Statistical Analysis Plan for Study 1878-701-008

A multicenter, evaluator-blinded, randomized, no-treatment controlled study to evaluate the safety and effectiveness of JUVÉDERM® VOLUMA® with Lidocaine for correction of temple hollowing in Chinese population

Date: 05 June 2023

Version 2.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) provides a technical and detailed elaboration of the statistical analyses of the effectiveness and safety data as outlined and specified in the final protocol of Study 1878-701-008 (Amendment 3, 18 January 2022). Specifications of tables, figures and data listings are contained in a separate document.

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

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

2.0 Study Design and Objectives

2.1 Study Objectives

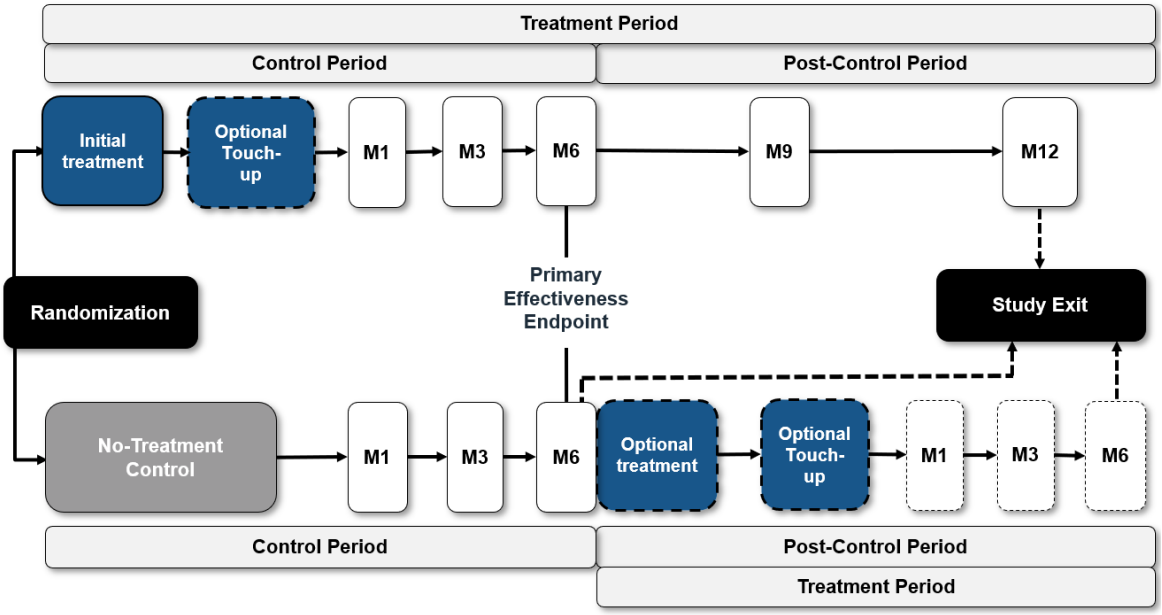
The objectives of this study are to evaluate the safety and effectiveness of VOLUMA with Lidocaine injectable gel in adult participants in Chinese population seeking correction of temple hollowing.

2.2 Study Design Overview

This is a prospective, multicenter, evaluator-blinded, randomized, parallel-group, no-treatment controlled study to evaluate the safety and effectiveness of VOLUMA with Lidocaine HA injectable gel to correct temple hollowing. See Protocol Section 4.1 Overall Design for details.

The schematic of the study is shown in Figure 1.

Figure 1. Study Schematic



2.3 Treatment Assignment and Blinding

Subjects will be randomized to VOLUMA with Lidocaine group (also referred as VOLUMA group in this document) or control group in a 2:1 ratio.

2.4 Sample Size Determination

The primary effectiveness measure is the change from baseline on the 5-point ATHS at Month 6. The primary effectiveness variable is the ATHS responder status, which is determined by an improvement of 1 grade or more (i.e., a change score of -4 to -1) in both temples.

Under a 2:1 randomization ratio, a total of 90 participants in the treatment group and 45 participants in the control group at the primary timepoint will provide at least 90% power to detect a difference of 30% in the responder rates between the groups. This calculation is based on a 2-sided Fisher's exact test at the 5% level, using responder rates

of 65% and 35% for both temples simultaneously for participants in the treatment and control groups, respectively. The assumption of a 65% treatment responder rate is a conservative estimate based on the observed responder rate in previous filler studies for other facial areas. Also, while a 30% difference from control is greater than a minimally important difference, a 35% responder rate is a reasonable upper limit of expectation for the control group, based on validation results of the ATHS and control group responder rates observed in previous filler studies. Furthermore, assuming a dropout rate of 20% after randomization through the Month 6 visit, 168 randomized participants are needed, with 112 participants in the treatment group and 56 participants in the control group (2:1 ratio).

With 7 investigational sites planned, it is recommended that no site randomize more than approximately 42 participants (i.e., twice the proportional share of 1 more site = 2 × [168/8]). The recommended minimum is 8 participants per site, which will guarantee that each treatment group has at least 2 participants at each site (allowing for mixed randomization blocks of 3 and 6 that incorporate the 2:1 randomization ratio).

The commercial software PASS (2008, Version 8.0.13) was used for the power calculation. The sample size calculation used an inequality test for 2 proportions to demonstrate that the treatment group is superior to the control group.

3.0 Endpoints

3.1 Primary Endpoint(s)

The primary endpoint is ATHS responder status based on the EI's live assessment at Month 6 during CP where a "responder" is a participant with at least 1-grade improvement on the ATHS in both temples.

3.2 Secondary Endpoint(s)

Secondary endpoints are:

- Responder status for EI assessment of GAIS in the temple area at Month 6 during CP
- Responder status for participant assessment of GAIS in the temple area at Month 6 during CP
- Change from baseline in participant responses to FACE-Q Satisfaction with Facial Appearance questionnaire, for the Rasch-transformed score (0-100) at Month 6
- Change from baseline in participant responses to FACE-Q Satisfaction with Temples questionnaire, for the Rasch-transformed score (0-100) at Month 6 during CP

3.3 Other Effectiveness Endpoint(s)

The additional effectiveness endpoints are for each visit unless specified otherwise.

- Primary and secondary endpoints at other visits
- Full-scale ATHS scores
- Full-scale EI assessment of GAIS
- Full-scale participant assessment of GAIS
- Responder status on the FACE-Q Satisfaction with Facial Appearance questionnaire, for the Rasch-transformed score, which will be summarized as a percentage of participants, where a responder is defined as a participant with an improved score from baseline
- FACE-Q Satisfaction with Facial Appearance questionnaire, for the Rasch-transformed score (0-100) (i.e., raw score)
- Responder status on individual items, A (How symmetric your face looks?), F (How rested your face looks?), J (How your face looks when first wake up?) of the FACE-Q Satisfaction with Facial Appearance questionnaire, which will be

summarized as a percentage of responders, where a responder is defined as a participant who responds as Somewhat Satisfied or Very Satisfied

- Responder status on the FACE-Q Satisfaction with Temples questionnaire, for the Rasch-transformed score, which will be summarized as a percentage of participants, where a responder is defined as a participant with an improved score from baseline
- FACE-Q Satisfaction with Temples questionnaire, for the Rasch-transformed score (0-100) (i.e., raw score)
- Responder status on individual items, 2 (how full your temples look?), 4 (how your temples look in a mirror (straight on)?), 11 (how well the shape of your temples compliments the shape of your face?), 12 (the age your temples make you look?), of the FACE-Q Satisfaction with Temples questionnaires, which will be summarized as a percentage of responders, where a responder is defined as a participant who responds as Somewhat Satisfied or Very Satisfied
- Participant's satisfaction with treatment questions and age perception question
- Volume change of each temple (left temple, right temple) from baseline as assessed by 3D imaging

3.4 Safety Endpoint(s)

The safety endpoints will include procedural pain, ISRs, AEs, jaw function assessment, vital signs, concomitant medications and procedures, and clinical laboratory values.

3.5 Additional Endpoint(s)

Not applicable.

4.0 Analysis Populations

The analysis populations will consist of participants as defined in Table 1 below:

Table 1. Analysis Sets

Population	Definition	Study Intervention
Modified Intent -to Treat (mITT)	All randomized participants who have non-missing baseline ATHS assessment for both temples.	As randomized
Observed Primary Endpoint (OPE)	All mITT population participants who have non-missing Month 6 ATHS assessments for both temples.	As randomized
Per-Protocol (PP)	All mITT population participants who have non-missing month 6 ATHS assessments for both temples and do not have any significant protocol deviations affecting the primary effectiveness endpoint. (i.e., a significant deviation that occurs after the primary endpoint does not lead to exclusion from the PP population.)	As randomized
Safety	All participants who are randomized and received study intervention (VOLUMA or no-treatment control).	As treated
VOLUMA Initially Treated (VIT)	All safety population participants who were randomized and received VOLUMA treatment at the beginning of the Control Period.	As treated
VOLUMA Optionally Treated (VOT)	All safety population participants who were randomized to control group and received VOLUMA optional treatment after the Control Period.	As treated

The number of participants in each of the 6 study populations (mITT, OPE, PP, Safety, VIT and VOT) will be summarized by treatment group, if applicable.

5.0 Participant Disposition

5.1 Study Period

The study periods are defined below in Table 2 for VOLUMA group and control group. For VOLUMA group, the Treatment Period (TP) is the Control Period (CP) plus the Post-control Period (PCP). For participants in the control group, the TP starts on the date of delayed treatment and ends on the day of the study exit.

Table 2. Definition of the Control and Treatment Periods by Analysis Group (as-randomized for effectiveness or as-treated for safety)

Analysis Group	Period	Analysis Group Label	Start Date	End Date
VOLUMA	CP	VOLUMA	Initial treatment date	The end date is the date of the Month 6 assessment: • If multiple assessments fall into Month 6 window, the date of the assessment included in the Month 6 analysis will be used as end date. • For participants who exit before Month 6, the end date is study exit date. • For participants with a missing assessment at Month 6, the end date is the target day for Month 6 (180 days from preceding treatment).
	TP	VOLUMA	Initial treatment date	Study exit date
Control	CP	Control	Randomization date	For participants who do not receive optional treatment: • If multiple assessments fall into Month 6 window, the date of the assessment included in the Month 6 analysis will be used as end date. • For participants who exit before Month 6, the end date is study exit date. • For participants with a missing assessment at Month 6, the end date is the target day for Month 6 (180 days from preceding treatment). For participants who receive optional treatment, the end date is the date when the treatment is received.
	TP	VOLUMA Post-Control	Optional initial treatment date	Study exit date

For CP, effectiveness analyses and baseline characteristics will be performed on the mITT population using the "as-randomized" assignment.

For TP, effectiveness and safety summaries will be performed on the VIT and VOT population for each treatment group.

5.2 Participant Disposition

The summary of study disposition will be done for all screened participants overall and by treatment group as randomized. Number and percentages are based on the numbers of participants in their randomized allocations. The following will be provided:

- Number of participants screened (overall)
- Number of participants randomized (this number will be used as denominator to compute the following percentages)
- Number of participants treated as randomized
- Number of participants not treated as randomized

For mITT population during CP:

- Number of participants completed CP
- Number of participants discontinued during CP
- Reasons for discontinuation during CP

For VIT and VOT population during PCP:

- Number of participants who continued into PCP
- Number of participants who completed PCP
- Number of participants who discontinued during PCP
- Reasons for discontinuation during PCP

For mITT population during the study:

- Number of participants who completed the study
- Number of participants who discontinued from the study
- Reasons for discontinuation from the study

Treatment group participants are considered to have completed the study if they complete the Month 12 visit in the Post-Control Period. Control group participants who elect not to receive optional treatment are considered to have completed the study if they complete the Month 6 visit in the Control Period. Meanwhile, control group participants who elect to receive optional treatment are considered to have completed the study if they complete the Month 6 visit in the Post-Control Period.

In addition, a listing will be provided for participant disposition.

6.0 Study Device Exposure and Administration

Treatment exposure-related variables will be summarized for the VIT and VOT populations. For the VOT population, data after participants received the initial treatment at Month 6 are included. Study treatment exposure will be measured by volume injected at each treatment (Table 3) and summarized by treatment group, treatment (initial, touch-up, initial and touch-up combined), and treatment area.

Table 3. Exposure to Study Treatment

Endpoint	Description	Timing	Methodology
Number (%) of participants received treatment	Summary by treatment group, treatment (initial, touch-up)	Initial, Touch-up	Frequency counts and percentages
Injection volume • Total • Left temple • Right temple	Summary by treatment group, treatment (initial, touch-up, initial and touch-up combined), and treatment area	Initial, Touch-up, Initial and Touch-up combined	Continuous descriptive statistics

The number of participants who received treatment anesthesia will be summarized for the VIT and VOT populations. Variables related to administration of treatment listed in Table 4 will be summarized for the VIT and VOT populations by treatment group, treatment area (right and left temple area), and treatment (initial, touch-up).

Table 4. Administration of Study Treatment

Endpoint	Description	Timing	Methodology
Pre-treatment anesthesia type • Ice • Topical • Local injection	Summary by treatment group and treatment (initial, touch-up)	Initial, Touch-up	Frequency counts and percentages
Pre-treatment anesthesia duration (minutes) for each pre-treatment anesthesia type • Anesthesia duration is computed as the earlier injection time of the two temples minus start of anesthesia administration time. Summarize by treatment group, treatment, and anesthesia type.	Summary by treatment group and treatment (initial, touch-up)	Initial, Touch-up	Continuous descriptive statistics
Treatment administration • Needle used[a] (27 Gauge 1/2 inch, 25 Gauge 1 inch, other) • Number of injections per temple • Planes of injection[a] (supraperiosteal, other) • Injection technique[a] (bolus, other) • Total number of syringes used[b] • Device/needle problem or malfunction[b]	Summary by treatment group, treatment (initial, touch-up) and treatment area (left temple, right temple)	Initial, Touch-up	Frequency counts and percentages, continuous descriptive statistics
Treatment characteristics • Injection ease • Product moldability	Summary by treatment group and treatment (initial, touch-up)	Initial, Touch-up	Frequency counts and percentages, continuous descriptive statistics

a. A participant may be counted in multiple planes of injection, injection techniques, and types of needles used.

b. Total number of syringes used and device/needle problem or malfunction will not summarized by treatment area.

A listing will be provided for extent of exposure and treatment administration.

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

7.1 Demographics and Baseline Characteristics

Demographic parameters (age; age group; sex; race) will be summarized descriptively in total and by treatment group for the mITT populations. Age (years) is calculated relative to informed consent date.

Baseline characteristics shown below will be summarized descriptively in total and by treatment group for the mITT populations:

- Weight (kg), height (cm), and BMI (kg/m^2)
- Allergan Temple Hollowing Scale
- FACE-Q Satisfaction with Facial Appearance questionnaire, for the Rasch-transformed score
- FACE-Q Satisfaction with Temples questionnaire, for the Rasch-transformed score
- Jaw Functional Limitation Scale
- Self-perception of age

Demographic and baseline characteristics will be provided in a listing.

7.2 Medical History

Abnormalities in participants' medical, surgical, cosmetic, and dental history, encompassing abnormalities, surgeries, and procedures reported as occurring before the Screening visit, 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.

Listing will be provided for the safety population.

7.3 Prior and Concomitant Medications

The medication data will be coded using the World Health Organization (WHO) Drug Dictionary. The actual version of the WHO Drug Dictionary will be noted in the statistical outputs and clinical study report. Prior medication is defined as any medication taken prior to the latter of randomization date or first treatment date in the control period. Concomitant medication is defined as any medication taken on or after the latter of randomization date or first treatment date in the control period.

Listing will be provided for the safety population. The Anatomical Therapeutic Chemical (ATC) class, and preferred drug name will be presented (i.e., 4th level, or most specific level available if 4th level is unavailable).

8.0 Effectiveness Analyses

8.1 General Considerations

- For CP, effectiveness analyses and baseline characteristics will be performed on the mITT population. Subjects will be included in the analysis according to the treatment groups that they are randomized to.
- For TP, effectiveness analysis will be performed on the VIT and VOT populations. Subjects will be included in the analysis according to the study treatment that they actually received.
- The baseline for effectiveness parameters will be the last non-missing assessment prior to or on Day 1, i.e., the latter of randomization or initial injection in Control Period.
- Day 1 is defined as the day on which the study intervention (VOLUMA or no-treatment control) is first received.
- Study day refers to Day 1 regardless of study period (e.g., CP, TP).
- The change from baseline values will be computed as the postbaseline value minus the baseline values.
- Continuous variables will be summarized by number of participants with observed values (n), mean, standard deviation (SD), median, 1st and 3rd quartiles (Q1, Q3), minimum (min), and maximum (max). Categorical

variables will be summarized by the number of participants with observed values or events (n) and the percentage of participants with observed values or events.

- All statistical hypothesis tests will be performed at the 2-sided, 5% significance level, unless stated otherwise.
- All CIs will be 2sided 95% CIs, unless stated otherwise.
- All statistical analyses will be performed using SAS Version 9.4 or subsequent.

8.2 Handling of Missing Data

Missing data will be imputed using the following methods for the effectiveness analyses:

- Multiple Imputation (MI): If there are any mITT participants with missing ATHS score at Month 6 on either side of temple, MI by the Fully Conditional Specification (FCS) method will be used to impute the missing data in ordinal scale by treatment group and by side of temple (left and right). Details of the MI method for primary analysis are provided in Section 8.3.2.
- Non-Responder Imputation (NRI): For sensitivity analysis, the NRI method will categorize any participants with missing Month 6 ATHS scores as non-responders.
- As Observed (AO): The AO method will analyze the data based on the observed values without imputation for the missing values. Thus, a participant who does not have an evaluation on a scheduled visit will be excluded from the AO analysis for that visit. The AO method will be used for sensitivity analysis of primary effectiveness endpoint using the OPE population.

8.3 Primary Effectiveness Endpoint(s) and Analyses

8.3.1 Primary Effectiveness Endpoint(s)

The primary effectiveness endpoint is the ATHS responder status at Month 6 in CP based on EI's assessment. A responder is defined as a participant with at least 1-grade

improvement (reduction) from baseline on the ATHS (described in Table 5) on both temples.

Table 5. Allergan Temple Hollowing Scale

Score	Grade	Description
0	Convex	Rounded temple
1	Flat	Flat temple; temporal fusion line may be visible
2	Minimal	Shallow depression or concavity with minimal volume loss; temporal fusion line may be visible
3	Moderate	Moderate depression or concavity with moderate volume loss; moderate prominence of temporal fusion line
4	Severe	Deeply recessed, sunken appearance; marked prominence of temporal fusion line and zygomatic arch

8.3.2 Handling of Missing Data for the Primary Effectiveness Endpoint(s)

If there are any mITT participants with missing scores for the primary effectiveness endpoint, the following missing-data handling method is planned.

Step 1: The missing data will be imputed by treatment group and by side of the temple using the FCS method for ordinal data (via SAS procedure MI with FCS logistic statement). Thirty (30) imputed datasets will be generated, using *seed = 1878701* and *nimpute = 30* based on the following model:

Month 6 ATHS = $\beta_0 + \beta_1 \times$ Baseline ATHS + $\beta_2 \times$ Month 1 ATHS + $\beta_3 \times$ Month 3 ATHS

Step 2: With the FCS logistic statement in the MI procedure, each imputed dataset will have ATHS scores of 0, 1, 2, 3, or 4. After imputation, the changes from baseline will be calculated, from which the responder status in ATHS at Month 6 of CP will be determined. Intermittent missing ATHS at Month 1 and Month 3 will be imputed by similar models using ATHS at previous visits as covariates.

Step 3: Each complete (imputed) dataset obtained from Steps 1 and 2 will be analyzed using SAS procedure FREQ.

Step 4: The pooled results will be obtained using PROC MIANALYZE.

The aggregated responder rates by treatment group and the associated p-value and 95% CI will be presented to assess the difference in the ATHS responder rate at Month 6 of CP between VOLUMA and the no-treatment control. If the 2-sided p-value is ≤ 0.05, then the ATHS responder rate at Month 6 in VOLUMA XC is statistically significantly different from that in the no-treatment control group.

8.3.3 Primary Effectiveness Analysis

Primary effectiveness analysis will be performed based on the mITT population. The study intervention will be determined to be clinically effective if the following criterion is met:

- The ATHS responder rate for the treatment group is statistically superior to the responder rate for the control group at Month 6

The null and alternate hypotheses are:

$$H_o: P_v = P_c \text{ versus } H_a: P_v \neq P_c$$

where P_v and P_c denote the responder rates for the treatment group at Month 6 after the latest of randomization, initial treatment and touch-up treatment, and control group at Month 6 after randomization, respectively.

8.3.4 Additional Analyses of the Primary Effectiveness Endpoint(s)

If there are any mITT participants with missing scores for the primary effectiveness endpoint, then a sensitivity analysis of the primary effectiveness endpoint will be performed for the OPE population using the Fisher's exact test for group comparison.

A second sensitivity analysis of the primary effectiveness endpoint will also be performed using Fisher's exact test for the mITT population with missing ATHS responder status at Month 6 imputed as non-responders.

The primary analysis using Fisher's exact test will also be performed for the PP population based on AO method for the missing data imputation as a third sensitivity analysis.

8.4 Secondary Effectiveness Analyses

There are 4 secondary effectiveness measures, to be recorded at the participant level rather than for each temple. The secondary effectiveness endpoints include:

- Responder status for EI assessment of GAIS (described in Table 6) in the temple area at Month 6 during CP
- Responder status for participant assessment of GAIS in the temple area at Month 6 during CP
- Change from baseline in participant responses to FACE-Q Satisfaction with Facial Appearance questionnaire, for the Rasch-transformed score (0 - 100) at Month 6 during CP
- Change from baseline in participant responses to FACE-Q Satisfaction with Temples questionnaire, for the Rasch-transformed score (0 - 100) at Month 6 during CP

A GAIS responder for EI assessment is defined as a participant who shows improvement in the overall aesthetic assessment (Improved or Much Improved on GAIS) based on EI assessment in the temple area at the Month 6 visit.

A GAIS responder for participant assessment is defined as a participant who shows improvement in the overall aesthetic assessment (Improved or Much Improved on GAIS) based on participant assessment in the temple area at the Month 6 visit.

For the responder status based on the EI's and participant's assessments of GAIS at Month 6, the responder rates with 95% exact CI will be summarized, and p-value for

between treatment group comparisons will be based on 2-sided Fisher's exact test using observed data.

Table 6. Global Aesthetic Improvement Scale

Grade	Description
Much Improved	Marked improvement in appearance
Improved	Improvement in appearance, but a touch-up or retreatment is indicated
No Change	The appearance is essentially the same as the original condition
Worse	The appearance is worse than the original condition
Much Worse	The appearance is much worse than the original condition

The other secondary effectiveness measures are the participant responses on the FACE-Q questionnaires: Satisfaction with Facial Appearance and Satisfaction with Temples.

Change from baseline on participant responses on FACE-Q Satisfaction with Facial Appearance and with Temples questionnaires, for the Rasch-transformed score (0-100) will be analyzed using a mixed-effects model for repeated measures, and the nominal p-value for the between treatment group comparison will be provided for Month 6.

For FACE-Q Satisfaction with Facial Appearance questionnaire and FACE-Q Satisfaction with Temples questionnaire, the responses to all the items will be summed and converted to a Rasch-transformed score that ranges from 0 to 100 (higher score indicates increased satisfaction) using the algorithm developed by the FACE-Q scale developers.[1]

For items with missing data (which includes selection of more than 1 response for an item, a response of non-applicable (N/A) or skipping a question), insert the unrounded mean of the completed items into the total sum score. The summed score including the imputation of missing items is rounded to the nearest integer and converted to the Rasch transformed score using the conversion table below. Higher scores reflect a better outcome. The converted score ranges from 0 (worst) to 100 (best). If less than 50% of the questions have been answered, then the overall score will be missing.

The secondary effectiveness endpoints will be analyzed based on mITT population based on the observed data.

8.5 Additional Effectiveness Analyses

The other effectiveness analyses will be performed for all variables listed in Table 7. Analyses on the primary and secondary variables at other visits will be performed. All analyses are based on observed data and missing data will not be imputed. All analyses are performed by treatment group and by visit. Additionally, no hypothesis will be tested and, thus, no multiplicity adjustment will be performed.

For CP, the mITT population will be used, including Month 1, 3 and 6 visits.

For TP of the VOLUMA group, the VIT population will be used, including Month 1, 3, 6, 9 and 12 visits.

For TP of the control group, the VOT population will be used, including Month 1, 3 and 6 visits.

In the responder analysis on FACE-Q Satisfaction with Facial Appearance and Satisfaction with Temples questionnaires, a responder is defined as any participant whose postbaseline Rasch-transformed score is greater than their baseline Rasch-transformed score.

For the descriptive summary of ATHS overall score, the temple with worse baseline score will be used to represent the participant; if both temples havethe same baseline score, the left side temple will be used to represent the participant.

For participant's self-perception of age in treatment group, change from baseline (e.g., -10 = "I look 10 years younger," 0 = "I look my age," and +6 = "I look 6 years older") will be calculated and summarized descriptively. The paired t-test nominal p-value will be calculated at Month 6 as descriptive guidance to consider if the change from baseline is different from zero for the primary timepoint of Month 6.

Table 7. Other Effectiveness Analyses

Endpoint	Description	Analysis Period	Methodology
ATHS responder status as assessed by EI	Number (%) of responders	CP, TP	Binomial descriptive statistics
Raw value and change from baseline in ATHS score as assessed by EI (based on the worse temple, or the left temple if both temples have the same baseline score)	Summary	CP, TP	Continuous descriptive statistics
GAIS responder status as assessed by EI	Number (%) of responders	CP, TP	Binomial descriptive statistics
Responses on GAIS as assessed by EI	Number (%) by response category	CP, TP	Frequency counts and percentages
GAIS responder status as assessed by participant	Number (%) of responders	CP, TP	Binomial descriptive statistics
Responses on GAIS as assessed by participant	Number (%) by response category	CP, TP	Frequency counts and percentages
Raw value and change from baseline in FACE-Q Satisfaction with Facial Appearance Rasch-transformed score as assessed by participant	Summary	CP, TP	Continuous descriptive statistics
Responder status in FACE-Q Satisfaction with Facial Appearance Rasch-transformed score as assessed by participant	Number (%) of responders (improved)	CP, TP	Binomial descriptive statistics
Responses on item A, F, J of FACE-Q Satisfaction with Facial Appearance	Number (%) by response category (including somewhat satisfied or very satisfied)	CP, TP	Binomial descriptive statistics
Responses on each item of FACE-Q Satisfaction with Facial Appearance	Number (%) by response category	CP, TP	Frequency counts and percentages

Table 7. Other Effectiveness Analyses (Continued)

Endpoint	Description	Analysis Period	Methodology
Raw value and change from baseline in FACE-Q Satisfaction with Temples Rasch-transformed score as assessed by participant	Summary	CP, TP	Continuous descriptive statistics
Responder status in FACE-Q Satisfaction with Temples Rasch-transformed score as assessed by participant	Number (%) of responders (improved)	CP, TP	Binomial descriptive statistics
Responses on item 2, 4, 11, 12 of FACE-Q Satisfaction with Temples	Number (%) by response category (including somewhat satisfied or very satisfied)	CP, TP	Binomial descriptive statistics
Responses on each item of FACE-Q Satisfaction with Temples	Number (%) by response category	CP, TP	Frequency counts and percentages
Participant responses to satisfaction with treatment on a 5-point scale	Number (%) by response category for treatment group only (including satisfied or very satisfied)	TP	Frequency counts and percentages
Participant responses to natural look of the treatment areas on a 5-point scale	Number (%) by response category for treatment group only (including satisfied or very satisfied)	TP	Frequency counts and percentages
Participant responses to natural feel of the treatment areas on a 5-point scale	Number (%) by response category for treatment group only (including satisfied or very satisfied)	TP	Frequency counts and percentages
Participant responses to treatment meeting expectation (yes/no)	Number (%) by response category for treatment group only	TP	Binomial descriptive statistics
Participant responses to interest to continue to use the treatment (yes/no)	Number (%) by response category for treatment group only	TP	Binomial descriptive statistics
Participant's willingness to recommend treatment (yes/no)	Number (%) by response category for treatment group only	TP	Binomial descriptive statistics
Volume change of each temple (left temple, right temple) from baseline as assessed by 3D imaging	Summary	CP, TP	Continuous descriptive statistics

Table 7. Other Effectiveness Analyses (Continued)

Endpoint	Description	Analysis Period	Methodology
Participant's self-perception of age	Change from baseline for treatment group; for control group at baseline only	CP, TP	Continuous descriptive statistics; with paired t-test within treatment group at Month 6

8.6 Effectiveness Subgroup Analyses

Not applicable.

9.0 Safety Analyses

9.1 General Considerations

- For CP, safety summaries will be performed on the safety population. Subjects will be included in the analysis according to the study treatment that they actually received.
- For TP, safety summaries will be performed on the VIT and VOT populations. Subjects will be included in the analysis according to the study treatment that they actually received.
- The baseline for safety parameters will be the last non-missing assessment prior to or on Day 1, i.e., the latter of randomization or initial injection in Control Period. Day 1 is defined as the day on which the study intervention (VOLUMA or no-treatment control) is first received.
- Study day refers to Day 1 regardless of study period (e.g., CP, TP).
- The change from baseline values will be computed as the postbaseline value minus the baseline values.
- 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 tables and clinical study report.

9.2 Adverse Events

9.2.1 Treatment-Emergent Adverse Events

An AE will be considered a treatment-emergent adverse event (TEAE) if the AE began or worsened (increased in severity or became serious) on or after the first administration of study intervention (i.e., on or after study Day 1 defined in Section 7.1).

A treatment-emergent serious adverse event (TESAE) is defined as a serious adverse event (SAE) that is also a TEAE.

9.2.2 Adverse Event Overview

TEAEs will be summarized by treatment group for CP using the safety population. TEAEs will also be summarized by treatment group for TP using the VIT and VOT populations, respectively as described in Table 8. The summaries will be presented using number and percentage of participants with TEAEs as well as the number of events. Additional summaries may be provided when applicable.

An overall summary of TEAEs will be presented as described in Table 8.

A TEAE will be considered a treatment-related TEAE if the TEAE is deemed as related to the study procedure or the study device by the TI. Treatment-related AEs will be summarized by duration, severity, time to onset on/after the most recent treatment, and outcome.

The duration of an AE is defined as the end date of AE minus the start date of AE plus 1.

If more than 1 AE is coded to the same preferred term for a participant, the participant will be counted only once for that preferred term using the most severe and most related occurrence for the severity and relationship to study intervention summaries, respectively.

If the severity of a TEAE is missing, the maximum severity will be assigned to the event for the summary on severity. The value will be displayed as missing in the data listing.

If the relationship to the study intervention is missing for a TEAE, the event will be considered related to the study intervention for the summary. The value will be displayed as missing in the data listing.

Listings of all AEs, TESAEs, TEAEs leading to discontinuation, and death will be presented.

Table 8. TEAE Summaries

Endpoint	Description	Timing	Methodology
Overall summary	Overall summary only for the following categories: • TEAEs • Treatment-related TEAEs • Treatment-related TEAEs at injection site • Treatment-related TEAEs not at injection site • TESAEs • Treatment-related TESAEs • Treatment-related TESAEs at injection site • Treatment-related TESAEs not at injection site • TESAEs leading to study discontinuation • Deaths	CP, TP	Frequency counts and percentages
TEAEs	• By SOC and PT • By SOC, PT, and maximum severity	CP, TP	Frequency counts and percentages
Treatment-related TEAEs	• By SOC and PT • By SOC, PT and maximum severity • By duration, time to onset from the most recent treatment, action taken and outcome and categorized by severity	CP, TP	Frequency counts and percentages
TESAEs	By SOC and PT	CP, TP	Frequency counts and percentages
AEs leading to study discontinuation	By SOC and PT	CP, TP	Frequency counts and percentages

9.2.3 Adverse Events of Special Interest

Not applicable.

9.3 Laboratory Data

Clinical laboratory values at baseline (screening) and changes from baseline at each assessment time point will be presented by data listings for each treatment group using Safety population.

9.4 Vital Signs

Vital signs (systolic and diastolic blood pressure, pulse rate, respiratory rate, and temperature) at baseline (screening) and changes from baseline at each assessment will be presented by data listings.

9.5 Height and Weight

Height and weight at Screening visit will be summarized as baseline characteristics. Weight will be listed.

9.6 Pregnancy Test

Participants with a positive result for the safety population will be listed.

9.7 Injection Site Response

ISRs are recorded in participant diaries after each treatment (initial and touch-up). Diary Day is derived as diary date - treatment date + 1 (e.g., the day of treatment is diary Day 1). ISR entry includes diary data and eCRF data on Ongoing ISR page. The maximum severity and ISR duration are based on diary data and eCRF data for the specified treatment (e.g., the maximum severity for touch-up treatment is based on diary data and eCRF data after touch-up treatment).

Number (%) of participants will be summarized for each treatment by predefined symptoms. The number of participants with ISR entries (diary data and eCRF data) for

each treatment is used as the denominator for the calculation of percentages. The ISR analyses are listed in Table 9.

Table 9. Injection Site Response Analyses

Endpoint	Description	Timing	Methodology
ISR severity	Number of participants with ISRs by symptom incidence and categorized by maximum severity reported on Diary and ongoing ISR eCRF	Initial, Touch-up	Frequency counts and percentages
ISR duration	Number of participants with ISRs by symptom incidence and categorized by duration. Duration is defined as number of days from first instance of the symptom to the last instance of the symptom reported on Diary or ongoing ISR eCRF (i.e., date of last ISR - date of first ISR + 1.)	Initial, Touch-up	Frequency counts and percentages, continuous descriptive statistics
Ongoing ISR	ISRs recorded on Ongoing ISR page		Listing

9.8 Procedural Pain

Participant assessment of procedural pain (pain during injection) on an 11-point scale ranging from 0 (no pain) to 10 (worst pain imaginable) after initial, touch-up be summarized descriptively.

9.9 Safety Subgroup Analyses

TEAEs and treatment-related TEAEs during CP and TP will be analyzed for all sites excluding site 11508 as well as site 11508 only. Details of the analyses are shown in Table 10.

Table 10. TEAE Subgroup Analyses

Endpoint	Description	Timing	Methodology
Overall summary	Overall summary only for the following categories: • TEAEs • Treatment-related TEAEs • Treatment-related TEAEs at injection site • Treatment-related TEAEs not at injection site • TESAEs • Treatment-related TESAEs • Treatment-related TESAEs at injection site • Treatment-related TESAEs not at injection site • TEAEs leading to study discontinuation • Deaths	CP, TP	Frequency counts and percentages
TEAEs	• By SOC and PT	CP, TP	Frequency counts and percentages
Treatment-related TEAEs	• By SOC and PT	CP, TP	Frequency counts and percentages

9.10 Other Safety Analyses

Other safety endpoints include global score of jaw functional and concomitant medication and concurrent procedures.

The other safety endpoints will be summarized by treatment groups, for VIT and VOT where applicable.

For Jaw Functional Limitation Scale, the 8-item form (JFLS-8) is used in the study. Raw and change from baseline values of global score (calculated as mean of the available items when no more than 2 items missing) will be summarized using descriptive statistics.

Concomitant medications (Section 7.3) and concurrent procedures (coded using MedDRA) will be listed.

10.0 Other Analyses

Not applicable.

11.0 Interim Analyses

Not applicable.

11.1 Data Monitoring Committee

Not applicable.

12.0 Overall Type-I Error Control

All statistical tests will be 2-sided hypothesis tests performed at the 5% level of significance for main effects. A gatekeeping procedure will be used for hypothesis testing of the primary endpoint (ATHS responder) and 2 of the secondary endpoints (responder rates for EI and participant assessments of GAIS) following a predefined sequence to control the overall type I error rate at the 0.05 level. The primary hypothesis must be rejected in order to test the responder rate for EI assessment of GAIS. Additionally, statistical significance must be established for the responder rate for EI assessment of GAIS in order to assess the hypothesis for the responder rate for participant assessment of GAIS.

13.0 Version History

SAP Version History Summary			
SAP Version	**Approval Date**	**Change**	**Rationale**
1	08 September 2020	Not Applicable	Original version
2	05 June 2023	Introduction	Section 1.0 added clarification of system information.
		Administration of Study Treatment	Section 6.0 added clarification footnote to Table 4 on treatment administration summary.
		Demographics and Baseline Characteristics	Section 7.1 removed ethnicity summary as this study will only include Chinese subjects.
		Handling of Missing Data for the Primary Effectiveness Endpoint(s)	Section 8.3.2 updated number of imputed datasets from 5 to 30, to improve robustness of multiple imputation results.
		Secondary Effectiveness Analyses	Section 8.4, added clarification for FACE-Q. Section 8.5, added clarification of analyses methods and description on additional effectiveness analyses.
		Adverse Events	Section 9.2 removed "maintenance treatment" related description as not applicable. Section 9.2.2, added clarification of treatment-related TEAE. Added clarification to Table 8 TEAE summaries.
		Injection Site Response	Section 9.7 removed "maintenance treatment" related description as not applicable and added clarification.

SAP Version History Summary			
SAP Version	**Approval Date**	**Change**	**Rationale**
		Safety subgroup analyses	Section 9.9 added safety subgroup analyses of TEAEs, to evaluate the impact of unusual AE reporting rule in Site 11508 (all ISRs were reported as AEs), subgroup analyses of AEs without this site and AEs within this site will be analyzed separately, to provide a holistic picture of AEs profile in this study.
		Appendix	Removed Appendix Scales and Scoring Algorithms and related descriptions in context to comply with current standard.

14.0 References

1. Pusic AL, Klassen AF, Scott AM, et al. Development and psychometric evaluation of the FACE-Q satisfaction with appearance scale: a new patient-reported outcome instrument for facial aesthetic patients. Clin Plast Surg. 2013;40(2):249-60.

Appendix A. Protocol Deviations

Significant protocol deviations will be identified. Unique participants reporting significant protocol deviations will be summarized in total and by treatment group for all randomized or treated participants as described in Table 11. If there are participants who are mistreated, tabulate participants as randomized.

A data listing of significant protocol deviations will be provided. Also, a listing of visits affected by COVID-19 will be included, even if there is no associated significant protocol deviation.

Table 11. Protocol Deviation Summary

Endpoint	Description	Timing	Methodology
Significant protocol deviations	Number (%) of participants with significant protocol deviation will be summarized (All randomized)	During study period	Frequency counts and percentages

Appendix B. Definition of Adverse Events of Special Interest

Not Applicable.

Appendix C. Changes to Protocol-planned Analysis

- Add Observed Primary Endpoint (OPE) population as "All mITT population participants who have non-missing Month 6 ATHS assessments for both temples," which will be used for additional sensitivity analysis of primary endpoint, to provide further information on robustness of results.
- Updated number of imputed datasets created from 5 to 30, on handling of missing data for primary effectiveness endpoint, to enhance robustness of multiple imputation results.

Appendix D. List of Abbreviations

Abbreviation	Definition
3D	3-dimensional
AE	adverse event
AO	as observed
ATHS	Allergan Temple Hollowing Scale
CI	confidence interval
CP	control period
EI	Evaluating Investigator
FCS	fully conditional specification
GAIS	Global Aesthetic Improvement Scale
ISR	injection site response
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
mITT	modified intent-to-treat
NRI	non-responder imputation
OPE	observed primary endpoint
PCP	post-control period
PI	Principal Investigator
PP	per-protocol
PT	prefer term
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SOC	system organ class
TEAE	treatment-emergent adverse event
TESAE	treatment-emergent serious adverse event
TI	Treating Investigator
TP	treatment period
VIT	VOLUMA initially treated
VOT	VOLUMA optionally treated
WHO	World Health Organization