Statistical Analysis Plan for Study CMA-MA-FAS-0580

HARMONY Brazil: Prospective, Multi-site, Study to Evaluate Subject Satisfaction with Facial Appearance Overall and the Aesthetic and Psychosocial Impact of Combined Facial Treatment

Date: 05 June 2024

Version 3.0

Table of Contents

1.0 Introduction

This Statistical Analysis Plan (SAP), based on Protocol Amendment 1 Version 2, 18-Oct-2021, describes the statistical analyses for AbbVie Study CMA-MA-FAS-0580 HARMONY Brazil: Prospective, Multi-site, Study to Evaluate Subject Satisfaction with Facial Appearance Overall and the Aesthetic and Psychosocial Impact of Combined Facial Treatment. Specifications of tables, figures, and data listings are contained in a separate document.

Study CMA-MA-FAS-0580 is a prospective, interventional, multicenter, combination filler medical device and drug post marketing study. Approximately 60 subjects will be enrolled consecutively at approximately 7 study sites in Brazil. Recruitment will also target men, who are typically under-represented in aesthetic studies. Each subject will be in the study for up to approximately 5 months and act as his/her own control. Subjects will be evaluated by the investigator to receive BOTOX, HArmonyCa Lidocaine facial filler, and Juvéderm facial fillers (JUVÉDERM VOLITE [redacted], JUVÉDERM VOLBELLA with Lidocaine, JUVÉDERM VOLIFT with Lidocaine, JUVÉDERM VOLUMA with Lidocaine, and JUVÉDERM VOLUX with Lidocaine). Selected subjects will receive staged treatment with the multiple products and will be evaluated using validated patient-reported outcome measurements to determine the psychological and emotional impact of an integrated treatment approach to reduce visible signs of aging. All products will be used as per their license only.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

Unless noted otherwise, all analyses will be performed using SAS Version 9.4 (SAS Institute Inc., Cary, NC 27513) or later.

This SAP includes changes to analyses described in the protocol. Details are outlined in Appendix C.

2.0 Study Design and Objectives

2.1 Objectives and Hypotheses

The primary objective of this study is to quantify the psychological, social, and emotional impact of comprehensive aesthetic treatment with a portfolio of AbbVie's products by measuring the change in subject's satisfaction with facial appearance from baseline to the final study visit

(Statistical Programming Plan [SPP] Appendix C FACE-Q™: Satisfaction with Facial Appearance). These products include BOTOX, HArmonyCa Lidocaine, and Juvéderm facial fillers, as described in Section 1.0 and Section 2.3.

The secondary objectives are:

- To determine the subject's self-appraisal of aging appraisal (SPP Appendix D FACE-Q™: Aging Appraisal), psychological function (SPP Appendix E FACE-Q™: Psychological Function), social function (SPP Appendix F FACE-Q™: Social Function) and satisfaction with skin (SPP Appendix G FACE-Q™: Satisfaction with Skin).

- To quantify the investigator's and subject's assessment of aesthetic improvement.

The clinical hypothesis of this study is that subjects treated with a combination of BOTOX, HArmonyCa Lidocaine facial filler, and Juvéderm facial fillers will be satisfied with their overall facial appearance at the final study visit as measured by the change from baseline on the FACE-Q™ Satisfaction with Facial Appearance Overall Scale.

2.2 Study Design Overview

This Phase 4 study is a prospective, interventional, multicenter, combination medical device and drug post marketing study. Each subject will be in the study for up to approximately 5 months and act as his/her own control. All products will be used as per their license only. The investigators will perform all treatments and evaluations.

The study is planned to enroll 60 subjects consecutively at approximately 7 study sites in Brazil. Recruitment will also target men, who are typically under-represented in aesthetic studies.

All subjects will sign an informed consent form (ICF) and begin the study at the screening visit (V0). Subjects following screening (V0) will enter the study and receive an initial treatment of HArmonyCa Lidocaine at Visit 1 (V1), followed by Juvéderm fillers at Visit 3 (V3), and BOTOX at Visit 5 (V5) as indicated in the study design diagram (Figure 1). If the decision is made to not treat with HArmonyCa Lidocaine, the subject may begin with Juvéderm fillers at V1. As indicated in the diagram, touch-ups of HArmonyCa at Visit 2 (V2), Juvéderm at Visit 5 (V5), and BOTOX at Visit 6 (V6) are optional. If a touch-up of HArmonyCa (V2) is not performed, the subject may advance in the study timeline to the next study visit procedures. If a touch-up of BOTOX is not performed, the subject will return on schedule for Visit 7 (V7). A

subject not receiving Juvéderm touch-up would continue with BOTOX treatment (and all study procedures) outlined for V5. To reduce the number of subject visits, V0 can be combined with the initial treatment visit (V1), if judged by the investigator to be appropriate and acceptable for the subject. End of study is the last clinical visit an enrolled subject has in this study, V7 (after approximately 5 months of treatment). Subjects who discontinue prematurely will be asked to return for the last clinical visit for end-of-study data collection. Identical study procedures will not be repeated where the visits are combined. The schedule of activities is shown in Table 1.

Figure 1. Study Design

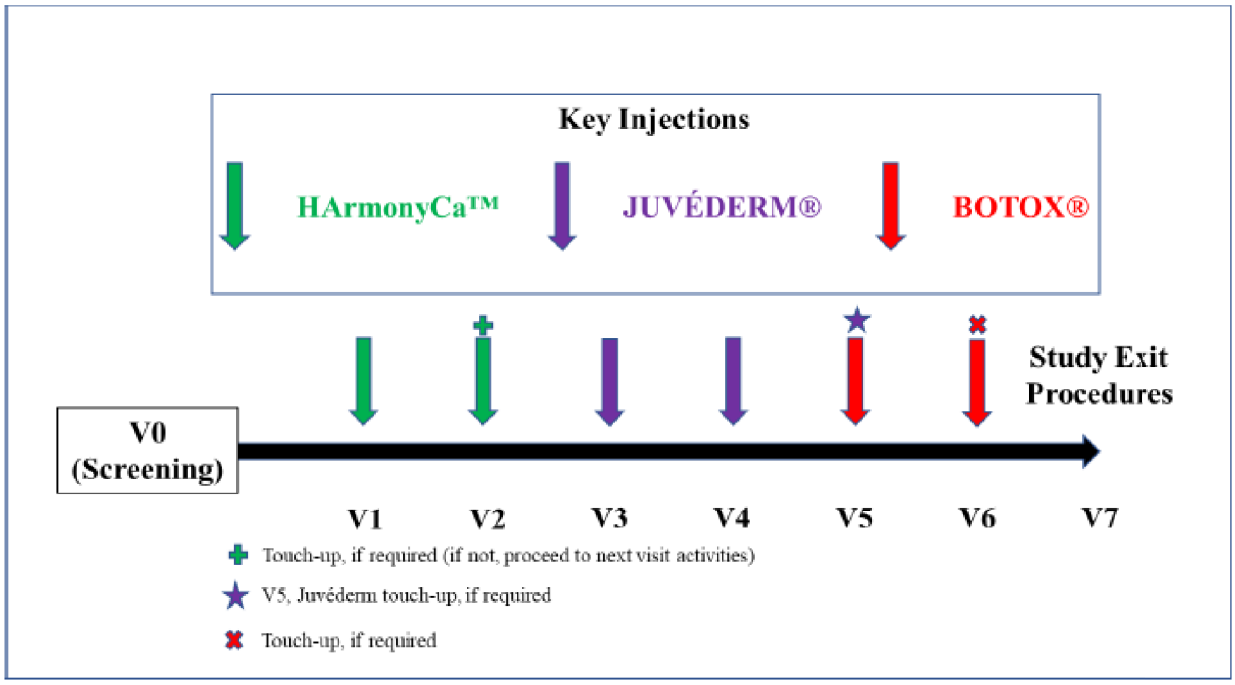

Table 2. Schedule of Activities (SOA)

Study Visit	V0[1]	V1	V2	V3	V4	V5	V6[2]	V7[3]
	Screening	Day 1	Day 30	Day 60	Day 90	Day 120	Day 135	Day 150
Visit window	(Day -14 to Day -1)	N/A	(30 ± 7 days after V1)	(30 ± 7 days after V2)	(30 ± 7 days after V3)	(30 ± 7 days after V4)	(15 + 7 days after V5)	(15 ± 7 days after V6)
	STUDY TREATMENTS							
HarmonyCa Injections[4, 13]		X	X[14]					
Juvéderm Injections[4]				X	X	X[5]		
BOTOX[4]						X	X[6]	
	STUDY PROCEDURES							
Provide subject information; review/verify study criteria	X							
Informed consent form	X							
Demographics	X							
Medical history	X							
Weight/BMI[7]	X							X
Pregnancy test[8]	X	X	X	X	X	X	X[9]	
Concomitant medications	X	X	X	X	X	X	X	X
2D facial photography	X							X
Standardized digital video	X							X
Adverse events	X	X	X	X	X	X	X	X
Fitzpatrick Skin Phototype	X							
Subject diary		X	X[10]	X	X	X[10]		

Study Visit	V0[1]	V1	V2	V3	V4	V5	V6[2]	V7[3]
	Screening	Day 1	Day 30	Day 60	Day 90	Day 120	Day 135	Day 150
Visit window	(Day -14 to Day -1)	N/A	(30 ± 7 days after V1)	(30 ± 7 days after V2)	(30 ± 7 days after V3)	(30 ± 7 days after V4)	(15 + 7 days after V5)	(15 ± 7 days after V6)
	STUDY RATING SCALES AND QUESTIONNAIRES							
Self-perception of age (SPA)	X							X
FACE Q[11]	X							X[12]
Periorbital aesthetic appearance questionnaire (PAAQ)	X							X
FWS – glabellar lines (GLs)	X							
FWS – crow's feet lines (CFLs)	X							
FWS –forehead lines (FHLs)	X							
Investigator global aesthetic improvement scale (GAIS)								X
Subject GAIS								X

Abbreviation: BMI, body mass index; CFLs, crow's feet lines; FHLs, forehead lines; FWS, facial wrinkle scale; GAIS, global aesthetic improvement scale; GLs, glabellar lines; PAAQ, periorbital aesthetic apperance questionnaire; SPA, self-perception of age.

1 Screening and Visit 1 can be combined in order to reduce the total number of visits, if judged by the investigator to be appropriate and acceptable for the subject. Identical study procedures will not be repeated where the visits are combined.

2 If no BOTOX touch-up is required, the subject will still return for V7.

3 All subjects attend V7 for study exit procedures and assessments.

4 Prior to attending study visits, subjects must not apply facial cosmetics.

5 Juvéderm filler, if required

6 BOTOX touch-up if required

7 BMI > 30 is exclusionary. Body weight will be measured at screening and end of study (V7) for all subjects. Drastic weight loss or gain may skew assessments of "before" and "after" photos.

8 Required for women of childbearing potential

9 Pregnancy test not required at V6 if no BOTOX touch-up is needed.

10 New subject diary only administered at touch-up if injection occurred. Collect completed diary at the next scheduled visit.

[11] FACE-Q Scales include Satisfaction with Facial Appearance Scale, Expectations Scale (baseline only), Aging Appraisal, Psychological Function Scale, Social Function Scale, and Satisfaction with Skin Questionnaire.

[12] FACE-Q Expectations subscale will not be collected at End of Study.

[13] If HArmonyCa injections are not to be administered to a subject, the subject will begin at Day 1 with V3 study activities (i.e., Juvéderm fillers).

[14] HArmonyCa touch-up, if required. If not required, proceed with V3 study activities.

2.3 Treatment Assignment and Blinding

Approximately 60 subjects will be enrolled consecutively at approximately 7 study sites in Brazil. Recruitment will also target men, who are typically under-represented in aesthetic studies. Study treatments include:

- **BOTOX**: The dose and the number of injection points may be adapted to the needs of the subject, based on their characteristics and location of the muscles to be treated. Subjects will receive at least one BOTOX treatment in the following areas: glabellar lines (GLs), crow's feet lines (CFLs), and forehead lines (FHLs).

- **HArmonyCa Lidocaine**: The investigator will determine the appropriate volume to be injected during initial and touch-up treatments based on his/her clinical experience. The maximum total volume of HArmonyCa Lidocaine allowed per subject for the initial and touch-up treatment combined is 4 syringes (5 milliliter [mL]).

- **JUVÉDERM**: The investigator will determine the appropriate volume to be injected during initial and touch-up treatments based on his/her clinical experience.

    - JUVÉDERM VOLITE [redacted]: The maximum total volume of JUVÉDERM VOLITE [redacted] allowed per subject for the initial and touch-up treatments combined is 4 syringes (4 mL).

    - JUVÉDERM VOLBELLA with Lidocaine: The maximum combined total volume of JUVÉDERM VOLBELLA with Lidocaine allowed per subject for all facial areas treated for both the initial and touch-up treatments combined is 2 syringes (2 mL).

    - JUVÉDERM VOLIFT with Lidocaine: The maximum total volume of JUVÉDERM VOLIFT with Lidocaine allowed per subject for the initial and touch-up treatments combined is 4 syringes (4 mL).

    - JUVÉDERM VOLUMA with Lidocaine: The maximum total volume of JUVÉDERM VOLUMA with Lidocaine allowed per subject for the initial and touch-up treatments combined is 4 syringes (4 mL).

- JUVÉDERM VOLUX with Lidocaine: The maximum total volume of JUVÉDERM VOLUX with Lidocaine allowed per subject for the initial and touch-up treatments combined is 4 syringes (4 mL).

The investigator should not inject greater than 20 mL of filler (Juvéderm and HArmonyCa Lidocaine) per 60 kg (130 lbs.) body mass per year for a single subject.

Blinding and randomization are not applicable. This is an open-label, single-arm study. Subjects are enrolled consecutively based upon subject need and all subjects are intended to receive all treatments. Investigators will perform all treatments and evaluations.

2.4 Sample Size Determination

A sample size of 48 subjects will provide 90% power to detect if change from baseline to end of study visit in the FACE-Q^{TM}: Satisfaction with Facial Appearance Overall Scale Rasch-transformed score is different from zero, assuming a mean change of 12.0 and a standard deviation of 25.0.

This calculation is based on a one-sample, 1-sided t-test at 0.025 significance level. The assumptions of mean and standard deviation are based on FACE-Q^{TM} results from the HARMONY Study GMA-CMB-14-001. Anticipating a potential 20% drop-out rate (for early withdrawals and/or subjects who failed to duly complete the FACE-Q^{TM}: Satisfaction with Facial Appearance Overall Scale assessment at end of study), 60 subjects will be enrolled.

PASS 2008 software was used to calculate the sample size.

3.0 Endpoints

3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline in the overall Rasch-transformed score of the FACE-Q^{TM}: Satisfaction with Facial Appearance Scale to the final study visit (V7).

3.2 Secondary Efficacy Endpoints

3.2.1 Key Secondary Efficacy Endpoints

- Subject's assessment of aging appraisal as measured by the change from baseline in the Rasch-transformed score of the FACE-Q^{TM}: Aging Appraisal to final study visit (V7)

- Subject's assessment of psychological function as measured by the change from baseline in the Rasch-transformed score of the FACE-Q^{TM}: Psychological Function Scale to final study visit (V7)

- Subject's assessment of social function as measured by the change from baseline in the Rasch-transformed score of the FACE-Q^{TM}: Social Function Scale to final study visit (V7)

- Subject's assessment of satisfaction with skin as measured by the change from baseline in the Rasch-transformed score of the FACE-Q^{TM}: Satisfaction with Skin to final study visit (V7)

- Subject's assessment of age-related facial appearance by the Self-Perception of Age (SPA) measure at baseline to final study visit (V7); responders are defined as having achieved a younger category.

3.2.2 Other Secondary Efficacy Endpoints

- Investigator's assessment of global facial aesthetic improvement as measured by the 5-point Global Aesthetic Improvement Scale (GAIS) at final study visit (V7)

- Subject's assessment of global facial aesthetic improvement as measured by the 5-point GAIS at final study visit (V7)

- Change in subject's satisfaction with appearance of periorbital area as measured by the change from baseline in the Periorbital Aesthetic Appearance Questionnaire (PAAQ) to final study visit (V7)

3.3 Safety Endpoints

The safety endpoints will include the number and percentage of subjects who have any adverse events (AEs), injection site reactions (ISRs) or device deficiencies.

All AEs and injection site reactions (ISRs) will be collected from when the ICF is signed through to the last study visit (V7). Serious adverse events (SAEs) will be reported within the specified time and format directly to the Sponsor.

4.0 Analysis Populations and Subgroup

4.1 Analysis Populations

4.1.1 Modified Intent-to-Treat (mITT) Population

The mITT Population is defined as all subjects who enrolled and had at least one post-baseline efficacy assessment at V7.

4.1.2 Evaluable Population

The Evaluable Population is defined as all subjects in the mITT Population who received study products (JUVÉDERM and BOTOX) as treatments in this study, had at least one efficacy assessment at the final visit (V7), and had no significant protocol violations or deviations that will affect any efficacy endpoint. Such significant protocol violations or deviations will be determined during data review.

4.1.3 Safety Population

The Safety Population is defined as all subjects who received any product as treatment in this study.

Unless otherwise specified, all efficacy analyses will be based on the Evaluable Population. The mITT Population will be used for demographics and baseline characteristics summary. The Safety Population will be used for the protocol deviation and violation, treatment exposure, all safety analyses, and listings unless otherwise specified.

4.2 Subgroup

Subgroup analysis will be performed for the primary efficacy analysis by gender, as male and female subjects may have differing expectations/perceptions of aesthetic improvement; they will be examined separately.

5.0 Subject Disposition and Treatment Disposition

The number and percentage of subjects who are in the mITT Population, Evaluable Population, and Safety Population will be summarized by treatment group (HArmonyCA, Non-HArmonyCA), gender (male, female), and overall, for all eligible subjects (subjects who signed the informed consent and met all eligibility criteria). Screen failure subjects and the reasons for screen failure will also be summarized by overall, for all screened subjects.

A listing of screen failure subjects, inclusion/exclusion criteria assessment and all analysis populations will be presented for all screened subjects.

The number and percentage of subjects who complete the study and who discontinue from the study will be presented by treatment group, gender, and overall, for all screened subjects. The number and percentage of subjects who complete all treatments (HArmonyCA, JUVÉDERM, and BOTOX) and discontinue from any of the treatments (HArmonyCA, JUVÉDERM, and BOTOX) as collected on the CRF will also be presented by treatment group, gender, and overall, for all screened subjects. The primary reason for study discontinuation and treatment discontinuation will be presented as collected on the eCRF. Reasons for discontinuation include:

- Adverse Event
- Lost to Follow-Up
- Pregnancy
- Withdrawal by Subject
- Study Terminated by Sponsor
- Physician Decision
- Protocol Deviation
- Other

A listing of study disposition and treatment disposition will be provided for the Safety Population.

The duration in study (days) will be summarized by treatment group, gender, and overall, for the Safety Population. It is calculated as (Last Assessment Date in the Study – First Assessment Date in the Study) + 1.

A listing of study visits will be provided for the Safety Population.

6.0 Treatment Exposure and Administration

Treatment exposure of HArmonyCa Lidocaine, JUVÉDERM VOLITE, JUVÉDERM VOLBELLA, JUVÉDERM VOLIFT, JUVÉDERM VOLUMA, JUVÉDERM VOLUX will be summarized by visit, point of injection (region/MD code), side, gender, and overall, for the Safety Population. The summaries will include the number of injections and total volume of injections (mL), and average volume of injections (mL).

The average volume of injections (mL) is calculated as total volume of injections (mL) / number of injections.

Treatment exposure of BOTOX will be summarized by visit, point of injection (MD code), side, number of injections, gender, and overall, for the Safety Population.

The investigator is responsible for ensuring that study data is properly recorded in compliance with the protocol at the investigational site.

All treatment exposure data will be listed for the Safety Population.

7.0 Demographics, Baseline Characteristics, Medical History, and Prior/Concomitant Medications/Procedures

7.1 Demographics and Baseline Characteristics

Demographic parameters will be summarized descriptively by treatment group, gender, and overall, for the mITT Population. Demographic parameters include age (years) and sex (if female, childbearing potential).

Baseline characteristics will be summarized descriptively by treatment group, gender, and overall, for the mITT Population. Baseline parameters include height (m), weight (kg), body mass index (BMI) (kg/m^2), and Fitzpatrick Skin Type (I, II, III, IV, V, VI).

Demographics and baseline characteristics will be provided in a listing for the Safety Population.

Facial wrinkle scale (FWS) will be collected for glabellar lines (GLs), crow's feet lines (CFLs), and forehead lines (FHLs) at Screening visit. The FWS will be summarized by the severity of each question, by treatment group, gender, and overall, for the Safety Population. FWS data will be provided in a listing for the Safety Population.

7.2 Medical, Surgical and Cosmetic Procedure History

Medical, surgical, and cosmetic procedural history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA will be noted in the statistical tables/listings and clinical study report.

All medical, surgical, and cosmetic procedural history data will be listed for the Safety Population.

7.3 Prior and Concomitant Medications

A prior medication is defined as any medication started before the date of first study treatment.

A concomitant medication is defined as any medication taken on or after the date of first treatment.

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO DD). The actual version of the WHO DD will be noted in the statistical tables/listings and clinical study report.

All prior and concomitant medications will be listed for the Safety Population.

7.4 Concomitant Procedures

All concomitant procedures will be listed for the Safety Population.

8.0 Efficacy Analyses

8.1 General Considerations

- All efficacy analyses will be conducted in the Evaluable Population.
- Day 1 is defined as the day on which the study treatment is first received.
- Baseline for efficacy is defined as the last non-missing assessment on or prior to Day 1.
- Study day is relative to Day 1.
- Change from baseline will be computed as post-baseline minus baseline.

- Categorical variables will be summarized descriptively using number of subjects with observed values or events (n) and the percentage of subjects with observed values or events.

- Continuous variables will be summarized descriptively using number of subjects with observed values (n), mean, standard deviation (SD), standard error of mean (SEM), median, minimum (min), and maximum (max).

- Unless otherwise specified, the estimated mean and median for a set of values will be printed out to 1 more significant digit than the original values; the SD and SEM will be printed out to 2 more significant digits than the original values; the min and max will report the same significant digits as the original values.

- Shapiro-Wilks test will be used to check the normality assumption. If the normality assumption is met, the statistical hypothesis test will be performed based on one-sample, one-sided t-test at 0.025 significance level. If the normality assumption is not met, Wilcoxon signed-rank test will be used instead and a 97.5% confidence interval (CI) for the median will be presented using the Hodges Lehmann method. Shapiro-Wilks and Wilcoxon signed-rank test will be performed by proc univariate procedure, and Hodges Lehmann CI of the median for paired samples will be performed based on a SAS macro by Han (2008) [1] as specified in SPP Section 3.3.

- All CIs will be 1-sided 97.5% CIs, and will be printed to 2 decimal places, unless stated otherwise. P-values output will be in the format: '0.xxxx', where xxxx is the value rounded to 4 decimal places. Every p-value less than 0.0001 will be presented as <0.0001. If the p-value is returned as >0.9999, then present it as it is.

- Unless otherwise specified, the summaries will be provided by treatment group, gender, and overall.

8.2 Handling of Missing Data

For all FACE-Q™ questionnaires, the sum of the items will be calculated and converted into an overall score (also known as the Rasch-transformed score) ranging from 0 (worst) to 100 (best). If the number of missing items (which includes selection of more than 1 response for an item, a response of non-applicable (N/A) or skipping a question) is at most 50% of all the items, then insert the unrounded mean of the completed items into the total sum score. Inversely, if the

number of missing items is more than 50%, then the overall score will be missing. The total score will not be calculated in this case and the particular visit will be excluded from the Rasch-transformed score calculation and summary.

For PAAQ questionnaire, all items will be transformed to a scale from 0 (no negative impact) to 100 (high negative impact). Domain scores will be calculated as the average of transformed scale scores using the 50% rule given below:

If the domain has at least 50% non-missing item responses it will be scored as the average of the non-missing responses (with reverse scoring and rescaling as appropriate). If the domain is missing more than 50% of item responses, then the domain score will be set to missing.

For all other questionnaires as specified below, no imputation will be performed.

- Self-Perception of Age Questionnaire
- Global Aesthetic Improvement Scale

8.3 Primary Efficacy Endpoint and Analyses

8.3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline in the overall Rasch-transformed score of the FACE-Q^{TM}: Satisfaction with Facial Appearance Scale to the final study visit (V7).

8.3.2 Handling of Missing Data for the Primary Efficacy Endpoint

For the FACE-Q^{TM}: Satisfaction with Facial Appearance questionnaire, if the number of missing items (which includes selection of more than 1 response for an item, a response of non-applicable (N/A) or skipping a question) is at most 50% of all the items (i.e., at least 5 items have been answered), then insert the unrounded mean of the completed items into the total sum score. Inversely, if the number of missing items is more than 50% (i.e., less than 5 items have been answered), then the overall score will be missing. The total score will not be calculated in this case and the particular visit will be excluded from the Rasch-transformed score calculation and summary.

8.3.3 Primary Efficacy Analyses

The FACE-Q™: Satisfaction with Facial Appearance questionnaire will be given at the Screening visit and the final study visit (V7). Change from baseline in the overall Rasch-transformed score to the final study visit will be summarized descriptively by treatment group and overall and will be analyzed using a one-sample t-test (or Wilcoxon signed-rank test if the normality assumption is not met) for the Evaluable Population. See SPP Appendix C for the calculation of the overall Rasch-transformed score.

The number and percentage of subjects who provide responses to the FACE-Q™ Satisfaction with Facial Appearance questionnaire will also be summarized descriptively by treatment group and overall based on the following categories for each FACE-Q™ question:

- 1. Very Dissatisfied
- 2. Somewhat Dissatisfied
- 3. Somewhat Satisfied
- 4. Very Satisfied

A listing for FACE-Q™ Satisfaction with Facial Appearance will be provided for the Safety Population.

8.3.4 Additional Analyses of the Primary Efficacy Endpoints

Not applicable.

8.4 Secondary Efficacy Endpoints and Analyses

8.4.1 Secondary Efficacy Endpoints

The key secondary efficacy endpoints are:

- Subject's assessment of aging appraisal as measured by the change from baseline in the Rasch-transformed score of the FACE-Q™: Aging Appraisal to final study visit (V7)
- Subject's assessment of psychological function as measured by the change from baseline in the Rasch-transformed score of the FACE-Q™: Psychological Function Scale to final study visit (V7)

- Subject's assessment of social function as measured by the change from baseline in the Rasch-transformed score of the FACE-Q^{TM}: Social Function Scale to final study visit (V7)

- Subject's assessment of satisfaction with skin as measured by the change from baseline in the Rasch-transformed score of the FACE-Q^{TM}: Satisfaction with Skin to final study visit (V7)

- Subject's assessment of age-related facial appearance by the Self-Perception of Age (SPA) measure at baseline to final study visit (V7); responders are defined as having achieved a younger category.

Other secondary efficacy endpoints are:

- Investigator's assessment of global facial aesthetic improvement as measured by the 5-point GAIS at final study visit (V7)

- Subject's assessment of global facial aesthetic improvement as measured by the 5-point GAIS at final study visit (V7)

- Change in subject's satisfaction with appearance of periorbital area as measured by the change from baseline in the Periorbital Aesthetic Appearance Questionnaire (PAAQ) to final study visit (V7)

8.4.2 Secondary Efficacy Analyses

Change from baseline in the Rasch-transformed score of FACE-Q^{TM} questionnaires (FACE-Q^{TM}: Aging Appraisal, FACE-Q^{TM}: Psychological Function, FACE-Q^{TM}: Social Function, FACE-Q^{TM}: Satisfaction with Skin) will be summarized descriptively by treatment group, gender, and overall and will be analyzed using a one sample t-test (or Wilcoxon signed-rank test if normality assumptions are not met) for the Evaluable Population. See SPP Appendix D, Appendix E, Appendix F, and Appendix G for the calculation of the overall Rasch-transformed score for each FACE-Q^{TM} questionnaire.

The number and percentage of subjects who provide responses to each FACE-Q^{TM} questionnaire will also be summarized descriptively by treatment group, gender, and overall, for the Evaluable Population, based on the following categories for each questionnaire:

- FACE-Q™ Aging Appraisal
  - 1. Definitely Agree
  - 2. Somewhat Agree
  - 3. Somewhat Disagree
  - 4. Definitely Disagree
- FACE-Q™ Psychological Function, FACE-Q™ Social Function
  - 1. Definitely Disagree
  - 2. Somewhat Disagree
  - 3. Somewhat Agree
  - 4. Definitely Agree
- FACE-Q™ Satisfaction with Skin
  - 1. Very Dissatisfied
  - 2. Somewhat Dissatisfied
  - 3. Somewhat Satisfied
  - 4. Very Satisfied

For the Self-Perception of Age (SPA) questionnaire, the number and percentage of subjects who look their current age will be summarized for baseline and the final study visit by treatment group, gender, and overall, for the Evaluable Population. The change from baseline in the number of years look younger and older to the final study visit will be summarized descriptively by treatment group, gender, and overall, for the Evaluable Population.

A shift table to summarize the changes into an improvement category from the baseline to the final study visit will be provided in frequency and percentage. No statistical testing for the shift table will be performed.

There are two assessment categories in GAIS at the final study visit: subject rating and investigator rating. Descriptive summaries will be provided in frequency and percentage of the scales at the final study visit, in the 5-point scale from -2 to +2 by treatment group, gender, and overall, for the Evaluable Population. The scales include:

- 2: Much Improved
- 1: Improved
- 0: No Change
- -1: Worse
- -2: Much Worse

Descriptive summaries will be provided in frequency and percentage for nine questions of PAAQ at baseline and the final study visit by treatment group, gender, and overall, for the Evaluable Population, with five categories of answers for each question. Five categories of answer include:

- 0. Never
- 1. Rarely
- 2. Some of the Time
- 3. Most of the Time
- 4. All of the Time

The PAAQ will also be summarized in the transformed total score and three domains, using descriptive statistics by treatment group, gender, and overall and will be analyzed using a one-sample t-test (or Wilcoxon signed-rank test if normality assumption is not met) for the Evaluable Population.

The transformed scale score is calculated as [(actual raw score – lowest possible raw score) / possible raw score range] * 100. This formula can be simplified as [(actual raw score – 0) / 4] * 100.

The transformed total score is calculated as [(actual total raw score - lowest possible total raw score) / (possible total raw score range)] * 100. This formula can be simplified as [(actual total raw score - 0) / 36] *100.

Three domains include psychological domain, appearance domain, and coping domain. The psychological domain is defined as (sum of transformed scale score in PAAQ questions of 2, 4, 6, 7, and 8) / (number of questions). This formula can be simplified as (sum of transformed scale score in PAAQ questions of 2, 4, 6, 7 and 8) / 5.

Similar manner will be applied to the appearance domain and the coping domain, but with different number of PAAQ questions included for each domain. The appearance domain includes PAAQ questions of 1, 3 and 5. The coping domain contains only PAAQ question 9. See SPP Appendix H for the details of domain score calculation and missing data imputation.

Listings will be provided for FACE-Q^{TM}: Aging Appraisal, FACE-Q^{TM}: Psychological Function, FACE-Q^{TM}: Social Function, FACE-Q^{TM}: Satisfaction with Skin, SPA, GAIS, and PAAQ for the Safety Population.

A figure will be provided for the change from baseline of each FACE-Q^{TM} questionnaires (Satisfaction with Facial Appearance, Aging Appraisal, Psychological Function, Social Function, and Satisfaction with Skin) in the Rasch-transformed score by treatment group and gender, for the Evaluable Population.

8.5 Subgroup Analyses

A subgroup analysis on the primary efficacy endpoints (change from baseline in the overall Rasch-transformed score of the FACE-Q^{TM}: Satisfaction with Facial Appearance Scale to the final study visit) will be performed by gender and overall.

9.0 Safety Analyses

9.1 General Considerations

- All safety analyses will be conducted by treatment group, gender, and overall, for the Safety Population.

- Adverse events (AEs) will be coded using the medical dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical outputs and clinical study report.

9.2 Adverse Events

Adverse events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs).

An AE will be considered a JUVÉDERM injection related AE if it has a reasonable possibility (Possible, Probable, or Causal relationship) that the JUVÉDERM injections caused the event. The JUVÉDERM injections include JUVÉDERM VOLITE, JUVÉDERM VOLBELLA, JUVÉDERM VOLIFT, JUVÉDERM VOLUMA, and JUVÉDERM VOLUX.

An AE will be considered a BOTOX related AE if it has a reasonable possibility (Reasonable Possibility) that the BOTOX cosmetic caused the event.

An AE will be considered a study procedure related AE if it has a reasonable possibility (Possible, Probable, or Causal relationship) that the study procedure caused the event.

An AE will be considered a device related AE if it is related to a device malfunction/deficiency.

If the causal relationship to the study treatment/device/procedure is missing for an AE that occurred on or after the date of the first study treatment, then a causal relationship will be imputed for analysis. The imputed values for relationship to study treatment/device/procedure will be used for the incidence summary; the actual values will be presented in data listings.

If severity is missing for an AE that occurred before the date of first study treatment, then a severity of mild will be imputed for analysis. If the severity is missing for an AE that occurred on or after the date of the first study treatment, then a severity of severe will be imputed. The imputed values for severity will be used for the incidence summary; the actual values will be presented in data listings.

9.2.1 Treatment-Emergent Adverse Events

Treatment-emergent adverse event (TEAE) is defined as any AE with the onset on or after the date (and time, if known) of the first study treatment. A TEAE will be considered a treatment-

emergent serious adverse event (TESAE) if it is a TEAE that additionally meets any SAE criterion.

All AEs will be included in a listing. Only TEAEs will be included for summaries, unless otherwise specified. Subjects with any TEAEs, TESAEs, severe TEAEs, TEAEs related to study treatment, TEAEs related to study procedure, TEAES related to device, TEAEs leading to study discontinuation, TEAEs leading to death, TESAEs related to study treatment, TESAEs related to study procedure, TESAEs related to device, TESAEs leading to study discontinuation, and TESAEs leading to death will be summarized.

9.2.2 Adverse Event Overview

Overall summaries of TEAEs will include:

- The number and percentage of subjects with any TEAE
- The number and percentage of subjects with any TESAE
- The number and percentage of subjects with any severe TEAE
- The number and percentage of subjects with any TEAE related to JUVÉDERM VOLITE
- The number and percentage of subjects with any TEAE related to JUVÉDERM VOLBELLA
- The number and percentage of subjects with any TEAE related to JUVÉDERM VOLIFT
- The number and percentage of subjects with any TEAE related to JUVÉDERM VOLUMA
- The number and percentage of subjects with any TEAE related to JUVÉDERM VOLUX
- The number and percentage of subjects with any TEAE related to BOTOX cosmetic
- The number and percentage of subjects with any TEAE related to study procedure
- The number and percentage of subjects with any TEAE related to device
- The number and percentage of subjects with any TEAE leading to study discontinuation

- The number and percentage of subjects with any TEAE leading to death
- The number and percentage of subjects with any TESAE related to JUVÉDERM VOLITE
- The number and percentage of subjects with any TESAE related to JUVÉDERM VOLBELLA
- The number and percentage of subjects with any TESAE related to JUVÉDERM VOLIFT
- The number and percentage of subjects with any TESAE related to JUVÉDERM VOLUMA
- The number and percentage of subjects with any TESAE related to JUVÉDERM VOLUX
- The number and percentage of subjects with any TESAE related to BOTOX cosmetic
- The number and percentage of subjects with any TESAE related to study procedure
- The number and percentage of subjects with any TESAE related to device
- The number and percentage of subjects with any TESAE leading to study discontinuation
- The number and percentage of subjects with any TESAE leading to death
    - Before BOTOX Treatment
    - On or After BOTOX Treatment

9.2.1 Treatment-Emergent Adverse Events by SOC and PT, and/or Severity

The number and percentage of subjects will be summarized by SOC and PT, and by maximum severity, SOC, and PT as follow:

- TEAEs by SOC and PT
- TEAEs by maximum severity, SOC, and PT

- TEAEs related to study treatment (JUVÉDERM VOLITE, JUVÉDERM VOLBELLA, JUVÉDERM VOLIFT, JUVÉDERM VOLUMA, JUVÉDERM VOLUX, and BOTOX cosmetic) by SOC and PT
- TEAEs related to study treatment (JUVÉDERM VOLITE, JUVÉDERM VOLBELLA, JUVÉDERM VOLIFT, JUVÉDERM VOLUMA, JUVÉDERM VOLUX, and BOTOX cosmetic) by maximum severity, SOC, and PT
- TEAEs related to study procedure by SOC and PT
- TEAEs related to study procedure by maximum severity, SOC, and PT
- TEAEs related to device by SOC and PT
- TEAEs related to device by maximum severity, SOC, and PT
- TEAEs leading to study discontinuation by SOC and PT
- TEAEs leading to death by SOC and PT

The TEAE summary tables will be presented with SOC in alphabetical order and in decreasing frequency within each SOC. If more than 1 AE is coded to the same SOC or PT for the same subject, the subject will be counted only once for that SOC or PT using the most severe occurrence for the summarization by severity.

In addition, the duration of TEAEs will be summarized by SOC and PT. The duration of a TEAE will be calculated as (AE End Date – AE Start Date) + 1, and summarized as:

- 1 – 3 Days
- 4 – 7 Days
- 8 – 14 Days
- 15 – 30 Days
- >= 30 Days

The duration summary table will be presented with SOC in alphabetical order and in decreasing frequency within each SOC. If more than 1 AE is coded to the same SOC or PT for the same subject, the subject will be counted only once for that SOC or PT under the maximum duration.

9.2.2 SAEs (Including Deaths) and Adverse Events Leading to Study Discontinuation

An SAE is defined as any AE that:

- led to death
- led to serious deterioration in the health of the subject, that either resulted in
    - a life-threatening illness or injury; or
    - a permanent impairment of a body structure or a body function; or
    - in-patient hospitalization or prolongation; or
    - medical or surgical intervention to prevent permanent impairment to body structure or a body function; or
- led to fetal distress, fetal death or a congenital abnormality or birth defect.

TESAEs, TESAEs related to study drug, TESAEs related to study procedure, TESAEs related to device, TESAEs leading to study discontinuation, and TESAEs leading to death will be summarized in the overall summary table, and by SOC and PT as follow:

- TESAE by SOC and PT
- TESAEs related to study treatment (JUVÉDERM VOLITE, JUVÉDERM VOLBELLA, JUVÉDERM VOLIFT, JUVÉDERM VOLUMA, JUVÉDERM VOLUX, and BOTOX cosmetic) by SOC and PT
- TESAEs related to study procedure by SOC and PT
- TESAEs related to device by SOC and PT
- TESAEs leading to study discontinuation by SOC and PT

- TESAEs leading to death by SOC and PT

9.2.3 Adverse Events of Special Interest

Not applicable.

9.3 Analysis of Laboratory Data

Not Applicable.

9.4 Analysis of Vital Signs

Not Applicable.

9.5 Injection Site Response

Subjects will maintain a diary record of the presence, location, frequency, severity (mild, moderate, and severe), and duration of any injection site response (ISR) for 30 days after each filler treatment.

At each visit, the duration of each IRS symptom (redness, pain after injection, tenderness to touch, firmness, swelling, lumps/bumps, bruising, itching, discoloration, and other) will be calculated as Last Occurrence Date in eDiary – First Occurrence Date in eDiary + 1. It will be summarized descriptively using number and percentage of subjects with observed symptom, mean, SD, SEM, median, min, and max, and by categories 1-3 days, 4-7 days, 8-14 days, and 15-30 days.

The number and percentage of subjects with any ISR will be summarized by visit, ISR symptom and maximum severity.

A listing of ISRs will be provided for the Safety Population.

9.6 Procedural Pain

Not Applicable.

9.7 Safety Subgroup Analyses

Not Applicable.

9.8 Other Safety Analyses

The numbers and percentage of subjects with device deficiencies will be summarized by suspect device type.

A listing will be provided for device deficiencies for the Safety Population.

10.0 Other Analyses

Not applicable.

11.0 Interim Analyses

Not applicable.

12.0 Overall Type-I Error Control

Not applicable.

13.0 SAP Version History Summary

SAP Version	Date	Summary
0.1	10 May 2023	First draft
0.2	28 June 2023	Second draft
1.0	05 September 2023	Final version
1.1	28 March 2024	Revision to modify the summaries for treatment exposure and ISR based on the data collected.
2.0	03 April 2024	Revised final version
2.1	04 June 2024	Revision with modifications below: 1. Remove Appendix D through Appendix I. 2. Add Hodges Lehmann method for the confidence interval for the median of paired samples. 3. Modify the missing data imputation for FACE-Q and PAAQ questionnaires.
3.0	05 June 2024	Revised final version

14.0 Reference

[1] Lingling Han, Merck & Co., Inc., North Wales, PA. Calculating the point estimate and confidence interval of Hodges-Lehmann's median using SAS® software.Paper ST-154, SESUG Proceedings. Available from http://analytics.ncsu.edu/sesug/2008/ST-154.pdf (access on May 28, 2024).

Appendix A. Protocol Deviations and Violations

A protocol deviation (PD) is any non-adherence to study procedures that does not result in additional risk to the subject. Protocol deviations will be reviewed and documented before database lock.

A protocol violation (PV) is any non-adherence to the protocol that may result in significant additional risk to the subject (e.g., enrollment of a subject who does not meet the study criteria). A protocol violation can also be an event of non-adherence to Good Clinical Practice (GCP) that may impact subject safety (e.g., failure to obtain proper consent before performing study procedures). Violations should be reported to the study sponsor and the Institutional Review Board (IRB) within 5 working days if they occur.

Significant PDs/PVs (equivalent to major PDs/PVs as collected in data) will be identified before database lock. Significant PDs/PVs which impact the study conduct, data validity or integrity for the study outcomes will be summarized by the following categories:

- Inclusion or exclusion criteria
- Withdrawal Criteria
- Treatment
- Concomitant Medications
- Timeframe Deviation

The number and percentage of subjects with significant PDs/PVs will be summarized for the Safety Population by gender and total, and by treatment group and overall, for the Safety Population.

All PDs/PVs will be listed for the Safety Population.

Appendix B. List of Abbreviations

Abbreviation	Definition
AE	Adverse Event
BMI	Body Mass Index
CFL	Crow's Feet Line
CI	Confidence Interval
DD	Drug Dictionary
eCRF	Electronic Case Report Form
FHL	Forehead Line
FWS	Facial Wrinkle Scale
GAIS	Global Aesthetic Improvement Scale
GCP	Good Clinical Practice
GL	Glabellar Line
ICF	Informed Consent Form
IRB	Institutional Review Board
ISR	Injection Site Response
max	Maximum
min	Minimum
mITT	Modified Intent-to-treat
mL	Milliliter
MedDRA	Medical Dictionary for Regulatory Activities
PAAQ	Periorbital Aesthetic Appearance Questionnaire
PD	Protocol Deviation
PT	Preferred Term
PV	Protocol Violation
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SEM	Standard Error of Mean
SOA	Schedule of Activities
SOC	System Organ Class
SPA	Self-Perception of Age
SPP	Statistical Programming Plan

Abbreviation	Definition
TEAE	Treatment-emergent Adverse Event
TESAE	Treatment-emergent Serious Adverse Event
WHO	World Health Organization

Appendix C. Changes to Protocol-planned Analyses

Protocol Planned Analysis	Modified Analysis
All efficacy analyses will be based on the mITT and Evaluable Populations.	All efficacy analyses will be based on the Evaluable Populations.
Subgroup analysis by gender and treatment touch-ups.	Subgroup analysis by gender only.
mITT (modified Intent-to-Treat) Population - all subjects who enrolled and had at least one post-baseline efficacy assessment.	mITT (modified Intent-to-Treat) Population - all subjects who enrolled and had at least one post-baseline efficacy assessment at V7.
Evaluable Population– all subjects as defined in the mITT Population who have received both study products as treatments in this study, had at efficacy assessment at the final visit (V7) and no major protocol violations or deviations.	Evaluable Population– all subjects as defined in the mITT Population who have received both study products as treatments in this study, had at efficacy assessment at the final visit (V7) and no major protocol violations or deviations that will affect any efficacy endpoint.
Exploratory efficacy analysis may be performed on the collected video and photographs.	Removed.