NCT03319719, redacted
version v1.0, 11Sep2019

Merz North America, Inc

Statistical Analysis Plan

Evaluation of Pain with Belotero® Balance with Integral Lidocaine for Correction of the Nasolabial Folds

Device Pre-Market

M930021001

Final Version 1.0

Date: 25-JAN-2018

Author: [redacted] Biostatistician

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] 26-JAN-2018 [redacted]
Author (print name) Date (dd-mmm-yyyy) Sign

[redacted] 26-JAN-2018 [redacted]
Peer Reviewer (print name) Date (dd-mmm-yyyy) Signature

[redacted] 26-JAN-2018 [redacted]
Study Medical Expert Merz (print name) Date (dd-mmm-yyyy) Si

TABLE OF CONTENTS

[redacted]

1 LIST OF ABBREVIATIONS

AE	Adverse event
ATC	Anatomical therapeutic chemical
BDRM	Blinded Data Review Meeting
CI	Confidence interval
CTR	Common treatment site response
[redacted]	[redacted]
FAS	Full analysis set
[redacted]	[redacted]
LOCF	Last observation carried forward
MedDRA	Medical Dictionary for Regulatory Activities
NLF	Nasolabial folds
OC	Observed cases
PPS	Per protocol set
SAE	Serious adverse event
SD	Standard deviation
SES	Safety evaluation set
TEAE	Treatment emergent adverse event
TFL	Tables / figures / listings
VAS	Visual analog scale
WHO	World health organization

2 GENERAL AND TECHNICAL ASPECTS

The objective of this statistical analysis plan is to specify the statistical analyses in more detail than stated in the clinical study protocol and to be precise enough to serve as a guideline for statistical programming and creation of tables, figures and listings (TFLs).

This statistical analysis plan is based on the clinical study protocol, dated 24-Jul-2017.

All programs will be written using SAS version 9.3 or higher for generating TFLs.

3 CLINICAL STUDY DESIGN AND OBJECTIVES

3.1 Clinical Study Design

This is a 6-week, prospective, randomized, split-face, double-blind, multicenter, pre-market study to demonstrate the effectiveness of Belotero Balance with integral lidocaine (Test product) for reduction of injection-related pain when compared to Belotero Balance without lidocaine (Control product) for the correction of NLFs. Approximately 52 subjects will be enrolled at approximately 3 investigational sites in the United States; a minimum of 20% of the enrolled subjects will be Fitzpatrick skin type IV, V, or VI. Subjects from the Fitzpatrick skin type IV, V, and VI group will be approximately evenly distributed within the IV, V, and VI categories. Subjects enrolled will have right and left NLF ratings of 2 or 3 (i.e., moderate or severe) on the Merz NLF Scale, as determined by the blinded evaluator (The Merz NLF Scale is an outcome measure with a 5-point ordinal rating for the assessment of NLF severity, ranging from 0=No Folds to 4=Very Severe Folds). Both NLFs must have the same rating (i.e., symmetrical NLFs) at baseline.

In this split-face design, a single randomization scheme will determine which NLF (subject's right or left) will be assigned to Test product or Control product and which NLF (subject's right or left) will be injected first. The treating investigator, the Merz NLF Scale evaluator, and the subject will be blinded to which NLF is randomized to Test and which NLF is randomized to Control.

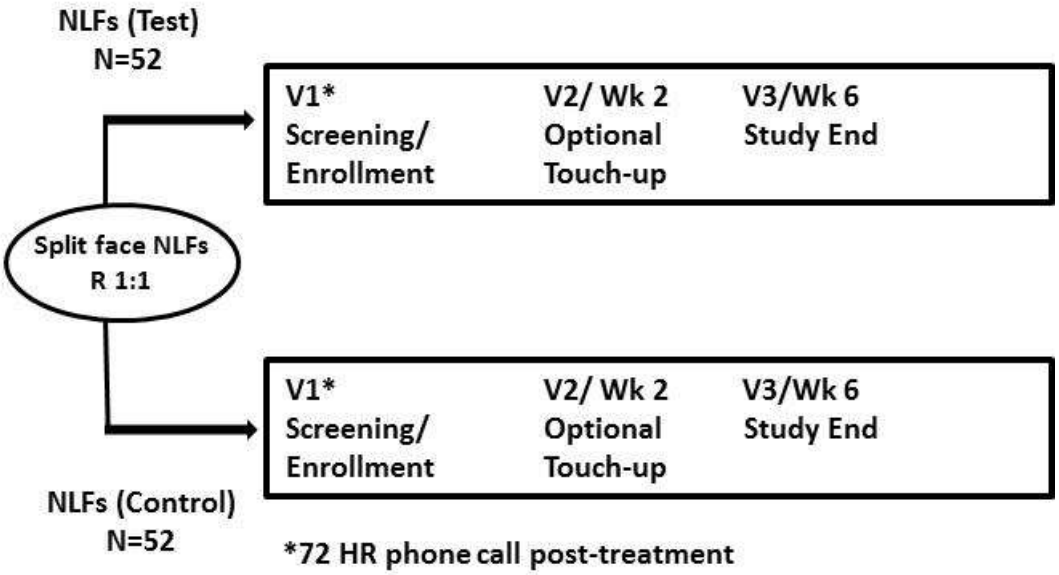
Study subjects will receive treatment with Belotero Balance with integral lidocaine in one NLF (Test) and Belotero Balance without lidocaine in the contralateral NLF (Control) on Day 1. All subjects without ongoing, treatment-related, investigator-reported AEs will be eligible for an optional touch-up treatment, at the investigator's discretion and to achieve optimal correction, at Week 2. Each treatment session will consist of multiple mid- to deep-dermal injections of 0.1 to 0.2 milliliters (mL) Belotero Balance with and without lidocaine (up to a maximum of 3 mL per NLF over the two treatment sessions) into the left or right NLF according to the randomization assignment and the instructions for use. Subjects will be evaluated at screening/enrollment, by phone at 72 hours after treatment, at Week 2 (optional touch-up treatment) and Week 6 (end of study). If subjects report a safety concern at the 72-hour phone call, an unscheduled visit will be arranged to address safety concerns. Of the 52

estimated subjects enrolled, each NLF will be analyzed as a separate unit of measure (i.e., either Test or Control NLF) for the primary and secondary efficacy endpoints, resulting in approximately 104 evaluable NLFs (52 Test, 52 Control). Pain levels will be compared between the Test NLF and Control NLF using a 10-cm VAS. Pain assessments will be evaluated at time zero (i.e., time the last injection needle is removed from each NLF, separately, after a full correction) [redacted].

Standard safety endpoints, including the incidence and nature of device- and/or injection-related AEs and SAEs observed during the study will be assessed. Furthermore, several pre-defined common treatment site responses (CTRs) will be evaluated via subject diaries.

Figure 1 illustrates the visit schedule associated with the split-face design.

Figure 1: Study design



NLFs: Nasolabial folds; N: sample size; V: Visit; Wk: Week; R: Randomization; HR: hour

Schedule of assessments is summarized in the following table:

Table 1 Scheduled of Assessments

Assessment	Visit 1	Phone call	Visit 2	Visit 3	
	Screening/ Enrollment	72-hours post treatment	Week 2	Week 6	Un-scheduled visit
	Day 1/ Baseline	Day 4 (± 1 d)	14 days (+ 3 d)	Day 42 (+ 3 d)	
Inclusion/Exclusion criteria	X				
Informed consent	X				
Urine pregnancy test[1]	X		X[6]	X	
Subject demographics	X				
Concomitant medication/Procedures	X	X	X	X	X
Medical history	X				
Standardized photographs	X			X	
Randomized Belotero Balance injection	X				
Optional touch-up injection[2]			X		
Subject VAS pain assessments[3]	X				
Subject pain-preference assessment[4]	X				
Blinded-evaluator NLF scale evaluation[5]	X			X	
[redacted]					
Safety/Adverse event assessment	X	X	X	X	X
Subject diary dispensed	X		X		
Subject diary review and collection			X	X	

[1] Only if female of childbearing potential; result is required to be negative prior to randomization.

[2] All subjects without ongoing investigator-reported AEs will be eligible for an optional touch-up treatment, at the investigator's discretion and to achieve optimal correction, at Week 2.

[3] VAS pain assessments are made for each NLF at multiple intervals during Visit 1 (i.e., time zero [the time the last injection needle is removed from each NLF, separately, after a full correction] [redacted]).

[4] This three-question assessment is conducted immediately after treatment of both NLFs (i.e., after completion of the time zero pain VAS) at Visit 1.

[5] The Merz NLF Scale evaluator will remain blinded throughout the study and is a qualified healthcare practitioner, delegated by the treating investigator and trained by the sponsor. Every effort should be made to ensure that the same blinded evaluator performs the Merz NLF Scale assessment at all visits.

[6] Pregnancy test is administered only to female subjects of childbearing potential and receiving a touch-up injection at Week 2.

3.2 Clinical Study Objectives

3.2.1 Effectiveness

The primary objective of the study is to demonstrate the effectiveness of Belotero Balance with lidocaine (Test product) for the reduction of injection-related pain in comparison to Belotero Balance without lidocaine (Control product) for the correction of NLFs.

3.2.2 Safety

The safety objectives include the identification and description of device- and/or injection-related adverse events (AEs) and serious adverse events (SAEs) as well as assessment of common treatment site responses (CTRs) during the course of the study.

4 DETERMINATION OF SAMPLE SIZE

Sample size calculation was performed using the nQuery+nterim version 4.0 statistical software. Based on a review of comparable clinical literature ([2-8]), it is assumed for this power calculation that a change of 3.8 cm on a 10-cm VAS represents a clinically meaningful difference in pain as experienced between lidocaine-treated and non-lidocaine-treated NLFs. Assuming a statistical power, 1–β of 95%, a one-sided level of significance of $\alpha = 0.025$, and a conservative estimate of standard deviation of difference of 7, a total of 47 subjects will be required to be enrolled into the trial. In order to accommodate likelihood of potential subject dropout after randomization and prior to treatment (up to 10% attrition), a total of 52 subjects will be enrolled, yielding approximately 104 total NLFs (52 Test and 52 Control). Additionally, a minimum of 20% of the enrolled total sample size will consist of subjects in the Fitzpatrick Skin Types IV, V, or VI group. Subjects from the Fitzpatrick skin type IV, V, and VI group will be approximately evenly distributed within the IV, V, and VI categories.

5 ANALYSIS SETS

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

Safety Evaluation Set (SES)

The Safety Evaluation Set (SES) will comprise subjects who are enrolled into the study, randomized, and receive an injection.

Full Analysis Set (FAS)

The Full Analysis Set (FAS) will consist of all randomized subjects and will be analyzed as randomized. This will be the primary population used for the efficacy analyses..

Per Protocol Set (PPS)

The Per Protocol Set (PPS) is a subset of subjects in the FAS without major protocol deviations. Final determination of what constitutes major or minor protocol deviations will be made prior to database lock during the Blinded Data Review Meeting (BDRM).

6 ENDPOINTS FOR ANALYSIS

6.1 Effectiveness Endpoints

6.1.1 Primary Effectiveness Endpoint

The primary effectiveness endpoint is the pain, measured using a 10-cm visual analogue scale (VAS), immediately after split-face NLF treatment (i.e., at time zero) on Day 1. Time zero is defined as the time the last injection needle is removed from each NLF, separately, after a full correction on Day 1.

6.1.2 Secondary Effectiveness Endpoint

The secondary effectiveness endpoint is the response according to the Merz NLF Scale for each NLF as assessed live by the blinded evaluator at Week 6. Response is defined as ≥ 1-point improvement on the Merz NLF Scale for each NLF compared to baseline. Each NLF is to be assessed separately. The blinded evaluator is a qualified healthcare practitioner, delegated by the treating investigator.

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

Continuous variables (values and changes from baseline) will be summarized by n, mean, standard deviation, median, minimum, and maximum. For qualitative variables, absolute and percent frequencies (n, %) will be displayed. Confidence limits and p-values will be given, where appropriate.

P-values when presented will be reported to four decimal places (e.g. p=0.0375). P-values below 0.0001 will be presented as '<0.0001'.

Minimum and maximum will be reported to the same number of decimal places as the data were collected; mean and median will be reported to one decimal place greater than the data were collected, and standard deviation to two decimal places greater. For derived data an adequate number of decimals will be chosen. Percentages will be calculated using the denominator of all subjects in a specified population or treatment group with available (non-missing) value of the variables in question. 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 Effectiveness Variables

The analysis of the primary and the secondary effectiveness variables will be based primarily on the FAS and additionally, for sensitivity purposes, on the PPS. All other effectiveness variables will be analyzed based on the FAS. Statistical tests will be two-sided at $\alpha = 0.05$

except for the primary effectiveness analysis where a one-sided $\alpha = 0.025$ will be used. Two-sided confidence limits (95%) and p-values associated with the appropriate test statistic will be provided where necessary.

7.1.1 Primary Effectiveness Variable

The analysis of the reduction in pain, measured using a 10-cm VAS, in the Test NLF compared to the Control NLF at time zero will be performed using a paired t-test to test the one-sided hypothesis that the mean of the differences in VAS scores between Test (μ_T) and Control (μ_C) folds is either greater than/equal to zero (≥ 0) or less than zero (< 0), as equivalently assessed using the following hypothesis-testing notations:

H_0 (null): $\mu_C \leq \mu_T$, (i.e. $\Delta = \mu_T - \mu_C \geq 0$) versus

H_a: (alternate): $\mu_C > \mu_T$, (i.e. $\Delta = \mu_T - \mu_C < 0$)

For testing this hypothesis using the t distribution, H_0 will be rejected if the upper limit of the 95% confidence interval (CI) around Δ is < 0 using a one-sided p-value ($\alpha = 0.025$), otherwise H_0 will not be rejected.

Descriptive summary statistics for continuous variables will be presented for VAS pain assessments in Test and Control NLFs as well as for the difference between Test and Control i.e. (Test – Control).

7.1.2 Secondary Effectiveness Variables

Number and percentage of Merz NLF Scale responders at Week 6 will be presented for the Test and Control NLF sides. The p-value from McNemar's test comparing responses paired by subject will be presented to compare both percentages.

[redacted]

7.2 Safety Variables

All safety analyses will be performed on the SES.

AEs will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA, version 20.0). Only treatment emergent adverse events (TEAEs) will be analyzed, which are defined as AEs with onset or worsening after the administration of the study drug. Incidences will be calculated for TEAEs on the system organ class level and on the preferred term level (total, by intensity, and by relationship – related or not related). Listings and tables displaying incidences for TEAEs leading to discontinuation and serious TEAEs will also be provided.

TEAEs associated with NLFs will be summarized by treatment. All TEAEs will be summarized without by-treatment breakdown.

In summaries by system organ class and preferred term each subject will be counted once per preferred term and treatment if applicable. Similarly in summaries by system organ class each subject will be counted once per system organ class and treatment if applicable. In summaries that are additionally broken down by intensity or relationship, subjects with multiple events will be summarized by the worst severity or worst relationship, as applicable.

Number and percentage of subjects with CTR, as well as the number of continuous CTR episodes, will be summarized by treatment and maximum severity as well as by treatment and maximum duration. Duration will be classified as 1-3 days, 4-7 days, 8-14 days 15-30 days and > 30 days.

7.3 Other Variables

Subject disposition, demographic characteristics and baseline characteristics will be presented using standard descriptive statistics for continuous variables (i.e. n, mean, SD, median, minimum, and maximum) and for qualitative variables (absolute and percent frequencies (i.e. n, %). Demographic and baseline characteristics will be summarized for the SES, the FAS, and the PPS.

Prior and concomitant treatments as well as non-pharmacological concomitant procedures will be listed. Prior and concomitant medications will be coded using WHO Drug Dictionary version Sept 2016 B2.

Frequencies of concomitant medications will be given based on different ATC code levels as well as by generic name for the SES. Indications for concomitant procedures will be coded and only listed. Medications will be considered prior of they have a stopping date that is prior to the first study injection. Medications that are ongoing or have stopped on or after the date of the first study injection will be considered concomitant.

Medical history and concomitant diseases will be summarized based on MedDRA (version 20.0) system organ class and preferred term levels for the SES. Medical history is defined as conditions that have a stopping date that is prior to the first study injection. Concomitant diseases are defined as conditions that are ongoing or have stopped on or after the date of the first study injection.

7.4 Special Statistical/Analytical Issues

7.4.1 Discontinuations and Missing Data

Imputation of effectiveness data

Observed Cases (OC)

- For the OC method, there will be no missing value imputations conducted. All effectiveness data will be analyzed as observed data.

Last Observation Carried Forward (LOCF)

- For the LOCF, in the analysis of the secondary effectiveness variable the subjects who have a missing score on the Merz NLF scale at Week 6 will have their latest available assessment (from the early discontinuation or unscheduled visit if available; failing

that, from the baseline visit) carried forward and imputed into the missing Week 6 assessment.

Imputation of safety data

For partial or completely missing start date of adverse events, the following imputation rules related to any treatment period will be used if applicable:

- If only the day is missing for any date used in a calculation, the 1st will be used to replace the missing day. If the new estimated date falls before the date of first dose, while the known month and year match the month and year of the first dose, the date of first dose will be used as the new estimated date.
- If the day and the month are missing, no estimation will be performed. However, if the year of onset is the same or greater than the year of the first dose date, the AEs will be considered treatment emergent and included in the summary tables.
- AEs with completely missing onset dates will be considered treatment emergent; no estimation of the dates will be performed.

Duration of CTRs would be computed using the following logic:

- Each subject is expected to have diary entries from Day 1 (day of the initial injection) up to and including Day 42 or the day of their completion/discontinuation, whichever comes first. If any subject does not have a diary entry for any of the required days, then for the purposes of CTR analysis it will be assumed that on the missing days they had all those CTRs (if any) that they had on the day before or on the day after the missing day, with the same severity.
- For any CTR with an uninterrupted single episode event, the duration will be computed as the days from the beginning of event to last date of event.
- For any CTR with multiple episodes, the CTR will be regarded as a continuous event if the date between paired episodes is not more than 1 day apart.
- For any CTR with multiple episodes, the CTR will not be regarded as a continuous event if the date between paired episodes is at least 2 days apart. In such instances, the duration with the longest duration will be used for that particular CTR.
- For any CTR with multiple continuous and not continuous events, the logic as aforementioned will be adopted, in which case the maximum of all CTR durations will be selected as the CTR duration to be used for summary purposes.

7.4.2 Interim Analyses

Not applicable.

7.4.3 Data Monitoring Committee

Not applicable

7.4.4 Multiple Comparisons/Multiplicity

Not applicable

7.4.5 Examination of Subgroups

Not applicable

8 CHANGES IN THE PLANNED ANALYSES

The original protocol (Version 1.0, dated 26APR2017) specified that "In the event of occurrence of missing primary effectiveness data on any of the NLFs, the multiple imputation (MI) will be used for imputation of missing data". However, by the definition of the FAS, there will not be any missing primary effectiveness data in the FAS. As a result, no imputation for missing data will be possible.

The secondary effectiveness variable will now be analyzed both with and without imputation. The post-baseline Last Observation Carried Forward (LOCF) imputation will thus be used for imputation of missing values.

9 REFERENCES

1. Study protocol: Evaluation of Pain with Belotero® Balance with Integral Lidocaine for Correction of the Nasolabial Folds, dated 24-JUL-2017.

2. Grunebaum LD, Elsaie ML, Kaufman J. Six-month, double-blind, randomized, split-face study to compare the efficacy and safety of calcium hydroxylapatite (CaHA) mixed with lidocaine and CaHA alone for correction of nasolabial fold wrinkles. *Dermatol Surg*. 2010; 36(supp 1):760-765.

3. Marmur E, Green L, Busso M. Controlled, randomized study of pain levels in subjects treated with calcium hydroxylapatite premixed with lidocaine for correction of nasolabial folds. *Dermatol Surg*. 2010; 36(3):309-15.

4. Weinkle S. Efficacy and tolerability of admixing 0.3% lidocaine with Dermicol-P35 27G for the treatment of nasolabial folds. *Dermatol Surg*. 2010; 36(3):316-320.

5. Schachter D, Bertucci V, Solish N. Calcium hydroxylapatite with integral lidocaine provides improved pain control for the correction of nasolabial folds. *J Drugs Dermatol*. 2016; 15(8):1005-1010.

6. Weiss R, Bank D, Brandt F. Randomized, double-blind, split-face study of small-gel-particle hyaluronic acid with and without lidocaine during correction of nasolabial folds. *Dermatol Surg*. 2010; 36(suppl 1): 750-759.

7. Hedén P, Fagrell D, Jernbeck J, Rylander R, Samuelson U, Sellman G, Stark B. Injection of stabilized hyaluronic acid-based gel of non-animal origin for the correction of nasolabial folds: Comparison with and without lidocaine. *Dermatol Surg*. 2010; 36(suppl 1): 775-781.

8. Brandt F, Bank D, Cross SL, Weiss R. A lidocaine-containing formulation of large-gel particle hyaluronic acid alleviates pain. *Dermatol Surg*. 2010; 36(suppl 3):1876-1885.