Statistical Analysis Plan for Study 1878-702-008

A multicenter, evaluator-blinded, randomized, parallel-group, controlled study of the safety and effectiveness of JUVÉDERM VOLUMA® XC injectable gel for correction of temple hollowing

Date: 22 September 2021

Version 3.0

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 1878-702-008 (Amendment 2, 12 May 2021). Specifications of tables, figures and data listings are contained in a separate document.

2.0 Study Design and Objectives

2.1 Study Objectives

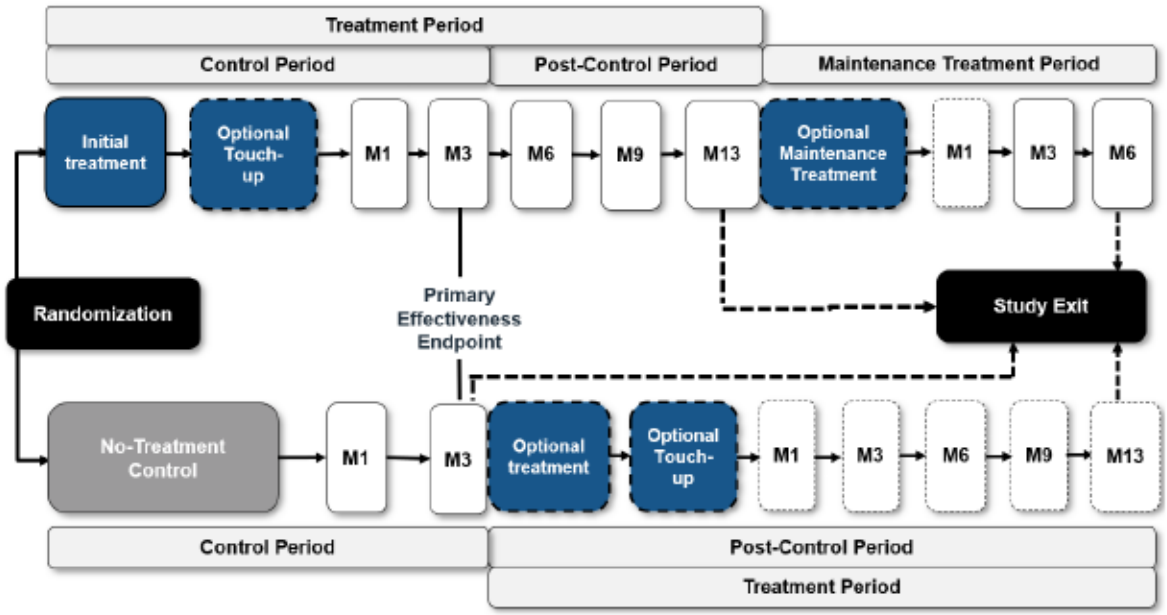
The objectives of this study are to evaluate the safety and effectiveness of VOLUMA XC injectable gel in adult participants seeking correction of temple hollowing.

2.2 Study Design Overview

This is a prospective, multicenter, evaluator-blinded, randomized, parallel-group, controlled study to evaluate the safety and effectiveness of VOLUMA XC 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 XC group or control group in a 2:1 ratio based on a [redacted]

[redacted]

[redacted]

[redacted]

3.0 Endpoints

3.1 Primary Endpoint(s)

Primary endpoint is ATHS responder status at Month 3.

3.2 Secondary Endpoint(s)

Secondary endpoints are:

- Responder status for EI assessment of GAIS in the temple area at Month 3
- Responder status for participant assessment of GAIS in the temple area at Month 3
- Change from baseline in participant responses to FACE-Q Satisfaction with Facial Appearance questionnaire, for the Rasch-transformed score (0-100) at Month 3
- Change from baseline in participant responses to FACE-Q Satisfaction with Temples questionnaire, for the Rasch-transformed score (0-100) at Month 3

3.3 Other Effectiveness Endpoint(s)

[redacted]

3.4 Safety Endpoint(s)

The safety endpoints will include procedural pain, ISRs, AEs, jaw function assessment, Snellen visual acuity, confrontational visual fields, ocular motility, and concomitant medications and procedures.

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
mITT	All randomized participants who have non-missing baseline assessment on the ATHS for both temples. This will be used as the full-analysis-set population.	As randomized
Observed Primary Endpoint (OPE)	All mITT population participants who have a Month 3 assessment on the ATHS for both temples.	As randomized
Safety	All participants who are randomized and received study intervention (VOLUMA XC or no-treatment control).	As treated
VOLUMA XC Treated (VT)	All safety population participants who received VOLUMA XC treatment as initial treatment at the beginning of the Control Period, or as optional treatment after the Control Period	As treated
VOLUMA XC Maintenance Treatment (VMT)	All safety population participants who received VOLUMA XC maintenance treatment	As treated

The number of participants in each of the 5 study populations (mITT, OPE, Safety, VT and VMT) 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 XC group and control group. For VOLUMA XC 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 optional treatment and ends on the same day as the study exit.

Table 2. Definition of the Control, Treatment and Maintenance 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 XC	Initial treatment date	The end date is the date of the Month 3 assessment: • If multiple assessments fall into Month 3 window, the date of the assessment included in the Month 3 analysis will be used as end date. • For participants who exit before Month 3, the end date is study exit date. • For participants with a missing assessment at Month 3, the end date is the target day for Month 3 (90 days from preceding treatment).
	TP	VOLUMA XC	Initial treatment date	Date of maintenance treatment or the study exit date, whichever is earlier
	MP	VOLUMA XC	Maintenance treatment date	Study exit date
Control	CP	Control	Randomization date	For participants who do not receive optional treatment: • If multiple assessments fall into Month 3 window, the date of the assessment included in the Month 3 analysis will be used as end date. • For participants who exit before Month 3, the end date is study exit date. • For participants with missing assessment at Month 3, the target day for Month 3 (Day 91) will be the end date. For participants who receive optional treatment, the end date is the date of optional treatment.
	TP	VOLUMA XC Post-Control	Optional initial treatment date	Study exit date

The start and end of treatment periods listed above for the mis-treated participants (ie.,

randomized to treatment and did not receive VOLUMA XC or randomized to control and was treated with VOLUMA XC prior to the primary timepoint) will be determined based on the treatment received.

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 VT population by treatment group.

For MP, effectiveness and safety summaries will be performed on the VMT population.

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 will be provided for the following:

- 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 VT 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 VMT population during MP:

- Number of participants who received maintenance treatment
- Number of participants who completed MP
- Number of participants who discontinued during MP
- Reasons for discontinuation during MP

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 who elect not to receive maintenance treatment are considered to have completed the study if they complete the Month 13 visit in the Post-Control Period. Meanwhile, treatment group participants who elect to receive maintenance treatment are considered to have completed the study if they complete the Month 6 visit in the Maintenance Period. Control group participants who elect not to receive optional treatment are considered to have completed the study if they complete the Month 3 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 13 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 VT and VMT populations. For control group, data after receiving optional initial treatment at Month 3 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, combined initial and touch-up, maintenance), 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, maintenance)	Initial, Touch-up, Maintenance	Frequency counts and percentages
Number (%) of participants who did not receive maintenance	Summary by reason	Maintenance	Frequency counts and percentages
Injection volume • Total • Left temple • Right temple	Summary by treatment group, treatment (initial, touch-up, initial and touch-up combined, maintenance), and treatment area	Initial, Touch-up, Initial and Touch-up combined, Maintenance	Continuous descriptive statistics

The number of participants who received treatment anesthesia will be summarized for the VT and VMT populations by treatment group. Variables related to administration of treatment listed in Table 4 will be summarized for the VT and VMT 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 	Summary by treatment group and treatment (initial, touch-up, maintenance)	Initial, Touch-up, Maintenance	Frequency counts and percentages
Pre-treatment anesthesia duration (minutes) for each pre-treatment anesthesia type 	Summary by treatment group and treatment (initial, touch-up, maintenance)	Initial, Touch-up, Maintenance	Continuous descriptive statistics
Treatment administration 	Summary by treatment group, treatment (initial, touch-up, maintenance) and treatment area (left temple, right temple) A participant may be counted in multiple planes of injection, injection techniques, and types of needles.	Initial, Touch-up, Maintenance	Frequency counts and percentages, continuous descriptive statistics
	Summary by treatment group and treatment (initial, touch-up, maintenance)	Initial, Touch-up, Maintenance	Frequency counts and percentages, continuous descriptive statistics

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; ethnicity) will be summarized descriptively in total and by treatment group for the mITT populations. Age (years) is calculated relative to informed consent date. For summary by race, participants who reported multiple races are only included in the 'Multiple' category.

Number (%) of participants who reported multiple races will be summarized as applicable.

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)
- Fitzpatrick Skin Phototype (each phototype, and by phototype groups, i.e., I and II, III and IV, V and VI)
- Exposure to sunlight (hours per day)
- Smoking history (status, tobacco product, frequency, and duration in months)
- Allergan Temple Hollowing Scale
- FACE-Q Satisfaction with Facial Appearance questionnaire, for the Rasch-transformed score
- FACE-Q Satisfaction with Temple 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, code, 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

- This section applies to both effectiveness and safety analyses when applicable.
- To account for different timing of VOLUMA XC treatment and comparable time frames of data collection and to enhance interpretability of effectiveness and safety analyses, the statistical analyses will be performed by the study periods defined in Table 2.
- 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 VT population, using as-treated grouping.
- For MP, effectiveness and safety summaries will be performed on the VMT population.
- Study day refers to Day 1 regardless of study period (e.g., CP, TP, MT).
- The baseline for effectiveness and 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 XC or no-treatment) is first received.
- The change from baseline values will be computed as the postbaseline value minus the baseline values.
- For vision assessments, baseline will refer to the most-recent measurement just prior to the most-recent treatment.
- 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 2-sided 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 3 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). [redacted]

- For TP, effectiveness and safety summaries will be performed on the VT population, using as-treated grouping.
- For MP, effectiveness and safety summaries will be performed on the VMT population.
- Study day refers to Day 1 regardless of study period (e.g., CP, TP, MT).
- The baseline for effectiveness and 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 XC or no-treatment) is first received.
- The change from baseline values will be computed as the postbaseline value minus the baseline values.
- For vision assessments, baseline will refer to the most-recent measurement just prior to the most-recent treatment.
- 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 2-sided 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 3 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). [redacted]

- Non-Responder Imputation (NRI): For sensitivity analysis, the NRI method will categorize any participants with missing Month 3 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 3 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. Each side of temple will be assessed independently by the blinded EI.

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

[redacted]

[redacted]

[redacted]

8.3.3 Primary Effectiveness Analysis

Primary effectiveness analysis will be performed based on the mITT population. The study intervention will be considered clinically effective if both of these criteria are met:

- The ATHS responder rate for the treatment group is statistically superior to the responder rate for the control group at Month 3
- At least 60% (a point estimate of 60% or greater) of participants treated with VOLUMA XC are ATHS responders at Month 3

For the first criterion, 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 3 after the latest of randomization, initial treatment and touch-up treatment, and control group at Month 3 after randomization, respectively.

If there are no mITT participants with missing scores for the primary effectiveness endpoint, a 2-sided Fisher's exact test with 5% significance level will be used to compare responder rates between treatment and the control group. Otherwise, the methods described in Section 8.3.2 will be applied. If the 2-sided p-value is ≤ 0.05, which implies that the responder rate between treatment and control is different, and the point estimate of the responder rate for the treatment group is greater than for the control group and is 60% or greater, then treatment will be considered clinically effective. In addition, responder rates and 95% CIs will be calculated for treatment group and control group. No hypothesis testing will be performed for the second criterion.

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 for the mITT population with missing ATHS responder status at Month 3 imputed as non-responders.

Summary of sensitivity analyses are presented in Table 6.

Table 6. Missing Data Handling Rules for the Sensitivity Analyses of the Primary Endpoint

Endpoint	Sensitivity Analysis No.	Missing Data Handling Rule	Population	Methodology
ATHS responder status at Month 3	1	AO method	OPE	Fisher's exact test
ATHS responder status at Month 3	2	NRI method	mITT	Fisher's exact test

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 7) in the temple area at Month 3 during CP
- Responder status for participant assessment of GAIS in the temple area at Month 3 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 3 during CP

- Change from baseline in participant responses to FACE-Q Satisfaction with Temples questionnaire, for the Rasch-transformed score (0 - 100) at Month 3 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 3 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 3 visit.

The GAIS responder rates based on EI assessment and participant assessments at Month 3 will be analyzed by descriptive statistics. The responder rates and the corresponding exact 95% CI will be estimated for treatment group and control group. The between-group difference in the responder rate and the corresponding 95% CI will be estimated, which will be considered descriptive and will be used to support the primary endpoint. No hypothesis testing procedure is planned to compare the treatment and control groups on the GAIS responder status based on EI and participant assessments.

Table 7. 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. The scales and scoring algorithms for FACE-Q questionnaires can be found in Appendix F.

For both FACE-Q scales, change from baseline to Month 3 on the Rasch-transformed score (0-100) of FACE-Q Satisfaction with Facial Appearance questionnaire, and change from baseline to Month 3 on participant responses on the Rasch-transformed score (0-100) of FACE-Q Satisfaction with Temples questionnaire will be tested within the treatment group with a 2-sided paired t-test at the 5% level to determine if the mean Rasch-transformed satisfaction score at Month 3 is statistically greater than that at baseline for the treatment group. The estimate, 95% CI, and p-value will be provided for the within-group change from Baseline to Month 3 for treatment group and control group.

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] (Appendix F-1, Appendix F-2).

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.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

9.0 Safety Analyses

9.1 General Considerations

Refer to Section 8.1 for general considerations apply to both effectiveness and safety analyses.

- For CP, safety analyses will be performed on the safety population, using as-treated grouping.
- For TP, safety summaries will be performed on the VT population, using as-treated grouping.
- For MP, safety summaries will be performed on the VMT 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 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 and MP using the VT and VMT populations, respectively as described in Table 9. 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 9.

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.

For participants who received maintenance treatment, their AEs reported during the first 6 months from the initial treatment will be summarized and presented in parallel with AEs reported during the maintenance treatment period. The summary for the initial treatment will be presented for participants who received initial treatment only and participants who received both initial and touch-up treatments.

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 9. 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 • Discontinued due to TEAEs • AESIs • Treatment-related AESIs • Deaths • TEAEs related to COVID-19 infection	CP, TP, MP	Frequency counts and percentages
TEAEs	• By SOC and PT • By SOC, PT, and maximum severity	CP, TP, MP	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, and outcome and categorized by severity	CP, TP, MP	Frequency counts and percentages
TESAEs	By SOC and PT	CP, TP, MP	Frequency counts and percentages
AEs leading to study discon-tinuation	By SOC and PT	CP, TP, MP	Frequency counts and percentages

9.2.3 Adverse Events of Special Interest

The following AESIs have been identified for the study intervention in this study: visual disturbance (including, but not limited to, any loss of vision, vision, double vision, pain in

or around eye [excluding pain in temple area], blind spot or shadow in the visual field, trouble moving eyes, etc.).

The incidence of AESIs will be summarized by SOC and PT and listed in a separate listing.

9.3 Analysis of Laboratory Data

Not applicable.

9.4 Vital Signs

Vital sign is collected at screening including blood pressure (systolic and diastolic; while participant is seated), pulse rate, and temperature. Vital sign will be listed.

9.5 Height and Weight

Height is collected at Screening visit. Weight is collected at Screening, Month 3 in CP for both treatment and control groups and at Month 13 in TP for treatment group. 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 recorded in participant diaries after each treatment (initial, touch-up, and maintenance). 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 10.

Table 10. Injection Site Response Analyses

Endpoint	Description	Timing	Methodology
ISR severity	Number of participants with ISRs by symptom incidence and categorized by maximum reported severity	Initial, Touch-up, Maintenance	Frequency counts and percentages
ISR duration	Number of participants with ISRs by symptom incidence and categorized by duration. Deration is defined as number of days from first instance of the symptom to the last instance of the symptom within the planned diary period. Duration is derived as date of last ISR - date of first ISR + 1. Duration of any ISR is calculated using the same algorithm regardless of ISR symptom.	Initial, Touch-up, Maintenance	Frequency counts and percentages, continuous descriptive statistics
Ongoing ISR	ISRs recorded on Ongoing ISR page		Listing

If a participant reported more than one "Other" ISR, the participant will be counted once in the summary for that treatment. For "Other" ISR, severity and duration analyses described in Table 10 will be performed. "Other" ISR may be summarized by category when free text categorization is applicable and relevant information is available.

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 and maintenance treatment will be summarized. The summary will be performed by treatment group for VT and VMT as described in Table 11.

Table 11. Procedural Pain Analyses

Endpoint	Description	Timing	Methodology
Procedural pain	Summary of pain scores as continuous scale	Initial, Touch-up, Maintenance	Categorical descriptive statistics

[redacted]

9.10 Other Safety Analyses

Other safety endpoints include the line change on Snellen visual acuity, confrontational visual fields and ocular motility, the global score of jaw functional (see Appendix F-3 Jaw Functional Limitation Scale Questionnaire), and concomitant medication and concurrent procedures.

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

Line change based on Snellen visual acuity will be calculated using the formula below. The number of lines in change from baseline in visual acuity will be tabulated. A positive

change represents an improvement, and a negative change represents a worsening. Baseline is the pre-treatment value on or before the most recent treatment.

Line change = 10 x [log10 ($d_{baseline}$/20) – log10 ($d_{follow-up}$/20)]

where $d_{baseline}$ = denominator of the Snellen equivalent score at baseline,

$d_{follow-up}$ = denominator of the Snellen equivalent score at follow-up visit

Positive findings in confrontation visual fields assessments (e.g., No is reported for "Is eye full to confrontation?" or Yes is reported for "Is there a change from pre-treatment?"), will be listed.

Positive findings in ocular motility (e.g., No is selected for "Does the eye have full duction and version?" or Yes is reported for "Is there a change from pre-treatment?") will be listed.

[redacted]

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

10.0 Other Analyses

Not applicable.

[redacted]

[redacted]

[redacted]

[redacted]

11.1 Data Monitoring Committee

No data monitoring committee is planned for this study.

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. All CIs will be 2-sided 95% CIs, unless stated otherwise. A gatekeeping procedure will be used for hypothesis testing of the primary endpoint (ATHS responder) and 2 of the secondary endpoints (FACE-Q Satisfaction with Facial Appearance and FACE-Q Satisfaction with Temples) 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 FACE-Q Satisfaction with Facial Appearance. Additionally, statistical significance must be established for FACE-Q Satisfaction with Facial Appearance in order to assess the hypothesis for FACE-Q Satisfaction with Temples.

For the secondary effectiveness endpoints of GAIS responder status at Month 3 as assessed by the EI and as assessed by the participant, the 95% CI will be provided for the proportion of responders within each treatment group, where applicable. No multiplicity

adjustments will be made for the 2 GAIS responder rates since these endpoints and associated statistics are intended to be descriptive rather than inferential.

13.0 Version History

SAP Version History Summary			
SAP Version	**Approval Date**	**Change**	**Rationale**
1	27-Aug-2020	Not Applicable	Original version
2	23-June-2021	Added Interim analysis	Per protocol amendment 2
		Added Appendix C	Per filler study standard
		Writing format change	Per AbbVie SAP template
3	22-September-2021	FACE-Q Satisfaction with Temple Scale update	Appendix F-2, per FACE-Q developer validation report
		Interim efficacy analysis includes visits up to Month 6	Section 11, Adding clarification.
		Volume change of each temple will be summarized by each side (left temple, right temple) as assessed by 3D imaging.	Section 8.5, to be consistent with injection volume data collected by each temple.

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 12. 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 and variables affected by COVID-19 will be included, even if there is no associated significant protocol deviation.

Table 12. 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

Refer to Section 9.2.3.

Appendix C. Changes to Protocol-planned Analysis

The subgroup analysis during the control period by volume injected (less than median vs. median or greater) will be only performed for treatment group and will not be done for control group. Volume injected does not apply to control participants during control period.

Appendix D. List of Abbreviations

Abbreviation	Definition
3D	3-dimensional
AE	adverse event
AESI	adverse event of special interest
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
MI	Multiple Imputation
mITT	modified intent-to-treat
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MP	Maintenance Treatment Period
NRI	Non-Responder Imputation
OPE	Observed Primary Endpoint
PCP	Post-control Period
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
VMT	VOLUMA XC Maintenance Treatment

Abbreviation	Definition
VT	VOLUMA XC Treated
WHO	World Health Organization

Appendix E. Changes to Protocol-planned Analyses

Not applicable.

Appendix F. Scales and Scoring Algorithms

F-1. FACE-Q Satisfaction with Facial Appearance questionnaire

FACE-Q SATISFACTION WITH FACIAL APPEARANCE

	Very Dissatisfied	Somewhat Dissatisfied	Somewhat Satisfied	Very Satisfied
a. How symmetric your face looks?	1	2	3	4
b. How balanced your face looks?	1	2	3	4
c. How well-proportioned your face looks?	1	2	3	4
d. How your face looks at the end of day?	1	2	3	4
e. How fresh your face looks?	1	2	3	4
f. How rested your face looks?	1	2	3	4
g. How your profile (side view) looks?	1	2	3	4
h. How your face looks in photos?	1	2	3	4
i. How your face looks when first wake up?	1	2	3	4
j. How your face looks under bright lights?	1	2	3	4

FACE-Q SATISFACTION WITH FACIAL APPEARANCE CONVERSION TABLE

Instructions: Ensure the data are rescored as follows: "Very Satisfied" = 4; "Somewhat Satisfied" = 3; "Somewhat Dissatisfied" = 2; "Very Dissatisfied" = 1. Higher scores reflect a better outcome. If 1 to 5 of a participant's 10 item scores were missing, the non-missing sum was prorated to a 10-item sum and rounded to the nearest integer. If 6 or more of the 10 item scores were missing, the sum was set to missing. Use the Conversion Table below to convert the raw summed scale score into a score from 0 (worst) to 100 (best).

SUM SCORE	EQUIVALENT RASCH TRANSFORMED SCORE (0-100)
10	0
11	11
12	16
13	20
14	23
15	26
16	29
17	31
18	33
19	35
20	38
21	40
22	42
23	44
24	46
25	48
26	51
27	53
28	55
29	58
30	61
31	64
32	66
33	69
34	72
35	76
36	79
37	82
38	87
39	92
40	100

F-2. FACE-Q Satisfaction with Temples questionnaire

FACE-Q SATISFACTION WITH TEMPLES

	Very Dissatisfied	Somewhat Dissatisfied	Somewhat Satisfied	Very Satisfied	Remark
1. How Well Temples Match Each Other	1	2	3	4	Removed
2. How Full Temples Look	1	2	3	4	
3. The Shape Of Temples	1	2	3	4	
4. How Temples Look In Mirror	1	2	3	4	
5. How Temples Look From The Side	1	2	3	4	
6. How Temples Look Slightly To Side	1	2	3	4	
7. How Temples Look In Photos	1	2	3	4	
8. How Temples Look Under Light	1	2	3	4	
9. Smooth Skin In Temple Area Looks	1	2	3	4	Removed
10. How Symmetric Temples Look	1	2	3	4	Removed
11. Temples Compliment Shape Of Face	1	2	3	4	
12. The Age Temples Make You Look	1	2	3	4	
13. Compared With Other Ppl Your Age	1	2	3	4	
14. How Well Temples Fit With Face	1	2	3	4	
15. The Contour Of Temples	1	2	3	4	Removed
16. How Youthful Temples Make You	1	2	3	4	

FACE-Q SATISFACTION WITH TEMPLES CONVERSION TABLE

Instructions: Ensure the data are scored as follows: "Very Satisfied" = 4; "Somewhat Satisfied" = 3; "Somewhat Dissatisfied" = 2; "Very Dissatisfied" = 1. Higher scores reflect a better outcome.

At the start of this study, 16 items were included in the FACE-Q Satisfaction with Temple Scale and data were collected throughout the study. Based on the result of validation analysis conducted using the baseline data of this study, items 1, 9, 10, 15 were excluded from the scale due to misfit to the Rasch model. The final analysis for this study will include 12 items (i.e., item 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 16).

The sum of the 12 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 less than 50% of all the items (i.e., at least 7 items have been answered) then insert the unrounded mean of the completed items into the total sum score. Otherwise, the overall score will be missing. The summed score including the imputation of missing items is rounded to the nearest integer and converted to the Rasch-transformed score using the below conversion table.

SUM SCORE	EQUIVALENT RASCH TRANSFORMED SCORE (0-100)
12	0
13	10
14	14
15	16
16	18
17	21
18	23
19	24
20	27
21	29
22	31
23	33
24	36
25	38
26	41
27	43
28	45
29	47
30	49
31	50
32	52
33	54
34	57
35	59
36	62
37	65
38	68
39	71
40	73
41	76
42	79
43	81
44	84
45	87
46	90
47	95
48	100

F-3. Jaw Functional Limitation Scale (JFLS) and Scoring Algorithm

Description

The JFLS was initially developed as an 8-item global scale for overall functional limitation of the masticatory system; based on the resultant items and supporting psychometric data, the instrument was re-developed in order to expand measured constructs to also include masticatory limitation, vertical mobility limitation, and verbal and non-verbal communication limitation, comprised within a 20-item instrument that also retained the items for the short global scale. Consequently, the full instrument could be used at baseline, from which all three subscales as well as the global score could be derived, and the short instrument could be used at follow-up, from which the global score could be derived; measurement congruence across time for a global score would be retained in addition to having subscale scores at baseline. Alternatively, one research group could use the short form and another group could use the long form, and the subscale scores would have measurement congruence across the two settings due to the very high reliability of the global score, whether derived from the full instrument or from the short instrument.

The 20 items are listed below. The score for each item is 0 to 10 with 0 being no limitation and 10 being severe limitation.

1. Chew tough food
2. Chew hard bread
3. Chew chicken
4. Chew crackers
5. Chew soft food
6. Eat soft food requiring no chewing
7. Open wide enough to bite from apple
8. Open wide enough to bite into a sandwich

9. Open wide enough to talk
10. Open wide enough to drink from a cup
11. Swallow
12. Yawn
13. Talk
14. Sing
15. Putting on a happy face
16. Putting on an angry face
17. Frown
18. Kiss
19. Smile
20. Laugh

Scoring

A single global score of "jaw functional limitation" can be computed as the mean of the available items.

Subscale scores for each type of functional limitation are computed, as follows:

- *Mastication:* mean of items 1-6.
- *Mobility:* mean of items 7-10.
- *Verbal and non-verbal communication:* mean of items 13-20.

A second type of global score can be obtained from the long form by computing the mean of the 3 subscale scores, as computed above. Note that all 3 subscale scores must be present in order to compute the global score in this manner.

Alternative scoring can be achieved through the use of Rasch software, but this is not further described in this manual.

Missing data

For the JFLS-20, scores can be computed based on no more than the following number of items with missing response: short form, 2 items missing allowed; mastication, 2 items missing allowed; mobility, 1 item missing allowed; and communication, 2 items missing allowed. For the JFLS-8, no more than 2 items may be missing. Computation of a score with missing items is adjusted by dividing by number of items present.

Interpretation

Norms have not yet been established for this instrument. Based on comparison of individuals who were lifetime negative for TMD to those with chronic TMD, observed scores were as follows:

	No lifetime TMD		Chronic TMD	
Scale	Mean	SE	Mean	SE
Mastication limitation	0.28	0.02	2.22	0.13
Mobility limitation	0.18	0.02	2.22	0.13
Verbal and Emotional Expression Limitation	0.14	0.02	0.72	0.10
Global	0.16	0.02	1.74	0.11

References

Ohrbach R, Larsson P, List T. The Jaw Functional Limitation Scale: Development, reliability, and validity of 8-item and 20-item versions. J Orofac Pain. 2008;22(3):219-30.

Ohrbach R, Fillingim RB, Mulkey F, et al. Clinical findings and pain symptoms as potential risk factors for chronic TMD: Descriptive data and empirically identified domains from the OPPERA case-control study. J Pain. 2011;12 (11 Suppl):T27-45.

[redacted]

[redacted]

8.3.3 Primary Effectiveness Analysis

Primary effectiveness analysis will be performed based on the mITT population. The study intervention will be considered clinically effective if both of these criteria are met:

- The ATHS responder rate for the treatment group is statistically superior to the responder rate for the control group at Month 3
- At least 60% (a point estimate of 60% or greater) of participants treated with VOLUMA XC are ATHS responders at Month 3

For the first criterion, 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 3 after the latest of randomization, initial treatment and touch-up treatment, and control group at Month 3 after randomization, respectively.

If there are no mITT participants with missing scores for the primary effectiveness endpoint, a 2-sided Fisher's exact test with 5% significance level will be used to compare responder rates between treatment and the control group. Otherwise, the methods described in Section 8.3.2 will be applied. If the 2-sided p-value is ≤ 0.05, which implies that the responder rate between treatment and control is different, and the point estimate of the responder rate for the treatment group is greater than for the control group and is 60% or greater, then treatment will be considered clinically effective. In addition, responder rates and 95% CIs will be calculated for treatment group and control group. No hypothesis testing will be performed for the second criterion.

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 for the mITT population with missing ATHS responder status at Month 3 imputed as non-responders.

Summary of sensitivity analyses are presented in Table 6.

Table 6. Missing Data Handling Rules for the Sensitivity Analyses of the Primary Endpoint

Endpoint	Sensitivity Analysis No.	Missing Data Handling Rule	Population	Methodology
ATHS responder status at Month 3	1	AO method	OPE	Fisher's exact test
ATHS responder status at Month 3	2	NRI method	mITT	Fisher's exact test

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 7) in the temple area at Month 3 during CP
- Responder status for participant assessment of GAIS in the temple area at Month 3 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 3 during CP

- Change from baseline in participant responses to FACE-Q Satisfaction with Temples questionnaire, for the Rasch-transformed score (0 - 100) at Month 3 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 3 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 3 visit.

The GAIS responder rates based on EI assessment and participant assessments at Month 3 will be analyzed by descriptive statistics. The responder rates and the corresponding exact 95% CI will be estimated for treatment group and control group. The between-group difference in the responder rate and the corresponding 95% CI will be estimated, which will be considered descriptive and will be used to support the primary endpoint. No hypothesis testing procedure is planned to compare the treatment and control groups on the GAIS responder status based on EI and participant assessments.

Table 7. 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. The scales and scoring algorithms for FACE-Q questionnaires can be found in Appendix F.

For both FACE-Q scales, change from baseline to Month 3 on the Rasch-transformed score (0-100) of FACE-Q Satisfaction with Facial Appearance questionnaire, and change from baseline to Month 3 on participant responses on the Rasch-transformed score (0-100) of FACE-Q Satisfaction with Temples questionnaire will be tested within the treatment group with a 2-sided paired t-test at the 5% level to determine if the mean Rasch-transformed satisfaction score at Month 3 is statistically greater than that at baseline for the treatment group. The estimate, 95% CI, and p-value will be provided for the within-group change from Baseline to Month 3 for treatment group and control group.

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] (Appendix F-1, Appendix F-2).

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.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

9.0 Safety Analyses

9.1 General Considerations

Refer to Section 8.1 for general considerations apply to both effectiveness and safety analyses.

- For CP, safety analyses will be performed on the safety population, using as-treated grouping.
- For TP, safety summaries will be performed on the VT population, using as-treated grouping.
- For MP, safety summaries will be performed on the VMT 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 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 and MP using the VT and VMT populations, respectively as described in Table 9. 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 9.

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.

For participants who received maintenance treatment, their AEs reported during the first 6 months from the initial treatment will be summarized and presented in parallel with AEs reported during the maintenance treatment period. The summary for the initial treatment will be presented for participants who received initial treatment only and participants who received both initial and touch-up treatments.

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 9. 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 • Discontinued due to TEAEs • AESIs • Treatment-related AESIs • Deaths • TEAEs related to COVID-19 infection	CP, TP, MP	Frequency counts and percentages
TEAEs	• By SOC and PT • By SOC, PT, and maximum severity	CP, TP, MP	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, and outcome and categorized by severity	CP, TP, MP	Frequency counts and percentages
TESAEs	By SOC and PT	CP, TP, MP	Frequency counts and percentages
AEs leading to study discon-tinuation	By SOC and PT	CP, TP, MP	Frequency counts and percentages

9.2.3 Adverse Events of Special Interest

The following AESIs have been identified for the study intervention in this study: visual disturbance (including, but not limited to, any loss of vision, vision, double vision, pain in

or around eye [excluding pain in temple area], blind spot or shadow in the visual field, trouble moving eyes, etc.).

The incidence of AESIs will be summarized by SOC and PT and listed in a separate listing.

9.3 Analysis of Laboratory Data

Not applicable.

9.4 Vital Signs

Vital sign is collected at screening including blood pressure (systolic and diastolic; while participant is seated), pulse rate, and temperature. Vital sign will be listed.

9.5 Height and Weight

Height is collected at Screening visit. Weight is collected at Screening, Month 3 in CP for both treatment and control groups and at Month 13 in TP for treatment group. 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 recorded in participant diaries after each treatment (initial, touch-up, and maintenance). 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 10.

Table 10. Injection Site Response Analyses

Endpoint	Description	Timing	Methodology
ISR severity	Number of participants with ISRs by symptom incidence and categorized by maximum reported severity	Initial, Touch-up, Maintenance	Frequency counts and percentages
ISR duration	Number of participants with ISRs by symptom incidence and categorized by duration. Deration is defined as number of days from first instance of the symptom to the last instance of the symptom within the planned diary period. Duration is derived as date of last ISR - date of first ISR + 1. Duration of any ISR is calculated using the same algorithm regardless of ISR symptom.	Initial, Touch-up, Maintenance	Frequency counts and percentages, continuous descriptive statistics
Ongoing ISR	ISRs recorded on Ongoing ISR page		Listing

If a participant reported more than one "Other" ISR, the participant will be counted once in the summary for that treatment. For "Other" ISR, severity and duration analyses described in Table 10 will be performed. "Other" ISR may be summarized by category when free text categorization is applicable and relevant information is available.

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 and maintenance treatment will be summarized. The summary will be performed by treatment group for VT and VMT as described in Table 11.

Table 11. Procedural Pain Analyses

Endpoint	Description	Timing	Methodology
Procedural pain	Summary of pain scores as continuous scale	Initial, Touch-up, Maintenance	Categorical descriptive statistics

[redacted]

9.10 Other Safety Analyses

Other safety endpoints include the line change on Snellen visual acuity, confrontational visual fields and ocular motility, the global score of jaw functional (see Appendix F-3 Jaw Functional Limitation Scale Questionnaire), and concomitant medication and concurrent procedures.

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

Line change based on Snellen visual acuity will be calculated using the formula below. The number of lines in change from baseline in visual acuity will be tabulated. A positive

change represents an improvement, and a negative change represents a worsening. Baseline is the pre-treatment value on or before the most recent treatment.

Line change = 10 x [log10 ($d_{baseline}$/20) – log10 ($d_{follow-up}$/20)]

where $d_{baseline}$ = denominator of the Snellen equivalent score at baseline,

$d_{follow-up}$ = denominator of the Snellen equivalent score at follow-up visit

Positive findings in confrontation visual fields assessments (e.g., No is reported for "Is eye full to confrontation?" or Yes is reported for "Is there a change from pre-treatment?"), will be listed.

Positive findings in ocular motility (e.g., No is selected for "Does the eye have full duction and version?" or Yes is reported for "Is there a change from pre-treatment?") will be listed.

[redacted]

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

10.0 Other Analyses

Not applicable.

[redacted]

[redacted]

[redacted]

[redacted]

11.1 Data Monitoring Committee

No data monitoring committee is planned for this study.

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. All CIs will be 2-sided 95% CIs, unless stated otherwise. A gatekeeping procedure will be used for hypothesis testing of the primary endpoint (ATHS responder) and 2 of the secondary endpoints (FACE-Q Satisfaction with Facial Appearance and FACE-Q Satisfaction with Temples) 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 FACE-Q Satisfaction with Facial Appearance. Additionally, statistical significance must be established for FACE-Q Satisfaction with Facial Appearance in order to assess the hypothesis for FACE-Q Satisfaction with Temples.

For the secondary effectiveness endpoints of GAIS responder status at Month 3 as assessed by the EI and as assessed by the participant, the 95% CI will be provided for the proportion of responders within each treatment group, where applicable. No multiplicity

adjustments will be made for the 2 GAIS responder rates since these endpoints and associated statistics are intended to be descriptive rather than inferential.

13.0 Version History

SAP Version History Summary			
SAP Version	**Approval Date**	**Change**	**Rationale**
1	27-Aug-2020	Not Applicable	Original version
2	23-June-2021	Added Interim analysis	Per protocol amendment 2
		Added Appendix C	Per filler study standard
		Writing format change	Per AbbVie SAP template
3	22-September-2021	FACE-Q Satisfaction with Temple Scale update	Appendix F-2, per FACE-Q developer validation report
		Interim efficacy analysis includes visits up to Month 6	Section 11, Adding clarification.
		Volume change of each temple will be summarized by each side (left temple, right temple) as assessed by 3D imaging.	Section 8.5, to be consistent with injection volume data collected by each temple.

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 12. 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 and variables affected by COVID-19 will be included, even if there is no associated significant protocol deviation.

Table 12. 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

Refer to Section 9.2.3.

Appendix C. Changes to Protocol-planned Analysis

The subgroup analysis during the control period by volume injected (less than median vs. median or greater) will be only performed for treatment group and will not be done for control group. Volume injected does not apply to control participants during control period.

Appendix D. List of Abbreviations

Abbreviation	Definition
3D	3-dimensional
AE	adverse event
AESI	adverse event of special interest
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
MI	Multiple Imputation
mITT	modified intent-to-treat
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MP	Maintenance Treatment Period
NRI	Non-Responder Imputation
OPE	Observed Primary Endpoint
PCP	Post-control Period
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
VMT	VOLUMA XC Maintenance Treatment

Abbreviation	Definition
VT	VOLUMA XC Treated
WHO	World Health Organization

Appendix E. Changes to Protocol-planned Analyses

Not applicable.

FACE-Q SATISFACTION WITH FACIAL APPEARANCE CONVERSION TABLE

Instructions: Ensure the data are rescored as follows: "Very Satisfied" = 4; "Somewhat Satisfied" = 3; "Somewhat Dissatisfied" = 2; "Very Dissatisfied" = 1. Higher scores reflect a better outcome. If 1 to 5 of a participant's 10 item scores were missing, the non-missing sum was prorated to a 10-item sum and rounded to the nearest integer. If 6 or more of the 10 item scores were missing, the sum was set to missing. Use the Conversion Table below to convert the raw summed scale score into a score from 0 (worst) to 100 (best).

SUM SCORE	EQUIVALENT RASCH TRANSFORMED SCORE (0-100)
10	0
11	11
12	16
13	20
14	23
15	26
16	29
17	31
18	33
19	35
20	38
21	40
22	42
23	44
24	46
25	48
26	51
27	53
28	55
29	58
30	61
31	64
32	66
33	69
34	72
35	76
36	79
37	82
38	87
39	92
40	100

FACE-Q SATISFACTION WITH TEMPLES CONVERSION TABLE

Instructions: Ensure the data are scored as follows: "Very Satisfied" = 4; "Somewhat Satisfied" = 3; "Somewhat Dissatisfied" = 2; "Very Dissatisfied" = 1. Higher scores reflect a better outcome.

At the start of this study, 16 items were included in the FACE-Q Satisfaction with Temple Scale and data were collected throughout the study. Based on the result of validation analysis conducted using the baseline data of this study, items 1, 9, 10, 15 were excluded from the scale due to misfit to the Rasch model. The final analysis for this study will include 12 items (i.e., item 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 16).

The sum of the 12 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 less than 50% of all the items (i.e., at least 7 items have been answered) then insert the unrounded mean of the completed items into the total sum score. Otherwise, the overall score will be missing. The summed score including the imputation of missing items is rounded to the nearest integer and converted to the Rasch-transformed score using the below conversion table.

SUM SCORE	EQUIVALENT RASCH TRANSFORMED SCORE (0-100)
12	0
13	10
14	14
15	16
16	18
17	21
18	23
19	24
20	27
21	29
22	31
23	33
24	36
25	38
26	41
27	43
28	45
29	47
30	49
31	50
32	52
33	54
34	57
35	59
36	62
37	65
38	68
39	71
40	73
41	76
42	79
43	81
44	84
45	87
46	90
47	95
48	100

F-3. Jaw Functional Limitation Scale (JFLS) and Scoring Algorithm

Description

The JFLS was initially developed as an 8-item global scale for overall functional limitation of the masticatory system; based on the resultant items and supporting psychometric data, the instrument was re-developed in order to expand measured constructs to also include masticatory limitation, vertical mobility limitation, and verbal and non-verbal communication limitation, comprised within a 20-item instrument that also retained the items for the short global scale. Consequently, the full instrument could be used at baseline, from which all three subscales as well as the global score could be derived, and the short instrument could be used at follow-up, from which the global score could be derived; measurement congruence across time for a global score would be retained in addition to having subscale scores at baseline. Alternatively, one research group could use the short form and another group could use the long form, and the subscale scores would have measurement congruence across the two settings due to the very high reliability of the global score, whether derived from the full instrument or from the short instrument.

The 20 items are listed below. The score for each item is 0 to 10 with 0 being no limitation and 10 being severe limitation.

1. Chew tough food
2. Chew hard bread
3. Chew chicken
4. Chew crackers
5. Chew soft food
6. Eat soft food requiring no chewing
7. Open wide enough to bite from apple
8. Open wide enough to bite into a sandwich

9. Open wide enough to talk
10. Open wide enough to drink from a cup
11. Swallow
12. Yawn
13. Talk
14. Sing
15. Putting on a happy face
16. Putting on an angry face
17. Frown
18. Kiss
19. Smile
20. Laugh

Scoring

A single global score of "jaw functional limitation" can be computed as the mean of the available items.

Subscale scores for each type of functional limitation are computed, as follows:

- *Mastication:* mean of items 1-6.
- *Mobility:* mean of items 7-10.
- *Verbal and non-verbal communication:* mean of items 13-20.

A second type of global score can be obtained from the long form by computing the mean of the 3 subscale scores, as computed above. Note that all 3 subscale scores must be present in order to compute the global score in this manner.

Alternative scoring can be achieved through the use of Rasch software, but this is not further described in this manual.

Missing data

For the JFLS-20, scores can be computed based on no more than the following number of items with missing response: short form, 2 items missing allowed; mastication, 2 items missing allowed; mobility, 1 item missing allowed; and communication, 2 items missing allowed. For the JFLS-8, no more than 2 items may be missing. Computation of a score with missing items is adjusted by dividing by number of items present.

Interpretation

Norms have not yet been established for this instrument. Based on comparison of individuals who were lifetime negative for TMD to those with chronic TMD, observed scores were as follows:

	No lifetime TMD		Chronic TMD	
Scale	Mean	SE	Mean	SE
Mastication limitation	0.28	0.02	2.22	0.13
Mobility limitation	0.18	0.02	2.22	0.13
Verbal and Emotional Expression Limitation	0.14	0.02	0.72	0.10
Global	0.16	0.02	1.74	0.11

References

Ohrbach R, Larsson P, List T. The Jaw Functional Limitation Scale: Development, reliability, and validity of 8-item and 20-item versions. J Orofac Pain. 2008;22(3):219-30.

Ohrbach R, Fillingim RB, Mulkey F, et al. Clinical findings and pain symptoms as potential risk factors for chronic TMD: Descriptive data and empirically identified domains from the OPPERA case-control study. J Pain. 2011;12 (11 Suppl):T27-45.