in MLFAS ratings, as assessed by a blinded evaluator at Visit V3, will be presented by treatment group for subgroups based on sex, race, ethnicity, FST, age group and site.

6.1.2 Secondary Effectiveness Endpoints

Secondary effectiveness endpoints will be analyzed primarily on the PPS using observed cases and additionally, for sensitivity purposes, on the ITT using observed cases.

Global Aesthetic Improvement Scale (GAIS)

GAIS scores at Visit V3, as completed by the investigator and the subject, will be summarized separately, for each treatment group, using metric statistics (number of observations, mean, standard deviation, minimum, quartiles, median, and maximum values). GAIS scores at Visit V3 completed by the investigator and the subject will also be summarized separately using absolute and percent frequencies for each possible score result and for the categories in Table 1.

Table 1: GAIS Score to Category Conversion

GAIS Score	Category
+3, +2, +1	any improvement
0, -1, -2, -3	no improvement

FACE-Q Satisfaction with Lips

Sum scores for all questions in the FACE-Q Satisfaction with Lips module will be transformed to a score ranging from 0 to 100, using Rasch-transformation (see [redacted]

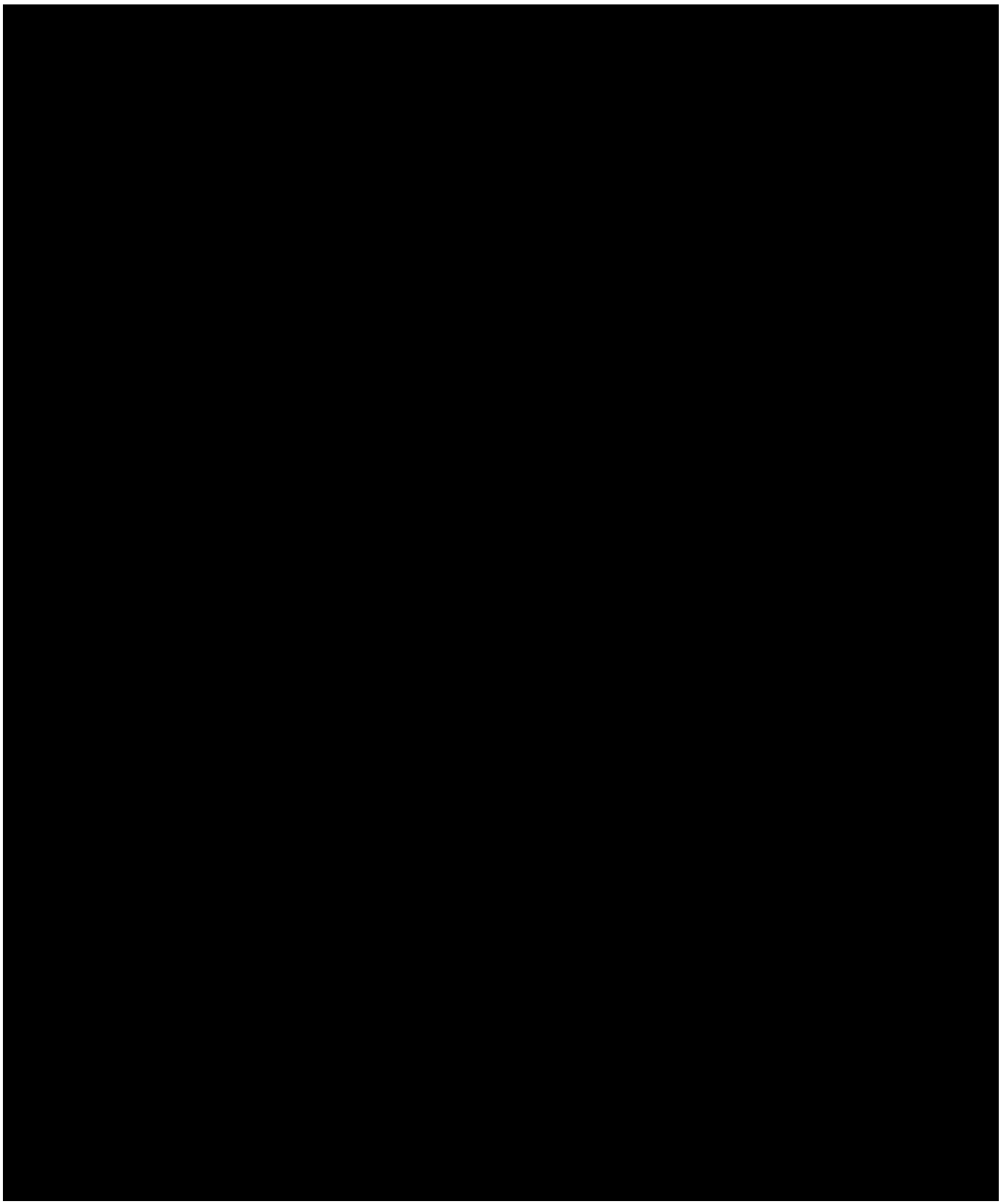
Missing-question scores will be imputed before Rasch transformation as follows:

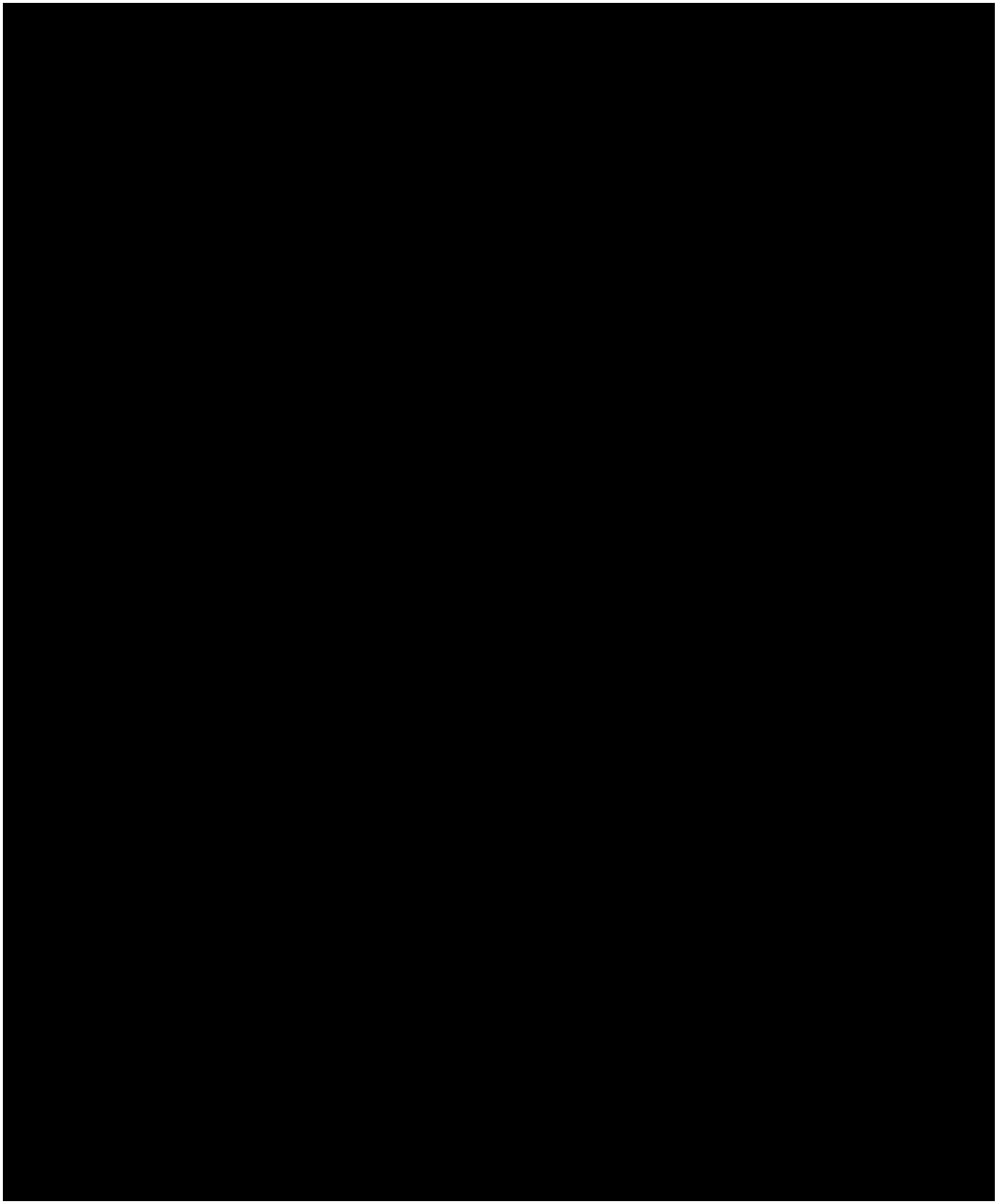
1. If less than 50% of the items in the module have missing values, the average score of the completed items will be used to replace the missing values and calculate the sum score.

2. If 50% or more items in the module are missing, no imputation will be performed and the sum score will not be derived (i.e., analysis value will be missing).

The Rasch-transformed scores of FACE-Q Satisfaction with Lips at Visit V3 will be summarized by treatment group using metric statistics for absolute values and change from baseline values.

A listing for FACE-Q scores and their derived Rasch-transformed scores will be presented.





6.2 Safety Endpoints

All safety analyses will be performed on the SES in total and by treatment groups. AEs will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA) version in effect at the time the database is closed.

TEAEs and non-TEAEs (if applicable) will be tabulated per

TEAEs were defined as AEs with onset or worsening on or after date of first administration of treatment (BIL or comparator). In this regard, an AE with onset prior to treatment that worsens at or after first administration of study treatment must have been documented as a new TEAE with onset at the time of worsening.

TEAEs with delayed onset are defined as AEs with onset after more than 21 days of last treatment as described in Section 5.2.2.

If the AE-onset day is incomplete, the event will be regarded as delayed onset, if the non-missing part of start does not imply onset before 22 days.

TEAE analyses will be reported overall and by treatment period. Two treatment periods will be defined: initial and retreatment period (BIL group only). TEAEs with onset on or after initial treatment and before retreatment will be assigned to the initial treatment period. TEAEs with onset on or after retreatment will be assigned to the retreatment period. TEAEs of subjects without retreatment will be assigned to the initial treatment period.

An AE will be regarded as treatment related if the causal relationship to product and/or to injection procedure is documented as “related” in the eCRF.

If intensity, causal relationship, and/or outcome is missing, the worst-case principle will be applied (i.e., missing intensity set to “severe”, missing causal relationship set to “related”, and missing outcome set to “unknown”).

6.2.1 Primary Safety Endpoints

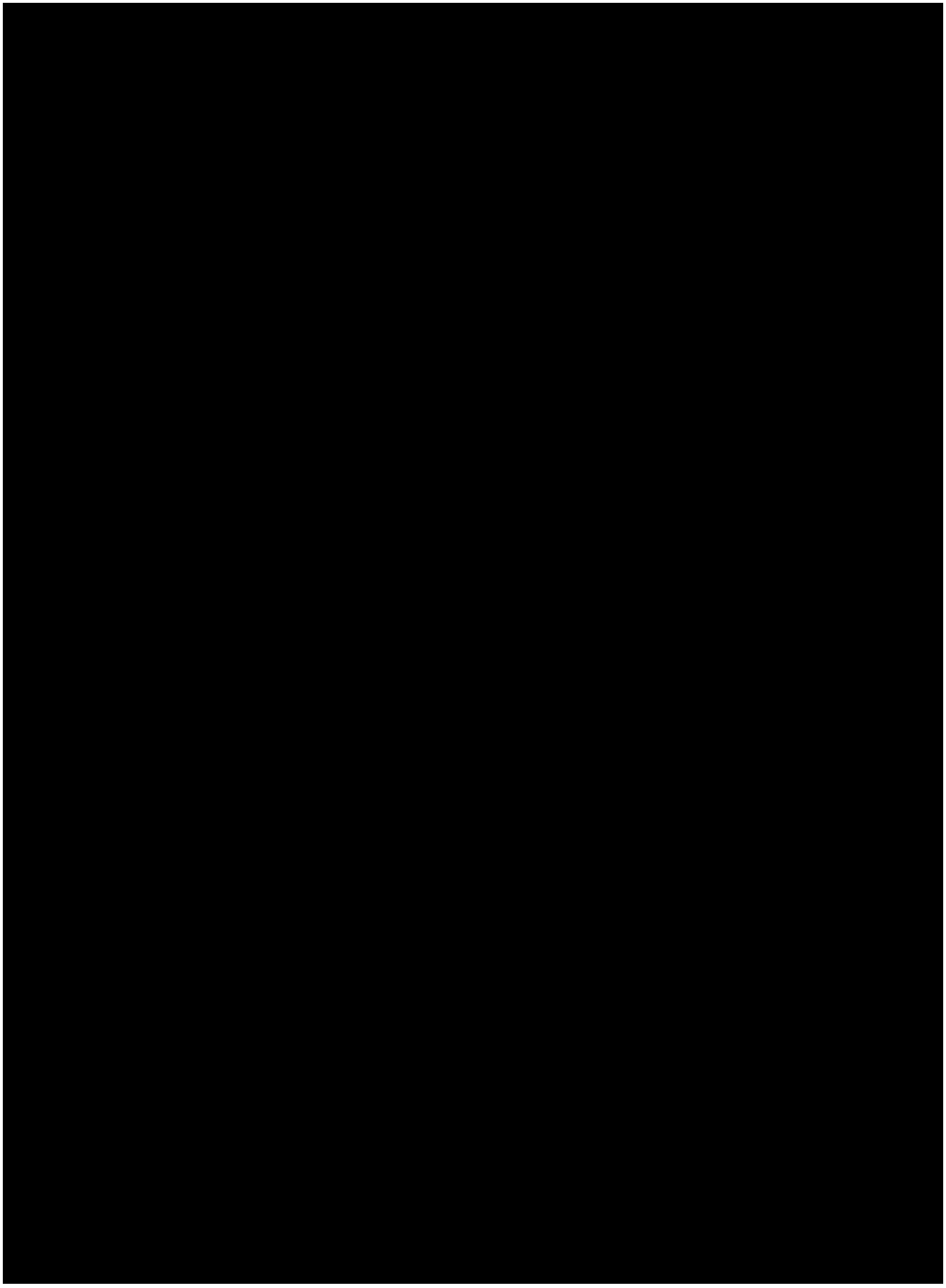
Not Applicable.

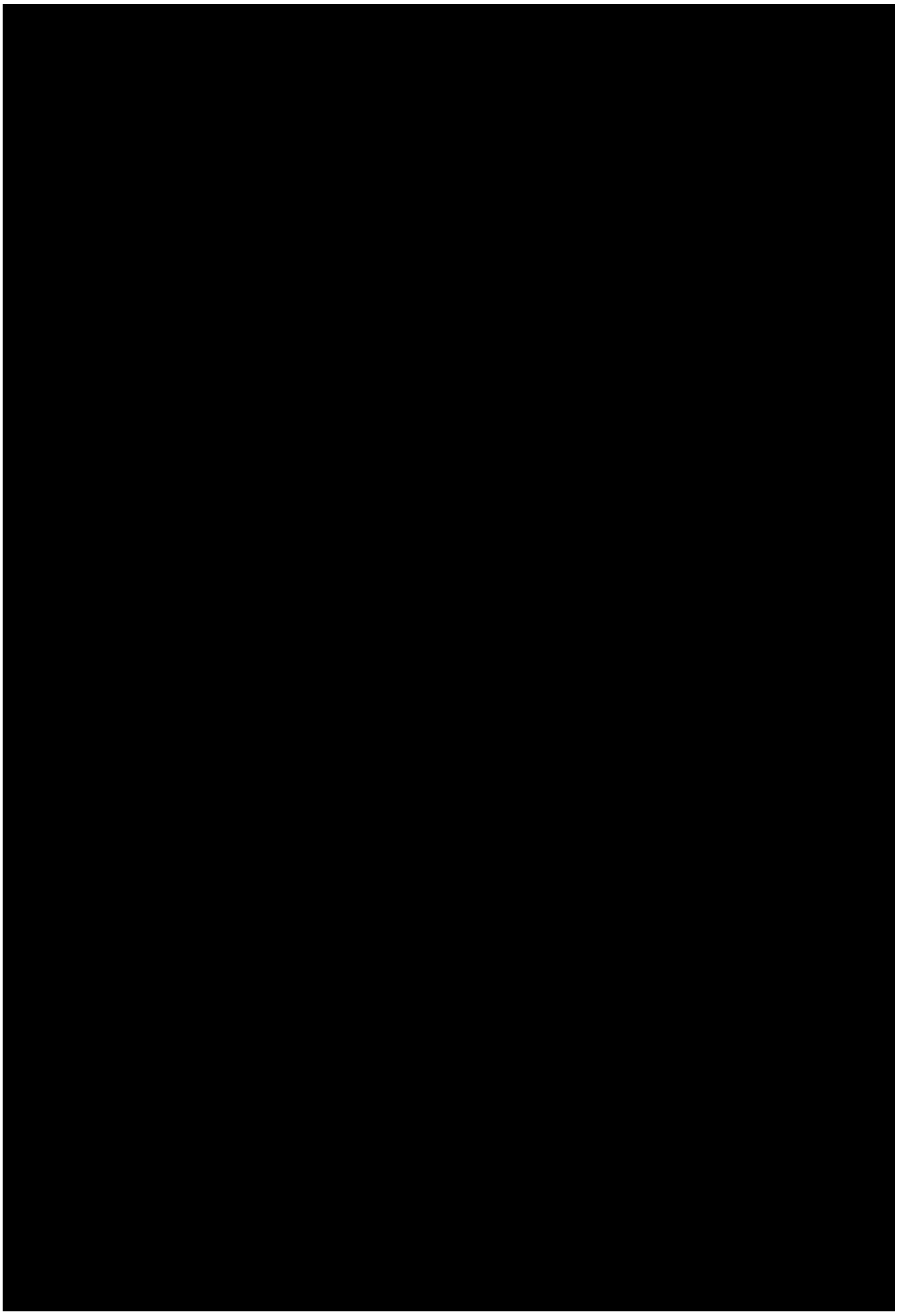
6.2.2 Secondary Safety Endpoints

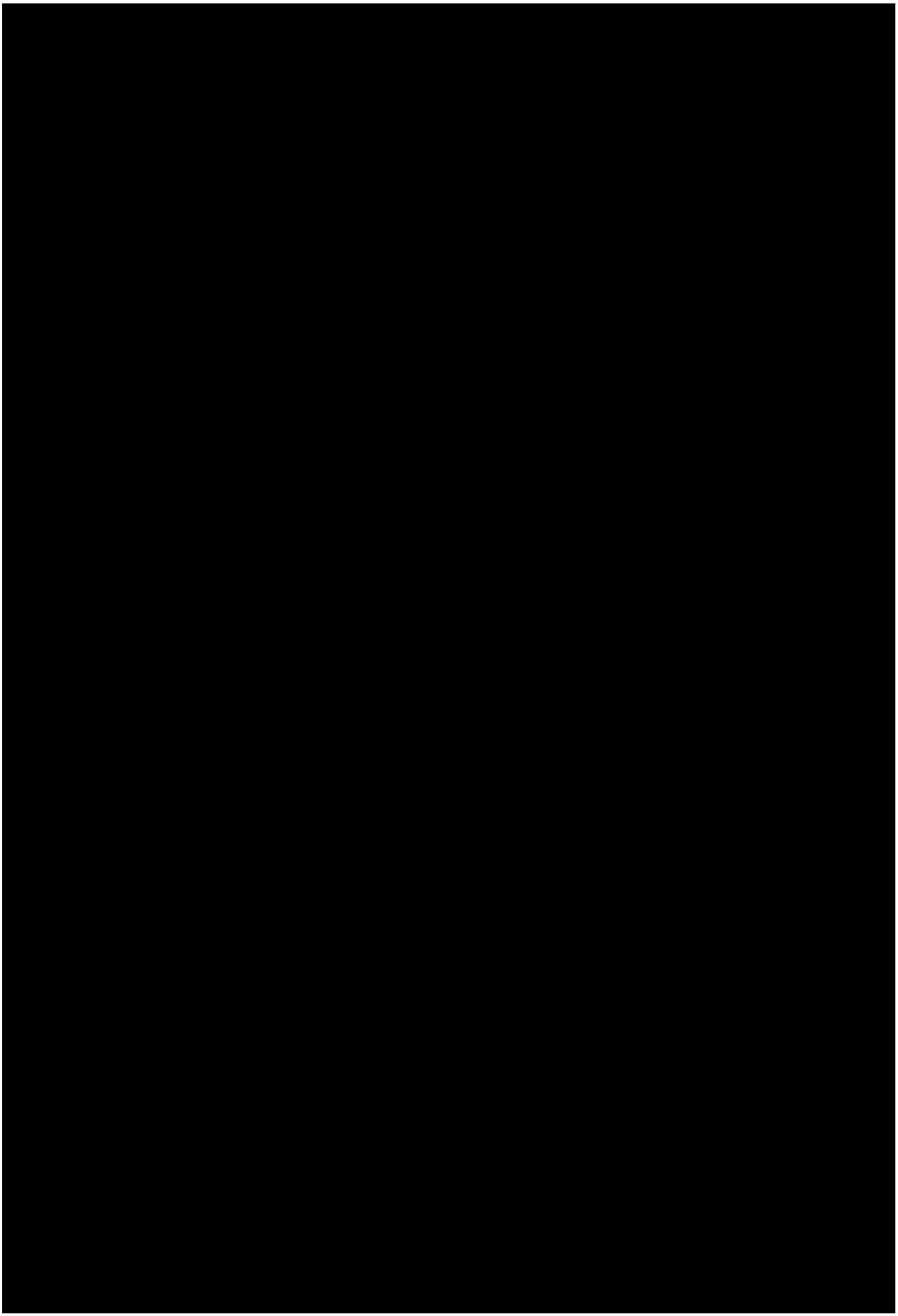
Frequency tables with absolute and percent frequencies of subjects and absolute frequency of events reported by investigators until Week 48 post-last treatment will be summarized for related serious TEAEs and related delayed-onset TEAEs. Summaries will be presented by System Organ Class (SOC) and Preferred Term (PT), and listings will be provided.

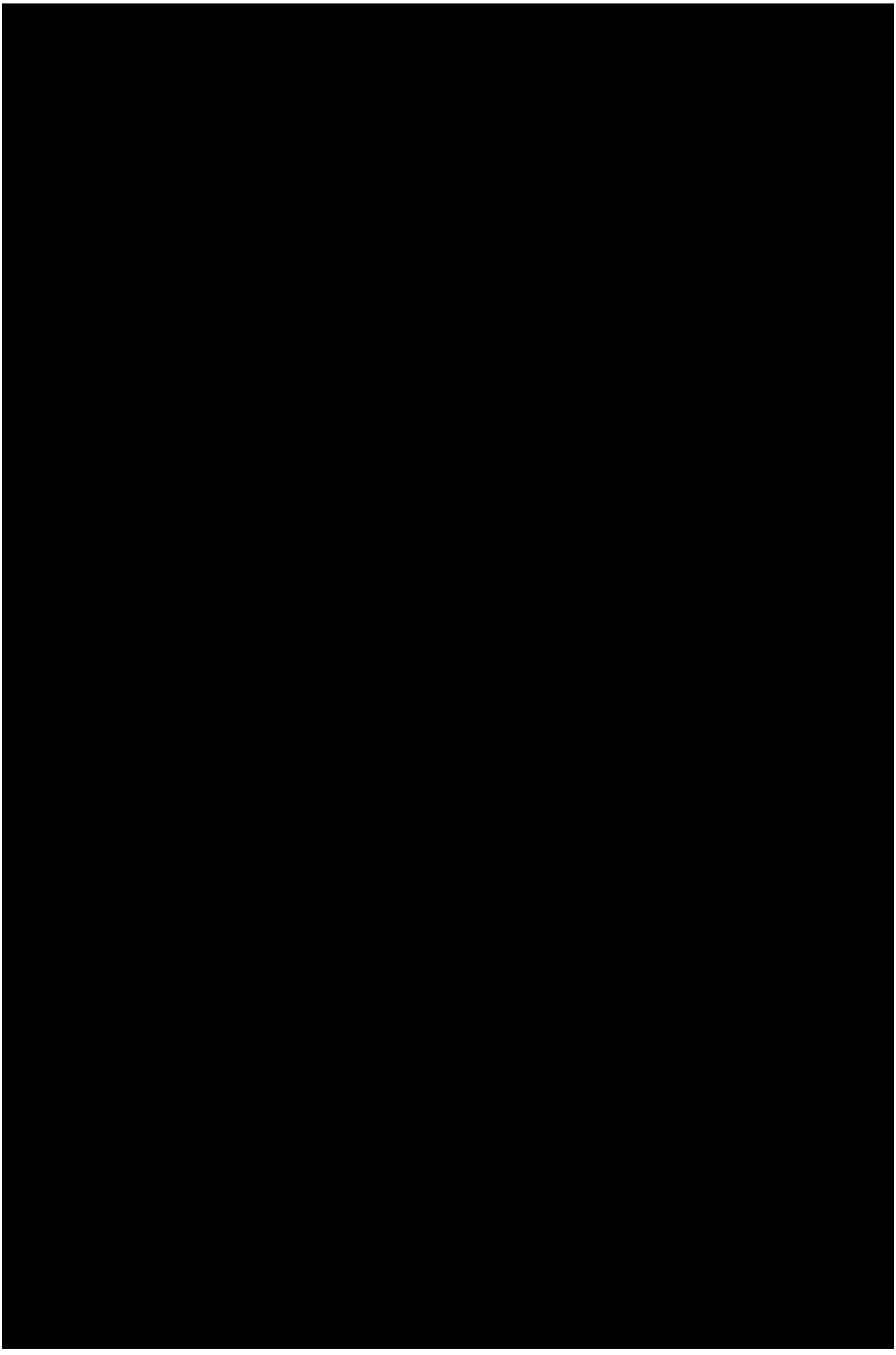
Additionally, a separate subject listing will be provided for related delayed-onset TEAEs.

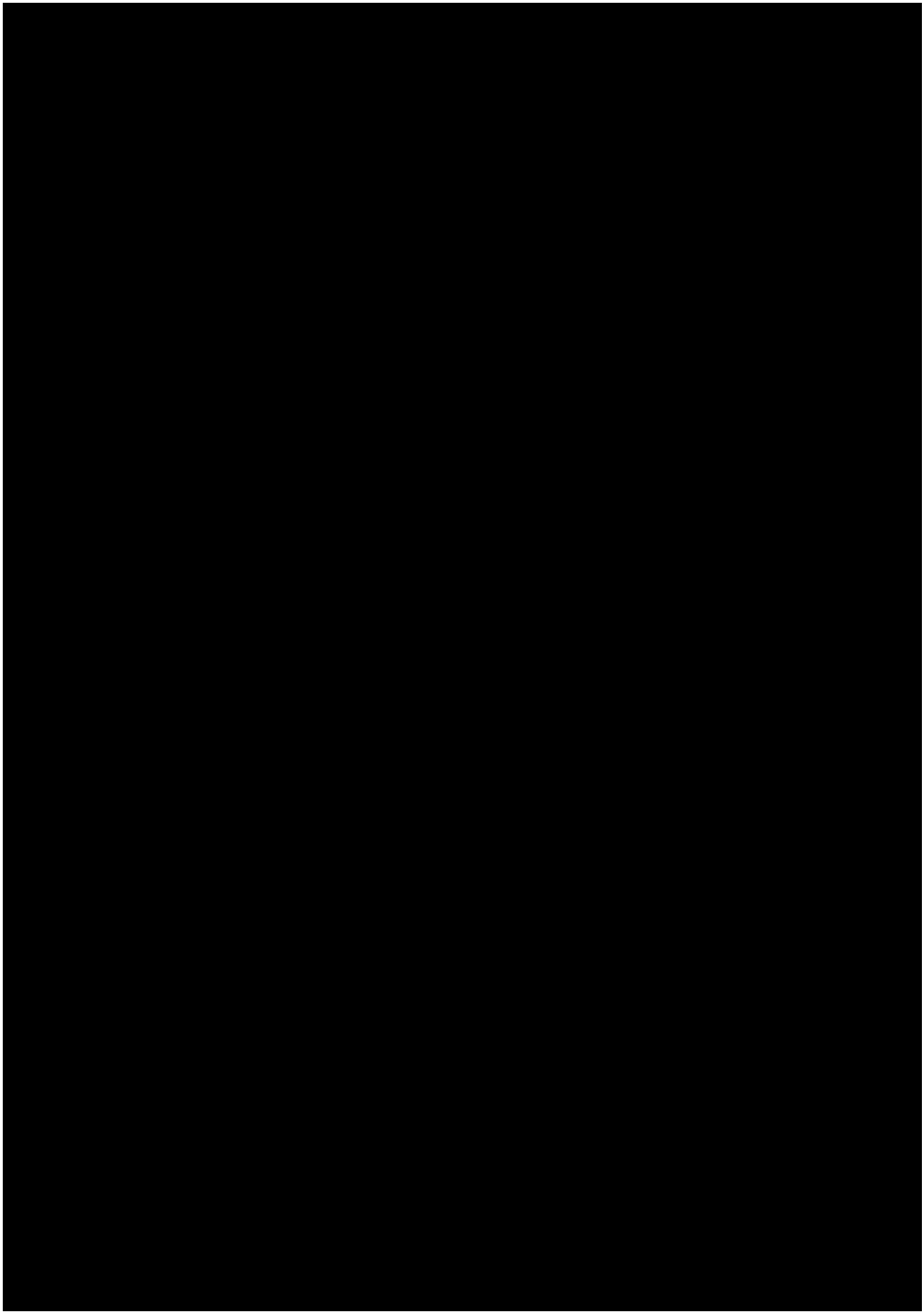












6.4 Special Statistical/Analytical Issues

6.4.1 Discontinuations and Missing Data

Discontinued subjects will not be replaced. In case of more than 10% missing primary effectiveness data due to a public health emergency, additional subjects could have been randomized.

In general, missing data will not be imputed and analyses will be based on observed cases.

For missing imputation techniques of the primary and secondary effectiveness endpoints, refer to Section 6.1.

For missing imputation techniques for AEs, refer to Section 6.2.3 and for CTRs, refer to [redacted]

6.4.2 Interim Analyses

No interim analysis is planned.

6.4.3 Multiple Comparisons/Multiplicity

Multiplicity in effectiveness analysis will be handled using a hierarchical-testing procedure, as detailed in Section 6.1.

6.4.4 Examination of Subgroups

Subgroup analyses for the primary and secondary effectiveness endpoints are specified in Sections 6.1.1.4 and 6.1.2.

Subgroup analyses for related TEAEs and CTRs are specified in Section 6.2.3.

6.4.5 Pooling of Sites

Pooling of investigational sites will be performed according to the algorithm described below. If convergence problems are observed in the ANCOVA/ANOVA model described above, the following steps will be performed:

1. The site with the lowest number of randomized subjects will be pooled with the site with second lowest number of subjects. If two or more sites with the same number of subjects are suited for pooling, the site with the lowest number of subjects in the control group will be pooled. If two or more sites with the same number of subjects and same number of subjects in the control group are suited for pooling, then the site to be pooled will be chosen randomly among all suited sites.

2. ANCOVA/ANOVA model will be re-run with pooled site as factor.

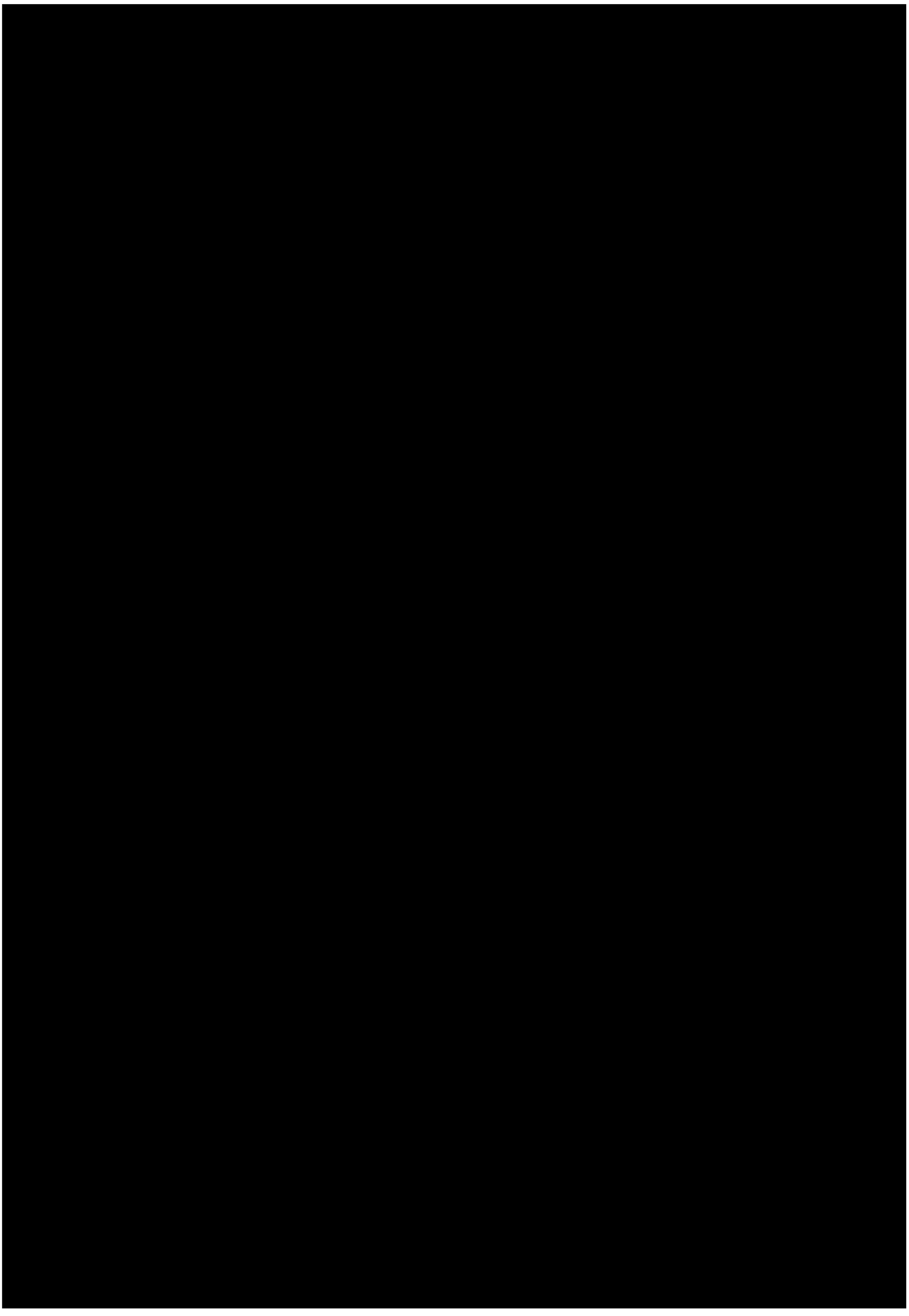
3. If convergence problems are resolved, no further pooling of investigational sites will be performed. Otherwise, steps 1) and 2) above will be repeated until convergence problems resolve.



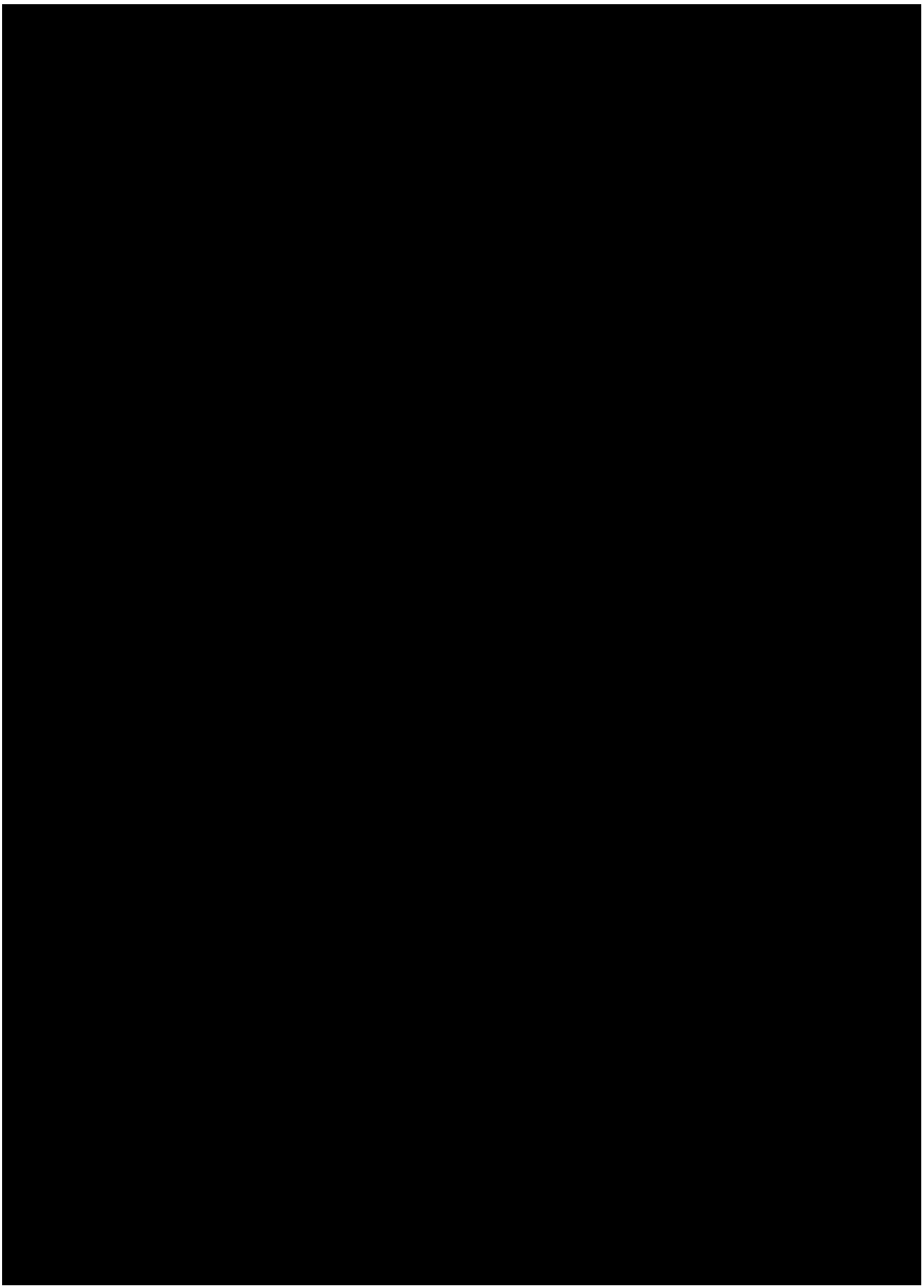
8 References

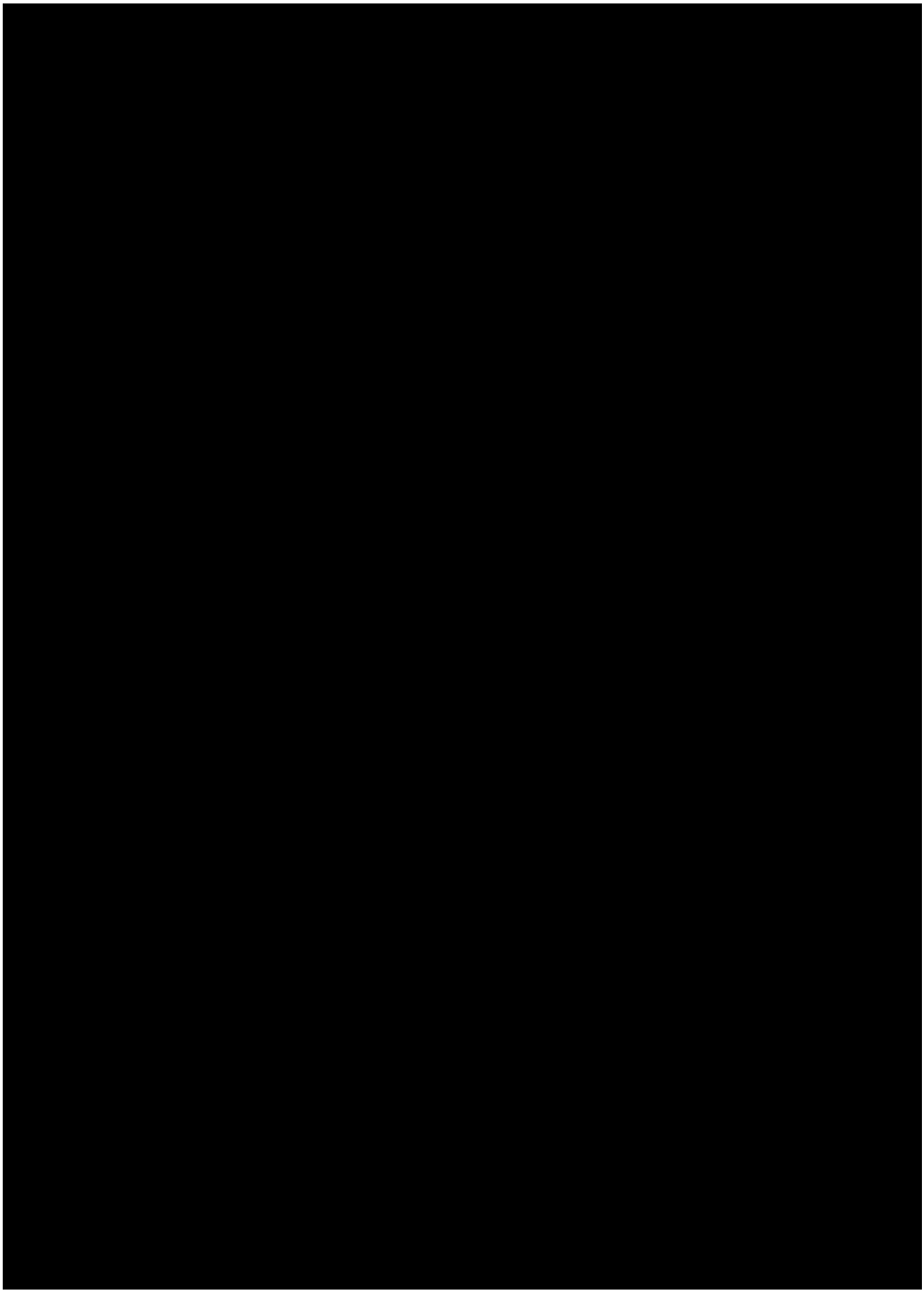
[1] M900721001 Clinical study report: Open-label, multicenter, uncontrolled, rater-blinded, post-market clinical follow-up (PMCF) study to conform performance and safety of Belotero Soft Lidocaine and Belotero Intense Lidocaine in the treatment of nasolabial folds, perioral fine lies and/or for lip volume enhancement, 21 February 2020

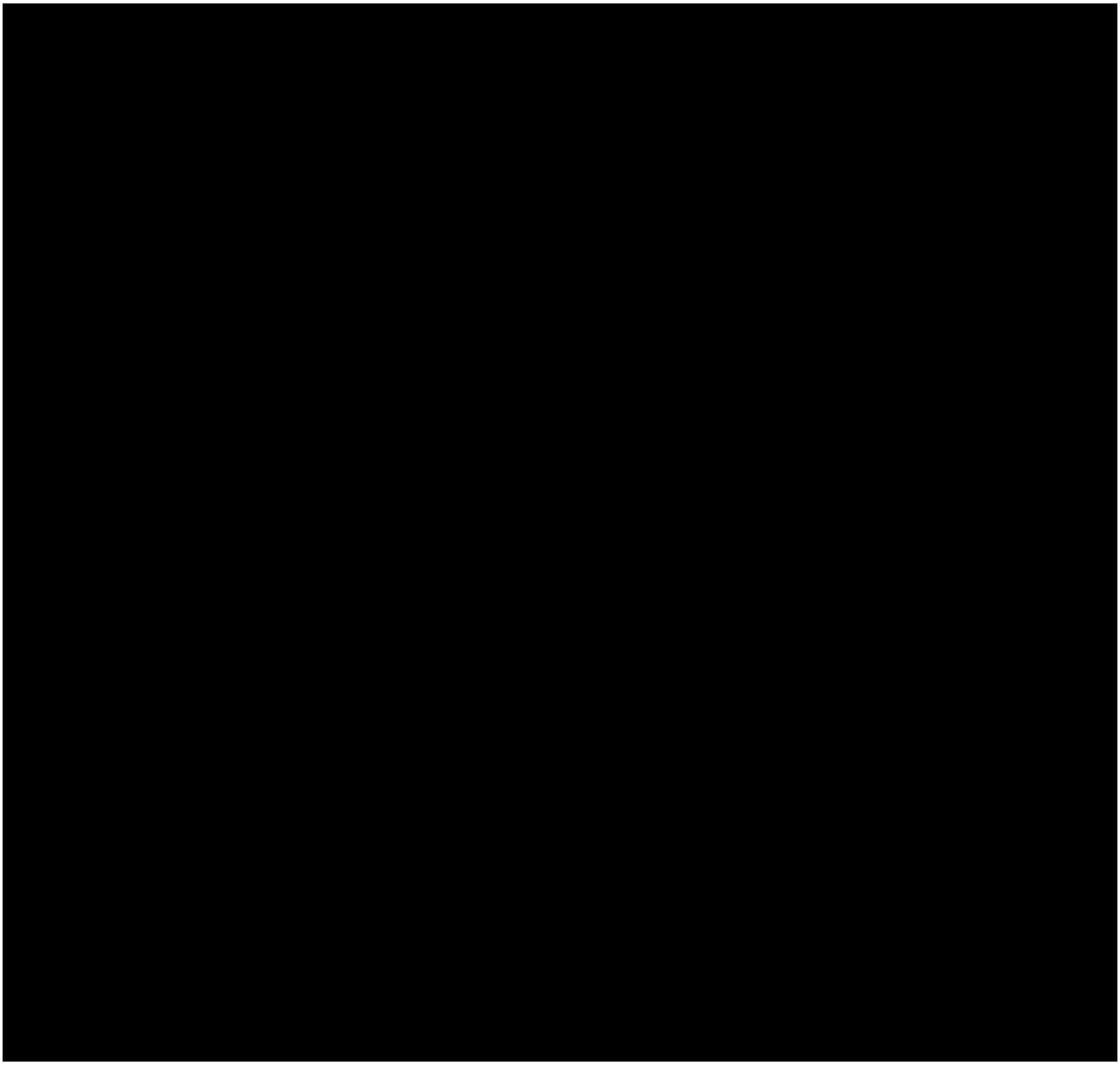
[2] MA-1700-04 Clinical Study: A randomized, evaluator blinded, no treatment-controlled study to evaluate the safety and effectiveness of Restylane Silk Injectable Gel with 0.3% Lidocaine for lip fullness augmentation and treatment of perioral rhytids. *In* SUMMARY OF SAFETY AND EFFECTIVENESS DATA (SSED) Restylane Silk Injectable Gel USA June 2014. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf4/P040024S072b.pdf











9.8 Appendix 8: Rasch Transformed FACE-Q Scores

