NCT04177212, redacted
version v1.0, 05Sep2023

# Statistical Analysis Plan

Prospective, multicenter, parallel-group, evaluator-blind, randomized study to investigate the effectiveness and safety of MRZF111 in the treatment of décolleté wrinkles

M930521001

***Version 1.0, Final***

Date: 22-APR-2021

Author: [redacted]

## SIGNATURE PAGE

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

[redacted]

| Author (print name) | Date (dd-mmm-yyyy) | Signature |
|---|---|---|

[redacted]

| Peer Reviewer (print name) | Date (dd-mmm-yyyy) | Signature |
|---|---|---|

[redacted]

| Study Medical Expert Merz (print name) | Date (dd-mmm-yyyy) | Signature |
|---|---|---|

# TABLE OF CONTENTS

# 1 LIST OF ABBREVIATIONS

| | |
|---|---|
| AE | Adverse event |
| ADE | Adverse device effect |
| ATC | Anatomical Therapeutic Chemical classification system of the World Health Organization |
| BMI | Body Mass Index |
| CRO | Contract research organization |
| [redacted] | |
| DRM | Data review meeting |
| eCRF | Electronic case report form |
| FAS | Full analysis set |
| G | Gauge |
| GAIS | Global Aesthetic Improvement Scale |
| iGAIS | Investigator's Global Aesthetic Improvement Scale |
| IMD | Investigational medical device |
| [redacted] | |
| LOCF | Last Observation Carried Forward |
| MAS | Merz Aesthetic Scale |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MI | Multiple Imputation |
| Nobs | Number of observations |
| pdf | Portable data format |
| PPS | Per protocol set |
| PT | Preferred term |
| SAE | Serious adverse event |
| SAP | Statistical analysis plan |
| SES | Safety evaluation set |
| sGAIS | Subject's Global Aesthetic Improvement Scale |
| SOC | System organ class |
| TEAE | Treatment emergent adverse event |
| TFLs | Tables, Figures, and Listings |
| WHO-DD | World Health Organization Drug Dictionary |

# 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 M930521001 (version 3.0), dated 29-Aug-2020 and the following amendment (version 4.0), dated 10-Jun-2020.

All programs will be written using SAS version 9.4. A preferred font size of 10 points, minimum font size of 8 points with a unique font size for the whole document required will be used for the tables and figures in section 14 for an output in A4 landscape format. For listings, a minimum 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 using the Metronomia internal procedures. These PDF files will be generated as needed to populate the subsections of Section 14 and Section 16.2 for the clinical study report. Each PDF file will include the corresponding table of contents, preceding the content of the file. Special attention will be paid to planning and performance of quality control measures as documented in the QC plan for the analysis of this study.

The Merz standard Tables, Figures, and Listings (TFLs) for medical devices, version 2.0 (18-Feb-2020) 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 investigation is a 52 weeks prospective, multicenter, parallel-group, randomized clinical investigation with [redacted] to the treatment assignment [redacted]. A [redacted] was chosen to minimize bias due to demand characteristics (i.e. what the evaluator expects). A parallel-group design was chosen to evaluate two different treatment regimens in one study. 9 sites were chosen to increase representativeness of the investigational results and to reduce site-associated biases. The treating investigators [redacted] that (will) participate in the investigation are board-certified dermatologists and/or plastic surgeons trained and qualified on the MAS. The [redacted] that (will) participate in the investigation are physicians or scientific experts, experienced in the area of aesthetic medicine, and also trained and qualified on the MAS.

Subjects have been randomized 1:1 either to treatment group A or treatment group B:

[redacted]

Number of visits:

Subjects have been assigned to attend overall 6 visits [redacted]. The treatment visits were followed by 24 or at latest 48 hours post-injection telephone contacts. On injection visits, all investigation assessments were conducted before injection.

When COVID-19 public health emergency measures were implemented in Germany in spring 2020, it was foreseeable that not all subjects could come to all their post-baseline visits in due time. Therefore, time windows of early post-baseline visits have been widened and time points of later post-baseline visits have been adjusted. [redacted]

## 3.2 Clinical Study Objectives

Primary objectives

- To evaluate the effectiveness and safety of MRZF111 treatment for improvement of décolleté wrinkles as assessed on the MAS Décolleté Wrinkles – At Rest.

Secondary objectives

- To evaluate the effectiveness of MRZF111 treatment for improvement of décolleté wrinkles as assessed on the MAS Décolleté Wrinkles – Dynamic.

- To evaluate the subject's and physician's treatment satisfaction assessment after MRZF111 treatment of the décolleté.

- To evaluate the subjects and physician's global aesthetic improvement (GAIS-Wrinkles) after MRZF111 treatment of the décolleté.

[redacted]

# 4 DETERMINATION OF SAMPLE SIZE

The primary effectiveness endpoint is the proportion of subjects with at least 1 point (≥ 1 point) improvement on the MAS Décolleté Wrinkles – At Rest scale at 16 weeks after last treatment compared to baseline (Day 1) [redacted]. By use of a one-sided binomial test, it will be tested at an error level of 2.5% whether the response rate is significantly larger than 50%. It is aimed to ensure a power of 90%. The binomial test is applied to the pooled data from both treatment groups. [redacted]

Per definition, the safety evaluation set (SES) will be at least as large as the FAS. Thus, according to the above outlined assumptions, at least 100 subjects are planned to be available for safety analyses. This sample size is sufficient to observe, with a probability of 80%, at least one adverse event with actual event probability of 1.6%.

With regard to safety analyses, it should be considered that the RADIESSE® in its original undiluted form is marketed since 2003. Since then, over 5 million units had been sold and used in a variety of subject populations. In this period, the majority of reported AEs were common expected injection site reactions that were mild in nature and resolved over time. [redacted] he large extent of the exposure to the undiluted RADIESSE® worldwide versus the relatively small number of complications led to the conclusion that its safety has already been established to a sufficient degree. As it is not expected that dilution of RADIESSE® Volume Advantage 1.5 mL Injectable Implant with saline will cause more AEs, it is assumed the safety can be transferred to the IMD.

Consequently, the sample size for this study was primarily determined to ensure an adequate power for the primary effectiveness analysis, however, still providing a good probability of risk detection.

[redacted]

# 5 ANALYSIS SETS

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

## Safety Evaluation Set (SES)

The SES is the subset of all randomized subjects who were exposed to the IMD at least once. Subjects will be analyzed as treated. [redacted]

## Full Analysis Set (FAS)

The FAS is the subset of subjects in the SES for whom at least the baseline value of the primary effectiveness variable is available (i.e., all subjects who have a baseline value on the MAS Décolleté Wrinkles – At Rest scale). In contrast to the SES, subjects will be analyzed as randomized.

## Modified Full Analysis Set

The modified FAS is the subset of the subjects in the SES for whom at least the baseline value of the primary effectiveness variable is available (i.e., all subjects who have a baseline value on the MAS Décolleté Wrinkles – At Rest scale). Subjects will be analyzed as treated. [redacted] Subjects who were treated only once with all missing treatments due to reasons not related to COVID-19 measures, will be analyzed in the Group they are randomized to. This includes premature withdrawal or PDs not related to COVID-19 measures. [redacted]

## Per Protocol Set (PPS)

The PPS is the subset of subjects in the modified FAS without major protocol deviations. Subjects of treatment group A switching to treatment group B will not be excluded from the PPS per se, as this is in line with protocol amendment 4. Further major protocol deviations and other reasons leading to exclusion will be defined prior to database lock during a data review meeting (DRM).

# 6 ENDPOINTS FOR ANALYSIS

## 6.1 Effectiveness Endpoints

### *6.1.1 Primary Effectiveness Endpoint*

The primary effectiveness endpoint is the proportion of subjects with at least 1-point (≥ 1-point) improvement on the MAS 'Décolleté Wrinkles – At Rest' scale at 16 weeks after last treatment compared to baseline (Day 1). [redacted] A response is defined as ≥ 1-point improvement on MAS 'Décolleté Wrinkles – At Rest' scale at 16 weeks after treatment compared to baseline (Day 1).

### *6.1.2 Secondary Effectiveness Endpoints*

The secondary effectiveness endpoints are:

- Proportion of subjects with at least 1-point (≥ 1-point) improvement on the MAS 'Décolleté Wrinkles – Dynamic' scale at 16 weeks after last treatment compared to baseline (Day 1). [redacted] A response is defined as ≥ 1-point improvement on MAS 'Décolleté Wrinkles – Dynamic' scale at 16 weeks after treatment compared to baseline (Day 1).

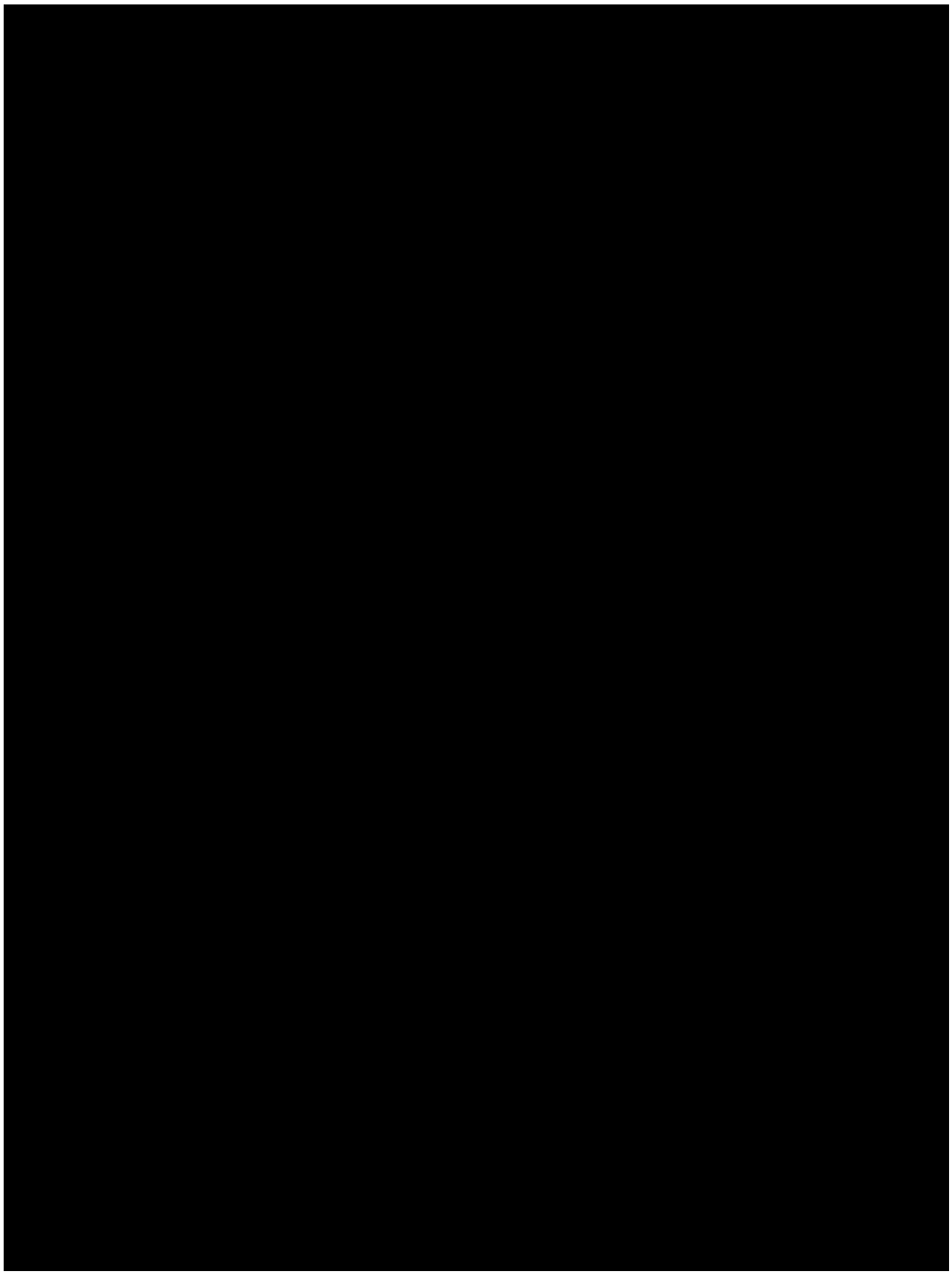
- Treating investigator's treatment satisfaction assessment of aesthetic improvement in the subject after the décolleté treatments 16 weeks after last treatment.

- Subject's treatment satisfaction assessment of aesthetic improvement after the décolleté treatments 16 weeks after last treatment.

- Treating investigator's evaluation of the global aesthetic improvement on the Investigator's Global Aesthetic Improvement Scale (iGAIS) [redacted].

- Subject's evaluation of the global aesthetic improvement on the Subject's Global Aesthetic Improvement Scale (sGAIS [redacted].

[redacted]



[redacted]

## 6.2 Safety Endpoints

### *6.2.1 Primary Safety Endpoints*

Not applicable.

### *6.2.2 Secondary Safety Endpoints*

- Incidence of TEAEs related to the treatment with the IMD.

[redacted]

# 7 STATISTICAL ANALYSIS METHODS

In the following, the term "treatment group" refers to the realized treatment group. This may differ from the randomized treatment group as those subjects randomized to group A [redacted] who did not manage to get their second injection in due time due to the COVID-19 public health emergency measures were allowed to skip this treatment and proceed with the last foreseen treatment. By this measure, these subjects switched to treatment group B [redacted]. Assuming that the mechanism of being prevented from performing the second treatment visit due to COVID-19 public health emergency measures is independent from any known or unknown, potentially confounding factors, the realized treatment groups should still be comparable in their baseline characteristics and analyses on them of high scientific value.

In the sections below, baseline severity refers to the severity of wrinkles on the MAS 'Décolleté Wrinkles – At Rest' as assessed [redacted].

For all analyses concerning '16 weeks after last treatment', assessments of the visit closest to 16 weeks after last treatment within the time frame of 12 to 20 weeks after last treatment will be used. In case no visit was performed in the specific time frame, the assessments for '16 weeks after last treatment' will be treated as missing.

## 7.1 Effectiveness Endpoints

The primary and secondary effectiveness endpoints will be summarized using the modified FAS population with missing data replaced as described in Section 7.4.1; for sensitivity purposes, these analyses will also be performed on the modified FAS with observed data and on the PPS. [redacted]

Statistical tests will be one-sided binomial tests at error level 2.5% for the responder rates in general. Adequate [redacted] statistics will be provided for each variable for the pooled data and by realized treatment group. The supportive exploratory analyses for the primary effectiveness endpoint will also be performed on the observed cases in the FAS. Metric variables will be summarized by number of observations (Nobs), number of missing values (nmiss), mean, standard deviation, median, minimum, Q1, Q3 and maximum. Categorical variables will be summarized by number and rate of events per category where the denominator will be chosen according to the adequate analysis population. Ordered categorical data may be summarized by metric and categorical statistics. All variables will be analyzed as absolute data and as change from baseline, as applicable.

Summary statistics for all visits and changes from baseline for all post-baseline visits will be displayed for each effectiveness endpoint for each treatment group and the pooled data.

If not otherwise specified, confidence intervals in the following refer to the two-sided 95% Wilson score confidence intervals.

All data captured in the electronic Case Report Form (eCRF) will be listed.

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, and standard deviation will be reported to one decimal place more than the data were collected; 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 group. 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(s)*

The primary effectiveness endpoint is the proportion of responders (responder rate) defined as subjects with ≥ 1-point improvement on the MAS 'Décolleté Wrinkles – At Rest' from baseline to 16 weeks after last treatment as assessed [redacted].

The primary analysis will be based on a one-sided binomial test at the one-sided type I error level of $\alpha = 0.025$ for the pooled data from both treatments against the null-hypothesis that the response rate (R) is ≤50%:

$H_0$ : R ≤50%

$H_1$ : R >50%.

$H_0$ is rejected if the resulting one-sided p-value is <0.025 and the observed proportion of responders is >0.5.

As further supportive explorative analyses, the effectiveness for the different time points of treatments with a 16 weeks post-treatment period in both treatment groups will be analyzed as for the primary endpoint:

- Proportion of responders (response rate) defined as subjects with ≥ 1-point improvement on the 'MAS Décolleté Wrinkles – At Rest' from last treatment application to 16 weeks after last treatment as assessed [redacted].
- Proportion of responders (response rate) defined as subjects with ≥ 1-point improvement on the MAS 'Décolleté Wrinkles – At Rest' from last treatment application to 16 weeks after last treatment as assessed [redacted].
- Proportion of responders (response rate) defined as subjects with ≥ 1-point improvement on the MAS 'Décolleté Wrinkles – At Rest' from baseline to 16 weeks after first treatment as assessed [redacted].

In addition, two-sided 95% Wilson score confidence intervals for response rates will be provided overall, and for each treatment group in the supportive analyses. A similar SAS code as given above will be used.

The primary effectiveness analyses will be based on the modified FAS with missing data replaced by a multiple imputation approach for each realized treatment group. [redacted]

Sensitivity analyses of the primary endpoint will be provided based on observed cases only in the modified FAS and on the observed cases in the PPS. The supportive explorative analyses will be performed on the observed cases in the modified FAS as well as on observed cases in the FAS. For analysis based on observed data, SAS code similar to the following will be used:

```
proc freq data = dataset order = freq;
      tables response / binomial (level="1" wilson p=.5) cl alpha=.05;
run;
```

, assuming the response is coded with '1' and no response is coded with '0'.

### *7.1.2 Secondary Effectiveness Endpoints*

**MAS ratings**

One secondary effectiveness endpoint is the proportion of responders (response rate) defined as subjects with ≥ 1-point improvement on the MAS 'Décolleté Wrinkles – Dynamic' from baseline to 16 weeks after last treatment as assessed [redacted]. Same analyses as for the primary endpoint will be performed on this endpoint (including sensitivity and supportive exploratory analyses on the same analysis sets and treatment groups).

**Satisfaction with décolleté appearance**

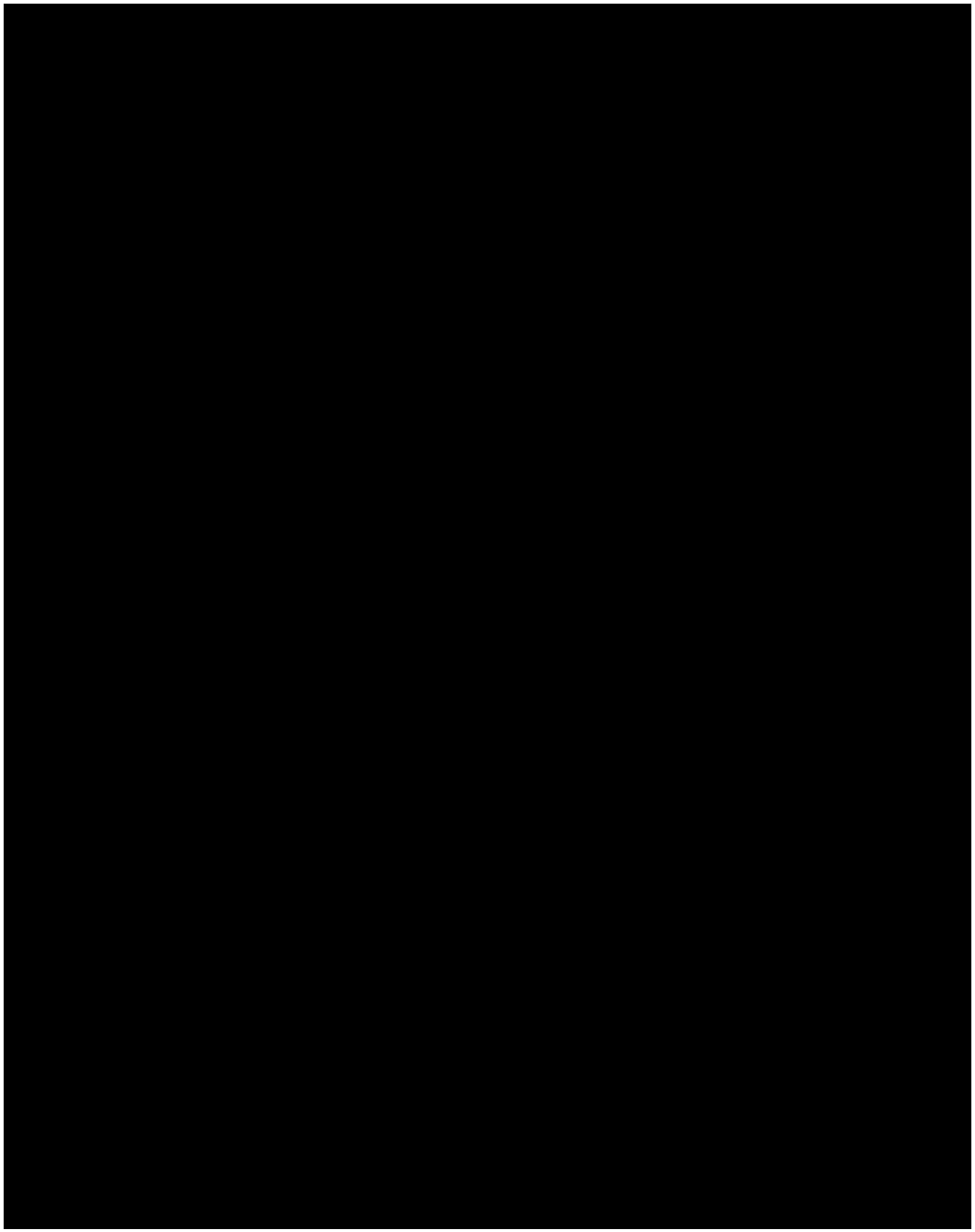
Number and percentage (n, %) of subjects with treatment satisfaction defined by a level of 1, 2 or 3 will be summarized and the corresponding two-sided 95% confidence intervals will be provided for the pooled treatment groups (overall).

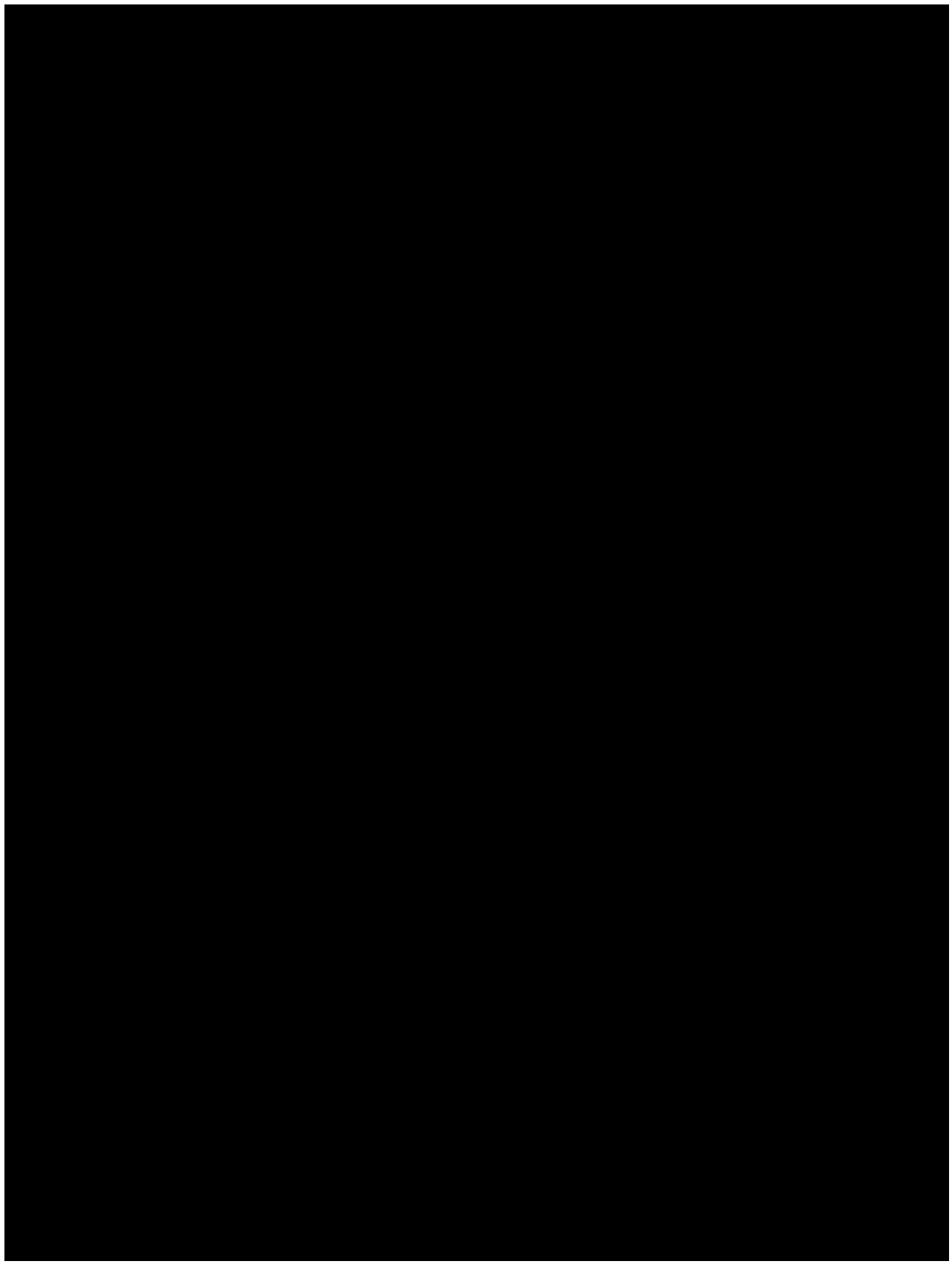
**GAIS**

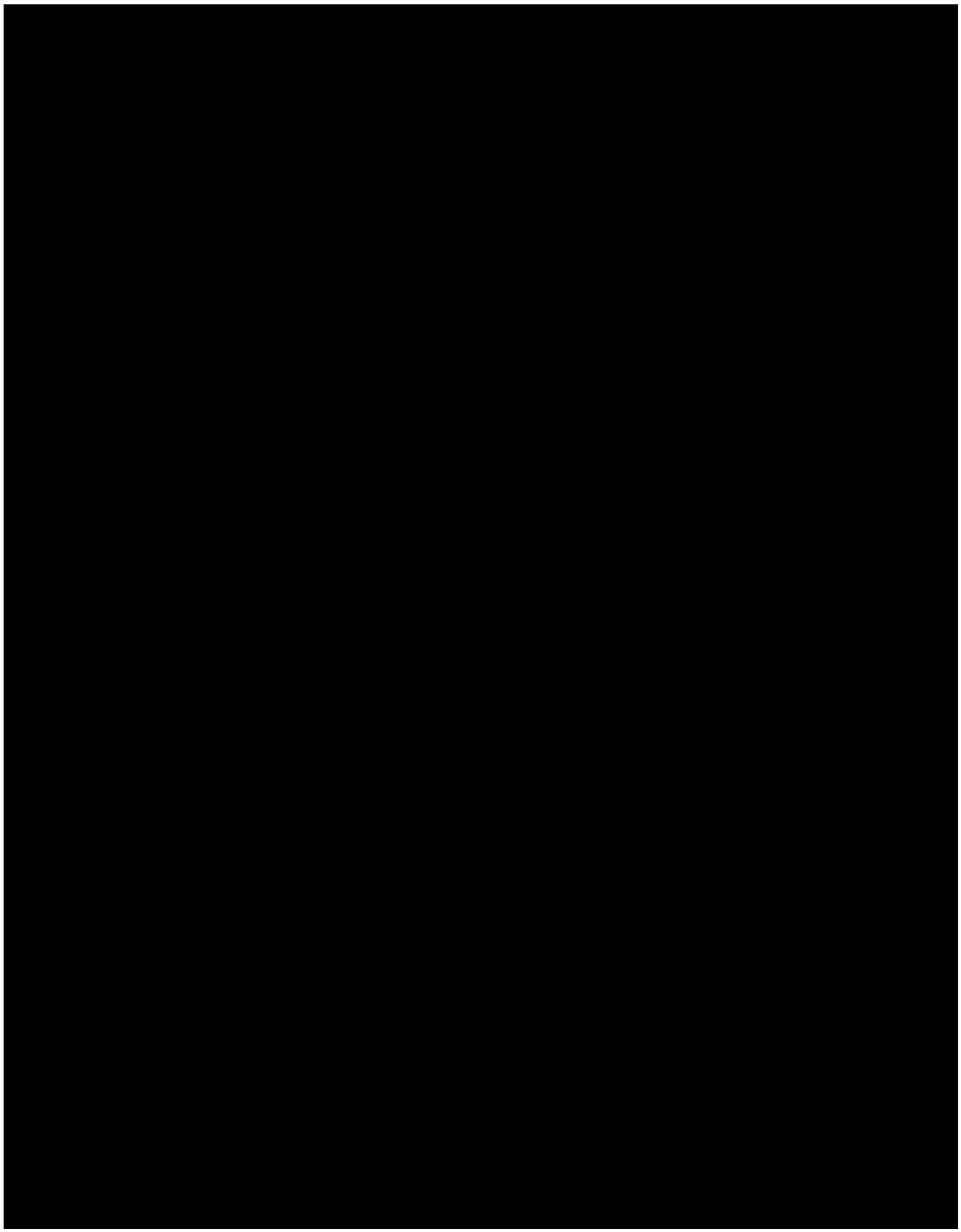
Number and percentage (n, %) of subjects with "improvement" in iGAIS-Wrinkles defined as a rating of 1, 2 or 3 at 16 weeks after last treatment will be summarized and two-sided 95% confidence intervals will be provided for the pooled treatment groups. Analogous methods will be used for sGAIS-Wrinkles.

[redacted]

[redacted]







[redacted]

## 7.2 Safety Endpoints

All safety analyses will be performed on the SES.

AEs will be coded according to the Medical Dictionary for Regulatory Activities (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 at or after date of first injection of the IMD. Incidences will be calculated for TEAEs on the system organ class (SOC) level and on the preferred term (PT) level (i.e., total and percentage, by intensity, and by relationship). Listings and (as

applicable) tables displaying incidences for TEAEs leading to discontinuation, serious TEAEs, and deaths will also be provided.

Each documented worsening of an AE will be aggregated with the previous corresponding AE record. The start date of the AE will be the date of the record reporting the onset of the AE with the first reported intensity and the stop date will be the last stop date of all aggregated worsening records or considered ongoing, if not yet resolved. For intensity, the worst attribute will be chosen.

CAVE: there is one exception, where an aggregation of the AE with corresponding worsening(s) is not allowed: if the AE record started before the first treatment and the worsening record started (worsened) at or after date of first treatment. In that case, only the AE record(s) starting at or after treatment will be aggregated and regarded as TEAE (see also below).

In case of missing intensity or missing causal relationship of an AE the worst case principle will be applied, i.e. a missing intensity will be set to “severe” and a missing causal relationship will be set to “related”. Missing data of the worst outcome will be imputed by “unknown”.

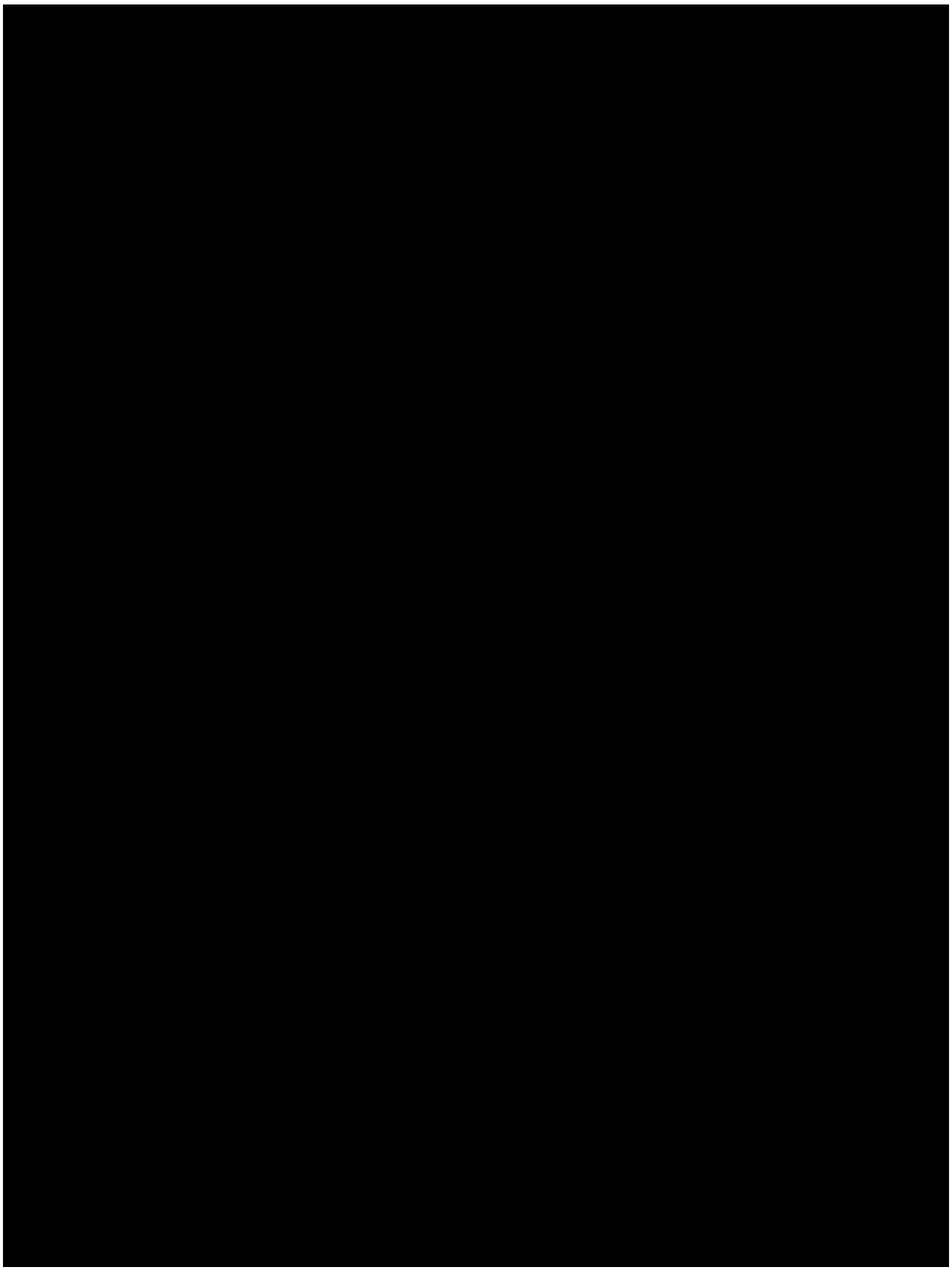
### *7.2.1 Primary Safety Endpoint(s)*

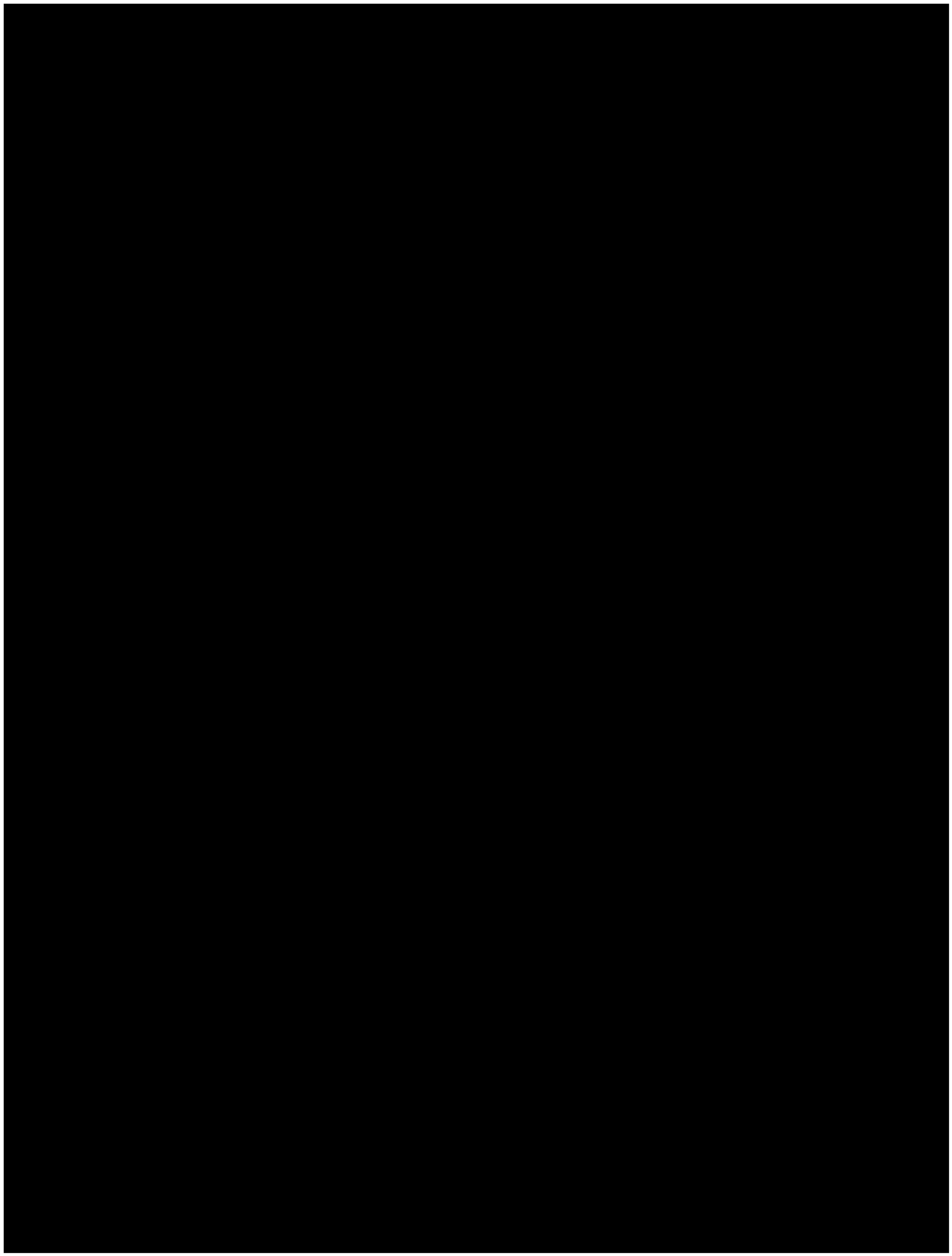
Not applicable.

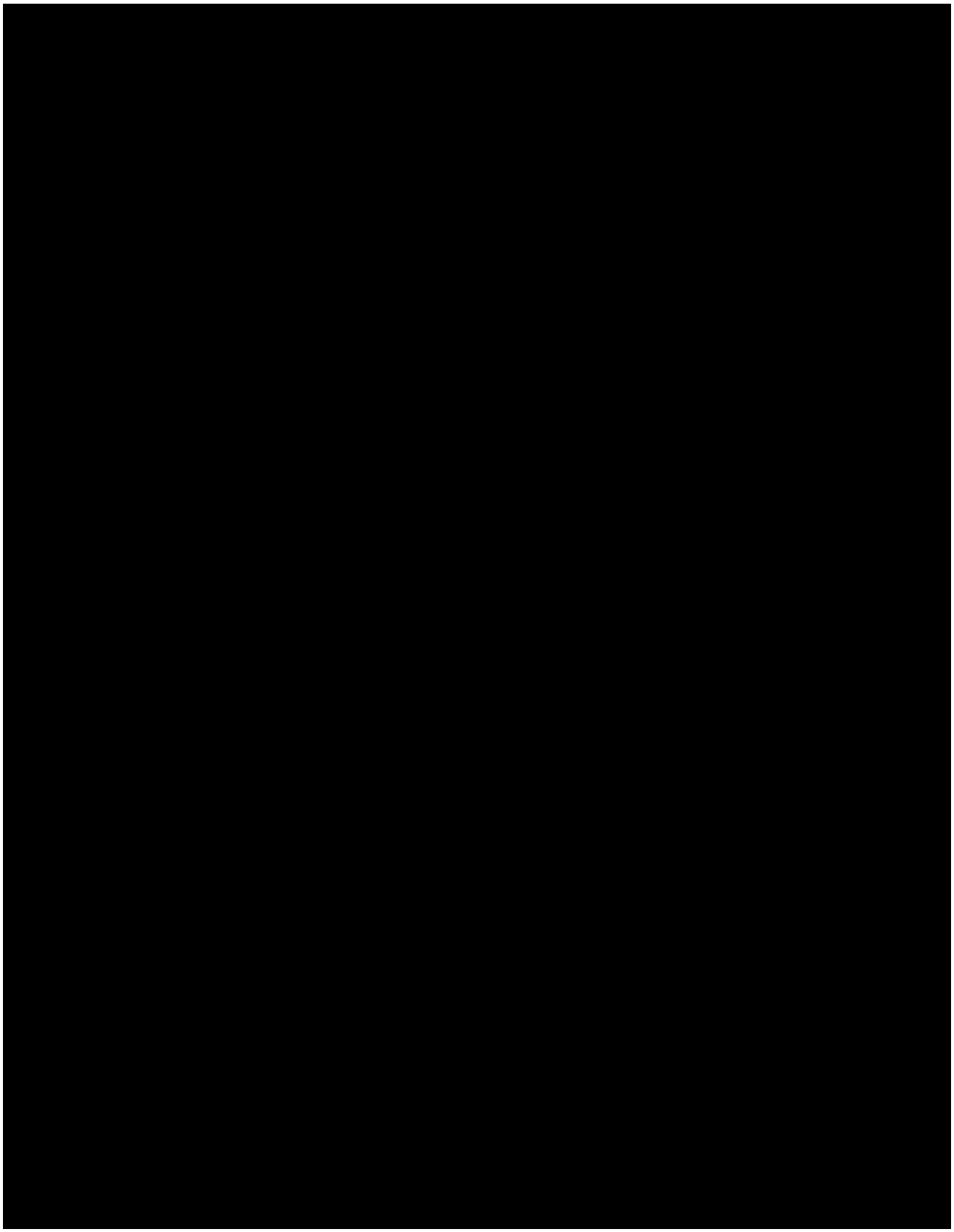
### *7.2.2 Secondary Safety Endpoints*

Incidences of related TEAEs will be provided by system organ class (SOC) and preferred term (PT).

[redacted]







[redacted]

## 7.4 Special Statistical/Analytical Issues

### *7.4.1 Discontinuations and Missing Data*

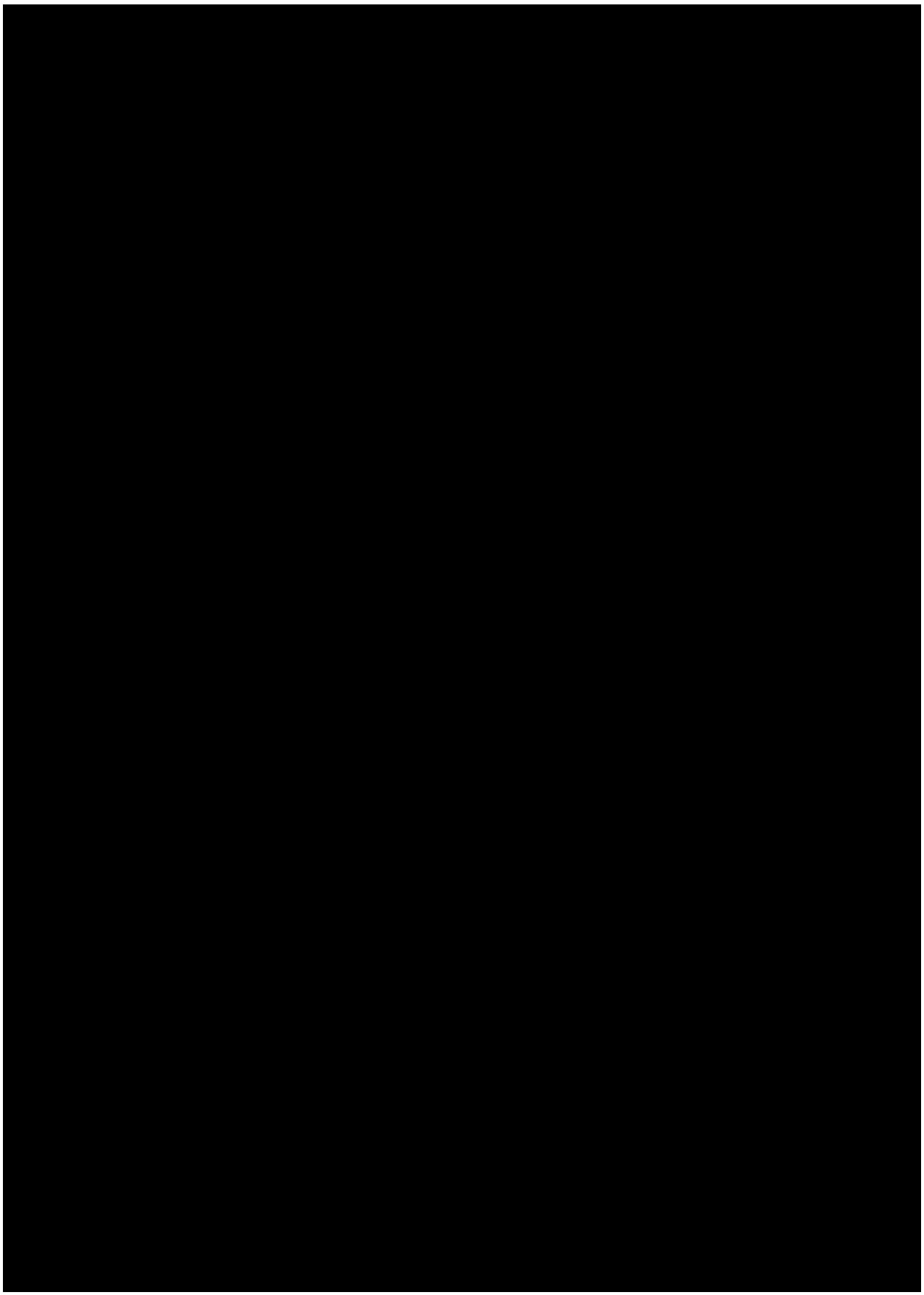
[redacted]

Multiple imputation (MI)

Missing post-baseline data for the primary endpoint will be replaced by multiple imputation. Multiple imputations will only be performed for subjects with a missing MAS 'Décolleté Wrinkles – At Rest' score at 16 weeks after last treatment who did not discontinue before 16 weeks after last treatment due to lack of effectiveness. Subjects who will discontinue due to the reason given above will be treated as non-responders.

All further specifications for MI are based on the assumption that the MAS 'Décolleté Wrinkles – At Rest' assessment at 16 weeks after last treatment is missing at random, i.e.; being missing will not depend on the MAS at 16 weeks after last treatment itself but will only depend on the factors in the imputation model. [redacted]

[redacted]



[redacted]

Last observation carried forward (LOCF)

For secondary effectiveness endpoints, missing values will be replaced by the last available value before 16 weeks after last treatment.

Observed cases (OC)

For the OC method, no missing value imputations will be conducted. All data will be analyzed as observed. This method will be used [redacted] to assess the sensitivity of the MI and the LOCF method for primary and secondary effectiveness endpoints.

[redacted]

[redacted]

### 7.4.2 Interim Analyses

Not applicable.

### 7.4.3 Data Monitoring Committee

Not applicable.

### 7.4.4 Multiple Comparisons/Multiplicity

[redacted] the type one error rate has not to be adjusted.

### 7.4.5 Examination of Subgroups

[redacted]

[redacted]

The supportive exploratory analyses for the primary effectiveness endpoint will also be performed for the observed cases in the FAS (i.e. on randomized treatment groups).

A further subgroup analysis for the following safety endpoints will be presented for the subgroup of subjects that had three injections with a cannula:

- Adverse events: Incidences of TEAEs, non-serious TEAEs and serious TEAEs will be provided by SOC and PT. Incidences of TEAEs will be provided overall and by worst intensity, by worst outcome, and by treatment area affected (yes/no). Treatment-related TEAEs will be presented by injection cycle.

[redacted]

### *7.4.6 Pooling of sites*

Pooling of investigational sites, will be performed when any investigational site does not have a sufficient number of effectiveness evaluable subjects in the modified FAS to avoid convergence problems in statistical models. Decision on pooling of sites will be taken during the DRM.

If for any of the models above convergence problems will be observed, the following steps for pooling of sites will be performed:

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

2) Regression models will be re-run with pooled site as factor

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

[redacted]

# 9 REFERENCES

[1] Ratitch, B., Lipkovich, I., O'Kelly, M. (2013) Combining Analysis Results from Multiply Imputed Categorical Data. Available at https://www.pharmasug.org /proceedings/2013/SP/PharmaSUG-2013-SP03.pdf [accessed August 25, 2020].

[2] Anne Lott & Jerome P. Reiter (2018): Wilson Confidence Intervals for Binomial Proportions With Multiple Imputation for Missing Data, The American Statistician, DOI: 10.1080/00031305.2018.1473796

[3] Barnard J, Rubin DB (1999). Small-sample degrees of freedom with multiple imputation, *Biometrika*, Volume 86, Issue 4, Pages 948–955, https://doi.org/10.1093/biomet/86.4.948