Statistical Analysis Plan for Study 2029-701-008

Multicenter, Evaluator-blinded, Randomized, Controlled Study of the Safety and Effectiveness of JUVÉDERM® VOLITE™ XC Injectable Gel for Improvement in Neck Appearance

Date: 07 October 2024

Version 5.0

Table of Contents

1.0 Introduction 5
2.0 Study Design and Objectives 5
2.1 Study Objectives and Hypothesis 5
2.2 Study Design Overview 5
2.3 Treatment Assignment and Blinding 6
2.4 Sample Size Determination 7
2.5 Sample Size Re-estimation 8
3.0 Endpoints 9
3.1 Primary Effectiveness Endpoint 9
3.2 Secondary Effectiveness Endpoints 10
3.3 Other Effectiveness Endpoints 10
3.4 Exploratory Effectiveness Endpoint 11
3.5 Safety Endpoints 11
4.0 Analysis Populations 11
5.0 Participant Disposition 12
6.0 Study Device Exposure and Administration 15
7.0 Demographics, Baseline Characteristics, Medical History, and Prior/Concomitant Medications 17
7.1 Demographics and Baseline Characteristics 17
7.2 Medical History 17
7.3 Prior and Concomitant Medications 18
8.0 Effectiveness Analyses 18
8.1 General Considerations 18
8.2 Handling of Missing Data 19
8.3 Primary Effectiveness Endpoint and Analyses 19
8.3.1 Primary Effectiveness Endpoint 19
8.3.2 Handling of Missing Data for the Primary Effectiveness Endpoint 20
8.3.3 Primary Effectiveness Analysis 21
8.3.4 Additional Analyses of the Primary Effectiveness Endpoints 23
8.4 Secondary Effectiveness Endpoints and Analyses 24
8.4.1 Secondary Effectiveness Endpoints 24

8.4.2 Secondary Effectiveness Analyses 25
8.5 Other Effectiveness Analyses 31
8.6 Effectiveness Subgroup Analyses 33
9.0 Safety Analyses 34
9.1 General Considerations 34
9.2 Adverse Events 34
9.2.1 Treatment-Emergent Adverse Events 35
9.2.2 Adverse Event Overview 35
9.2.3 Treatment-Emergent Adverse Events by SOC and/or PT 38
9.2.4 Treatment-Emergent SAEs (Including Deaths) and Treatment-Emergent Adverse Events leading to Withdrawal of Study Treatment 38
9.2.5 Adverse Events of Special Interest 38
9.3 Pregnancy Test 38
9.4 Analysis of Vital Signs 39
9.5 Injection Site Response 39
9.6 Procedural Pain 40
9.7 Safety Subgroup Analyses 40
9.8 Other Safety Analyses 41
10.0 Other Analyses 41
11.0 Interim Analyses 42
11.1 Data Monitoring Committee 42
12.0 Overall Type-I Error Control 42
13.0 Version History 45
14.0 References 47

List of Tables

Table 1. Analysis Sets 12
Table 2. Definition of the Control and Treated Periods by Analysis Group (as-randomized for Effectiveness or as-treated for Safety) 13
Table 3. Study Device Exposure 15
Table 4. Administration of Study Treatment 16

Table 5. Allergan Transverse Neck Lines Scale 20
Table 6. Missing Data Handling Rules for the Sensitivity Analyses of the Primary Endpoint 23
Table 7. Global Aesthetic Improvement Scale 24
Table 8. Secondary Effectiveness Analyses 30
Table 9. Other Effectiveness Analyses 33
Table 10. TEAE Summaries 37
Table 11. Injection Site Response Analyses 40
Table 12. Procedural Pain Analyses 40
Table 13. Protocol Deviation Summary 48

List of Figures

Figure 1. Study Schematic 6

List of Appendices

Appendix A. Protocol Deviations 48
Appendix B. Definition of Adverse Events of Special Interest 49
Appendix C. Changes to Protocol-planned Analysis 50
Appendix D. List of Abbreviations 51

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for VOLITE XC Study 2029-701-008 Multicenter, Evaluator-blinded, Randomized, Controlled Study of the Safety and Effectiveness of JUVEDERM® VOLITE™ XC Injectable Gel for Improvement in Neck Appearance (Amendment 5).

Study 2029-701-008 examines the safety and effectiveness of VOLITE XC in adult participants with moderate or severe transverse neck lines [Allergan Transverse Neck Lines Scale (ATNLS) Grades 2 or 3].

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

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

This SAP includes changes to analyses described in the protocol. Details are outlined in Section 13.0.

2.0 Study Design and Objectives

2.1 Study Objectives and Hypothesis

The objectives of this study are to evaluate the safety and effectiveness of VOLITE XC in adults seeking improvement in neck appearance.

The statistical hypothesis is that the ATNLS responder rate for the VOLITE XC group is statistically superior to the responder rate for the control group at Month 1 and greater than 50%.

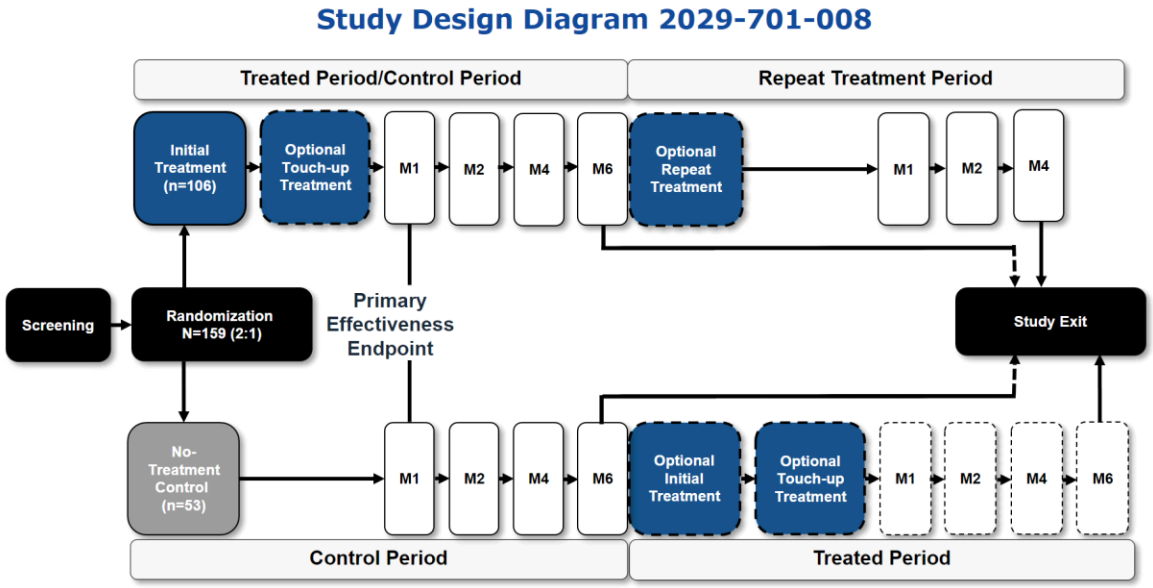
2.2 Study Design Overview

This is a multicenter, evaluator-blinded, randomized, controlled study to evaluate the safety and effectiveness of VOLITE XC for improvement in neck appearance. Safety data

for the first 20 treated participants through the 20th treated participant's 14th day after initial treatment will be submitted to the FDA prior to proceeding with enrollment of additional participants. See protocol Section 4.1 Overall Design for details.

The schematic of the study is shown in Figure 1.

Figure 1. Study Schematic



Note: The dotted lines and box outlines in the above schema indicate that the corresponding activity are optional based on the study design.

2.3 Treatment Assignment and Blinding

Participants will be randomized to VOLITE XC group or control group in a 2:1 ratio.

The study design is single-blind where the Evaluating Investigators (EIs) and image analysis technicians will remain blinded to each participant's assigned study intervention throughout the course of the study. The Treating Investigators (TIs), study coordinators, and participants will not be blinded to study intervention. The TIs and study coordinators

will secure the randomization and other records (e.g., records of study interventions) from potential discovery by the blinded EIs. The TIs, study coordinators, and participants will not discuss the randomized treatment assignments with or in the presence of the EIs.

2.4 Sample Size Determination

Up to 212 participants will be screened. Assuming a screen failure rate of 25%, a total of 159 participants will be randomized to either the treatment group or the control group in a 2:1 randomization ratio. Assuming an attrition rate of 10%, 96 and 48 participants in the treatment group and the control group, respectively, will provide at least 93% power to detect a between-group difference of 30% in the ATNLS responder rate at Month 1 of Control Period (CP). This calculation is based on a Chi-squared test at 5% significance level assuming 65% vs. 35% for the ATNLS responder rate of the treatment group vs. the control group, respectively. In addition, 96 participants in the treatment group will provide at least 84% power to detect a statistical difference in the ATNLS responder rate of the treatment group from 50% at Month 1 of CP. This calculation is based on a 1-sided 1-sample Z-test at 2.5% significance level assuming 65% for the ATNLS responder rate of the treatment group. The overall power to detect both a between-group difference of 30% in the ATNLS responder rate and the ATNLS responder rate of the treatment group greater than 50% at Month 1 of CP is at least 80%. This calculation is based on a simulation of 10,000 iterations for the hypothesis testing assumptions described above. With 20 investigational sites planned, it is recommended that no site randomize more than 20 participants.

Assuming a 5% drop-out rate between randomization and initial treatment for the treatment group, and a 10% drop-out/study completion rate between randomization and optional treatment for the control group, it is estimated that at least 147 participants will be treated in the study. With 147 treated participants, the probability to observe at least 1 incidence of treatment-related AE is 94.9% assuming an incidence rate of 2% in the population based on the cumulative distribution function for the binomial distribution.

The modules of PTTOU-1 for the two-group Chi-square test of equal proportions (unequal n's) and POT0-1 for the 1-sample Chi-square test of commercial software nQuery were used for the sample size and power calculation.

2.5 Sample Size Re-estimation

A blinded sample size re-estimation will be conducted when at least 50% of randomized participants have completed the Month 1 visit of the CP. The sample size re-estimation will be conducted in a blinded fashion without breaking the randomization code. Thus, there is no impact to overall type I error. Study participants and investigators will not be aware of the process or the results of the sample size re-estimation process both during and after the re-estimation. The results of the re-estimation will be for AbbVie's evaluation and will not be communicated to the study participants or investigators. Thus, there will be no impact to neither the study conduct nor bias to the study outcome after the sample size re-estimation. Therefore, the study integrity will be maintained. The sample size re-estimation process will be performed by the steps described below:

1. The overall ATNLS responder rate at the Month 1 of CP, $\hat{P}$, will be estimated based on the pooled data without breaking the randomization code.
2. The ATNLS responder rate of the control group at Month 1 of CP, $\hat{P}_c$, will be estimated by 3 sources of data:

    a. *Cumulative rater qualification data of protocol 2029-701-008*
    b. *Scale validation data of protocol 2029-601-000*
    c. *35% as specified for the sample size estimation*

    Due to the non-interventional nature, participants who were rated as moderate or severe (ATNLS Grade 2 or 3) in Session 1 of the rater qualification assessment of protocol 2029-701-008 and the scale validation study 2029-601-008 can be likened to the no-treatment control participants of this pivotal study as indicated in inclusion criteria 3.03 of the study protocol. Thus, the proportion of participants with at least 1-grade improvement in ATNLS from Session 1 to Session 2 (referred as ATNLS responder rate) and the corresponding 95% CI will be

calculated using the source data a and b listed above. $\hat{P}_c$ will be estimated by the median of the following 3 values:

   a. Upper bound of the 95% CI of the ATNLS responder rate based on the cumulative rater qualification data of the pivotal trial
   b. Upper bound of the 95% CI of the ATNLS responder rate of the scale validation study 2029-601-000
   c. 35% as specified for the sample size estimation

3. The ATNLS responder rate for the treatment group, $\hat{P}_v$, will be calculated based on the estimates obtained at steps No. 1 and 2 and the randomization ratio of 2:1 using the following formula.

$$\hat{P}_v = \frac{3}{2}\hat{P} - \frac{1}{2}\hat{P}_c$$

4. The probability to reject both hypothesis testing for the primary effectiveness endpoint will be calculated via a simulation based on the estimates obtained at steps 2 and 3.
5. If the probability estimated at step 4 is less than 80%, then the sample size will be increased to achieve at least 80% power for the analyses of primary effectiveness endpoint.

If the estimated sample size at step 5 is greater than the maximum sample size allowed for this study, 200 randomized participants, then the study will remain unchanged with no additional participants to be randomized (159 randomized participants). The study will conclude after the last participant completes the last visit or discontinue from the trial.

3.0 Endpoints

3.1 Primary Effectiveness Endpoint

The primary endpoint is the responder status based on the EI's live assessment of transverse neck lines using ATNLS at Month 1 after the last treatment (initial or touch-up)

for treatment group and after randomization for control group, where a "responder" is a participant with at least 1-grade improvement from baseline on the ATNLS.

3.2 Secondary Effectiveness Endpoints

Secondary endpoints are:

- Responder status based on the EI's assessments of global aesthetic improvement on the neck using the GAIS at Month 1
- Change from baseline to Month 1 on FACE-Q Appraisal of the Neck questionnaire
- Responder status based on participant's assessment of global aesthetic improvement on the neck using the GAIS at Month 1

3.3 Other Effectiveness Endpoints

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

- Responder status based on the EI's live assessment of transverse neck lines using ATNLS at other applicable visits
- Responder status based on the EI's assessment of the neck using GAIS at other applicable visits
- Responder status based on participant's assessment of the neck using GAIS at other applicable visits
- Change from baseline in participant responses to FACE-Q Appraisal of the Neck questionnaire at other applicable visits
- Participant responses to FACE-Q Appraisal of the Neck questionnaire at each measurement visit
- Percentage of participants showing improvement on the FACE-Q Appraisal of the Neck questionnaire at each measurement visit
- Change from baseline in neck skin hydration measured by MoistureMeter®D instrument at each measurement visit

- Participant responses to questionnaires (ANSS Feel, ANSS Appearance, ANSS Psychosocial, ANTOS Satisfaction with Treatment, ANTOS Recommend) at each measurement visit

3.4 Exploratory Effectiveness Endpoint

- Change from baseline in neck skin topography measured by PRIMOS 3D images at each measurement visit (applicable study sites only)

3.5 Safety Endpoints

The safety endpoints will include:

- Participant assessment of procedural pain after study injection on an 11-point scale ranging from 0 (no pain) to 10 (worst pain imaginable)
- The presence and severity of ISRs
- Snellen visual acuity, confrontational visual fields, and ocular motility assessments by the TI (or designee)
- AEs, including AESIs, as determined by the TI
- Concomitant medications and concurrent procedures

4.0 Analysis Populations

The population sets defined in Table 1 below will be used for the analyses.

Table 1. Analysis Sets

Population	Definition	Study Intervention
Intent-to-Treat (ITT)	All randomized participants.	As randomized
Observed Primary Endpoint (OPE)	All ITT participants with baseline and Month 1 assessments on the ATNLS.	As randomized
Safety	All randomized participants who have at least 1 study intervention (i.e., VOLITE XC or no-treatment)	As treated
VOLITE Treated (VT)	• All randomized participants who receive VOLITE XC treatment at the beginning of the control period • All randomized participants who receive optional VOLITE XC treatment after the control period	As treated
VOLITE Repeat Treatment (VRT)	Participants in the VOLITE treated population who receive repeat treatment	As treated

5.0 Participant Disposition

The study periods are defined below in Table 2 for VOLITE XC group and control group. For VOLITE XC group, the Treated period (TP) is the Control Period (CP) and prior to Repeat Treatment Period. For participants in the control group, the TP starts on the date of optional initial treatment and ends on the same day as the study exit.

Table 2. Definition of the Control and Treated Periods by Analysis Group (as-randomized for Effectiveness or as-treated for Safety)

Analysis Group	Period	Analysis Group Label	Start Date	End Date
VOLITE	CP and TP	VOLITE XC	Latter of initial treatment date or randomization date	Date of repeat treatment or the study exit date, whichever is earlier.
	RTP	VOLITE XC	Repeat treatment date	Study exit date
Control	CP	Control	Randomization date	For participants who do not receive optional treatment, the end date is the study exit date. For participants who receive optional treatment, the end date is the date when the treatment is received.
	TP	VOLITE XC Post-Control	Optional initial treatment date	Study exit date

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 CP:

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

For TP:

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

For RTP:

- Number of participants continued into repeat treatment
- Number of participants completed RTP
- Number of participants discontinued during RTP
- Reasons for discontinuation during RTP

During the study:

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

Treatment group participants who elect not to receive repeat treatment are considered to have completed the study if they complete the Month 6 visit in the CP. Meanwhile, treatment group participants who elect to receive repeat treatment are considered to have completed the study if they complete the Month 4 visit in the RTP. 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 CP. 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 TP.

6.0 Study Device Exposure and Administration

In general, unless otherwise specified, 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 number of participants with observed values or events (n), frequency count (N1), and percentage of participants with observed values or events.

Treatment exposure-related variables will be summarized for the VT and VRT populations. For control group, data after participants received the optional initial treatment at Month 1 are included. Study treatment exposure will be measured by volume injected at each treatment (Table 3) and summarized by treatment group and treatment (initial, touch-up, initial and touch-up combined, repeat).

Table 3. Study Device Exposure

Endpoint	Description	Timing	Methodology
Number (%) of participants received treatment	Summary by treatment group, treatment (initial, touch-up, repeat)	Initial, Touch-up, Repeat	Frequency counts and percentages
Injection volume	Summary by treatment group, treatment (initial, touch-up, initial and touch-up, repeat)	Initial, Touch-up, Initial and Touch-up, Repeat	Continuous descriptive statistics

The number of participants who received treatment anesthesia will be summarized for the VT and VRT population. Variables related to administration of treatment listed in Table 4 will be summarized for the VT population by treatment group and treatment (initial, touch-up, repeat).

Table 4. Administration of Study Treatment

Endpoint	Description	Timing	Methodology
Pre-treatment anesthesia type • Ice • Topical	Summary by treatment group and treatment (initial, touch-up, repeat)	Initial, Touch-up, Repeat	Categorical descriptive statistics
Pre-treatment anesthesia duration (minutes) • Anesthesia duration is computed as the injection time minus start of anesthesia administration time. Summarize by treatment group, treatment, and anesthesia type.	Summary by anesthesia type, treatment group and treatment (initial, touch-up, repeat)	Initial, Touch-up, Repeat	Continuous descriptive statistics
Treatment administration • Needle used (32 Gauge 1/2 inch) • Cannula used (25 Gauge 1½ inch) • Planes of injection[a] (intradermal, subdermal) • Injection technique [a] (fanning, tunneling, microdroplet)	Summary by treatment group, treatment (initial, touch-up, repeat)	Initial, Touch-up, Repeat	Categorical and continuous descriptive statistics
Treatment characteristics • Injection ease (11-point scale) • Product moldability (11-point scale)	Summary by treatment group and treatment (initial, touch-up, repeat)	Initial, Touch-up, Repeat	Categorical and continuous descriptive statistics

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

A listing will be provided for study device exposure, treatment administration, and treatment characteristics.

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

Demographics, baseline characteristics, medical history, and prior and concomitant medications will be summarized overall and by treatment group. Categorical variables will be summarized with the number and percentage of participants; percentages will be calculated based on the number of non-missing observations. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum and maximum).

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 ITT population.

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

- Weight (kg), height (cm), and BMI (kg/m^2)
- Fitzpatrick skin phototype (each phototype, and by phototype groups: I and II, III and IV, V and VI)
- Exposure to sunlight (hours per day)
- Alcohol use history (status, amount/frequency, and duration in years)
- Nicotine use history (status, amount/frequency, nicotine product, and duration in years)
- Allergan Transverse Neck Lines Scale (ATNLS)

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 initial treatment date in the CP. Concomitant medication is defined as any medication taken on or after the latter of randomization date or initial treatment date in the CP. Listing will be provided for the safety population. The Anatomical Therapeutic Chemical 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

- For CP, effectiveness analyses and baseline characteristics will be performed on the ITT population.
- For TP, effectiveness analysis will be performed on the VT population.
- Day 1 is defined as the day on which the study intervention (VOLITE XC or no-treatment control) is first received (i.e., the latter of randomization or initial treatment in CP).
- Baseline for effectiveness parameters will be the last non-missing assessment prior to or on Day 1.
- Study day is always relative to Day 1 regardless of study period (e.g., CP, TP, RTP).
- Change from baseline will be computed as the post-baseline minus the baseline.
- 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 higher.

8.2 Handling of Missing Data

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

- Multiple Imputation (MI): The missing ATNLS score at Month 1 in CP will be imputed using SAS PROC MI with Fully Conditional Specification (FCS) method separately by treatment group.
- Multiple Imputation (MI): The missing GAIS score at Month 1 in CP will be imputed using SAS PROC MI with Fully Conditional Specification (FCS) method separately by treatment group.
- Multiple Imputation (MI): The missing FACE-Q score at Month 1 in CP will be imputed using SAS PROC MI with Fully Conditional Specification (FCS) method separately by treatment group.
- Non-Responder Imputation (NRI): the NRI method will categorize any participants with missing Month 1 ATNLS scores as non-responders.
- As Observed (AO): The AO method will not impute values for missing ATNLS scores, thus, a participant who does not have an ATNLS score at Month 1 will be excluded from the AO analysis.

8.3 Primary Effectiveness Endpoint and Analyses

8.3.1 Primary Effectiveness Endpoint

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

improvement (reduction) from baseline on the ATNLS (described in Table 5) on the neck area.

Table 5. Allergan Transverse Neck Lines Scale

Score	Grade	Description
0	None	No transverse neck lines
1	Minimal	Superficial transverse neck lines
2	Moderate	Moderate, effaceable transverse neck lines
3	Severe	Deep, non-effaceable transverse neck lines
4	Extreme	Non-effaceable transverse neck furrows with redundant skin

8.3.2 Handling of Missing Data for the Primary Effectiveness Endpoint

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

Missing values will be imputed using the multiple imputation method. After imputing 30 data sets, change from baseline in ATNLS at Month 1 will be calculated, from which the responder status will be determined. By pooling the results from the 30 imputed data sets using PROC MIANALYZE, the following statistics will be calculated: the pooled responder rates and corresponding 2-sided 95% CI in the VOLITE XC group and in the control group, the pooled between-group difference in responder rates, and the corresponding 2-sided 95% CI and p-value.

Specifically, the missing ATNLS score at Month 1 in CP will be imputed as follows:

Step 1: The missing data will be imputed by treatment group using the FCS method for ordinal data (via SAS procedure MI with FCS logistic statement along with the FIRTH option). Thirty (30) imputed datasets will be generated, using *seed 2029701* and *nimpute 30* based on the following model:

Month 1 ATNLS $\beta 0 + \beta 1 \times$ Age $+ \beta 2 \times$ Sex $+ \beta 3 \times$ Fitzpatrick skin type $+ \beta 4 \times$ Body mass index $+ \beta 5$ Site $+ \beta 6 \times$ Baseline ATNLS

If complete or quasi-complete separation data issues arise from the above procedure, the LIKELIHOOD AUGMENT option will be implemented in the FCS logistic statement to address the issue.

Step 2: After imputation, change from baseline will be calculated, from which the responder status in ATNLS at Month 1 of CP will be determined.

Step 3: Each complete (imputed) dataset obtained from Steps 1 and 2 will be analyzed using SAS procedure FREQ where the standard error for the responder rate within each treatment group and between-group difference will be estimated based on the normal approximation.

Step 4: SAS PROC MIANALYZE will be used to generate the final inferences of the between-group difference in the responder rate.

8.3.3 Primary Effectiveness Analysis

The primary effectiveness analysis will be performed based on the ITT population. There are 2 hypotheses testing for the primary effectiveness endpoint, ATNLS responder status at Month 1 in CP.

Hypothesis 1: The null hypothesis is that the ATNLS responder rate for the VOLITE XC group is equal to the responder rate for the control group. The alternative hypothesis is that the responder rates are not equal for the treatment group and the control group. These hypotheses are stated as:

$$H_0: P_v \quad P_c$$

$$H_a: P_v \neq P_c$$

where P_v and P_c denote the responder rates for the VOLITE XC group at Month 1 after the last treatment (initial or touch-up) and for the control group at Month 1 after randomization, respectively.

Hypothesis 2: The null hypothesis is that the ATNLS responder rate for the treatment group is less than or equal to 50%, and the alternative hypothesis is that the responder rate for the treatment group is greater than 50%. These hypotheses are stated as follows:

$$H_0: P_v \leq 50\%$$

$$Ha: P_v > 50\%$$

If there are no ITT participants with missing primary effective endpoint, the first hypothesis will be tested via a 2-sided Fisher's exact test at the 0.05 level to compare responder rates between VOLITE XC and control groups. The second hypothesis will be tested against 50% by the 2-sided 95% CI based on the Clopper-Pearson exact binomial method.

If there are ITT participants with missing primary effective endpoint, the first and second hypotheses will be tested based on the method and derived statistics described in Section 8.3.2. If the resulting 2-sided p-value for the between-group comparison is ≤ 0.05 and the ATNLS responder rate of VOLITE XC group is greater than that of the control group, then VOLITE XC will be considered statistically significantly higher (better) than the control group. Furthermore, if the lower bound of the 2-side 95% CI for VOLITE XC group is > 50%, then the ATNLS responder rate for VOLITE XC group at Month 1 will be considered statistically significantly greater (better) than 50%.

If both hypotheses are rejected, then VOLITE XC will be considered clinically effective.

8.3.4 Additional Analyses of the Primary Effectiveness Endpoints

If there are any ITT participants with missing scores for the primary effectiveness endpoint, a sensitivity analysis of the primary effectiveness endpoint will be performed based on 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 ITT population with missing ATNLS responder status at Month 1 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
ATNLS responder status at Month 1	1	AO method	Observed Primary Endpoint	Fisher's exact test
ATNLS responder status at Month 1	2	NRI method	ITT	Fisher's exact test

The poolability of site for the primary effectiveness analysis will be evaluated by including investigational site, treatment, and the interaction of site with treatment as predictor variables in a multivariate logistic regression analysis along with the FIRTH option, with responder status (1 yes and 0 no) on the ATNLS at Month 1 as the dependent variable. The interaction between site and treatment will be considered significant if the p-value of this interaction is less than 0.15. Investigational sites with less than 6 participants in the ITT population will be excluded.

The primary effectiveness endpoint will also be summarized using descriptive statistics by site regardless of the sample size for each site.

8.4 Secondary Effectiveness Endpoints and Analyses

8.4.1 Secondary Effectiveness Endpoints

The secondary effectiveness endpoints will be analyzed based on ITT population. The missing data of secondary effectiveness endpoints with hypothesis testing will be imputed and analyzed by the multiple imputation method described in Section 8.4.2.

The secondary effectiveness endpoints include:

- Responder status based on EI's assessments of global aesthetic improvement on the neck using the GAIS (described in Table 7) at Month 1 in CP
- Change from baseline to Month 1 in CP on FACE-Q Appraisal of the Neck questionnaire
- Responder status based on participant's assessments of global aesthetic improvement on the neck using the GAIS (described in Table 7) at Month 1 in CP

A GAIS responder for EI's assessment is defined as a participant who shows improvement in the overall aesthetic assessment (Improved or Much Improved on GAIS) based on EI's assessment on the neck at Month 1.

A GAIS responder for participant's assessment is defined as a participant who shows improvement in the overall aesthetic assessment (Improved or Much Improved on GAIS) based on participant's assessment on the neck at Month 1.

Table 7. Global Aesthetic Improvement Scale

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

The other secondary effectiveness measure is the participant response on the FACE-Q Appraisal of the Neck questionnaire.

For FACE-Q Appraisal of the Neck, the sum of the items will be calculated and converted into a 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., number of items representing at least 5 have been answered) then insert the unrounded mean of the completed items into the total sum score. Otherwise, the Rasch-transformed 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 algorithm developed by the FACE-Q scale developer.

8.4.2 Secondary Effectiveness Analyses

The analysis for responder status based on EI assessments of GAIS between VOLITE XC and control groups at Month 1

The corresponding null and alternative hypotheses are:

$$H_0: P_v \quad P_c$$

$$H_a: P_v \neq P_c$$

where P_v and P_c denote the GAIS responder rates based on EI assessments for the VOLITE XC group at Month 1 after the last treatment (initial or touch-up) and for the control group at Month 1 after randomization, respectively.

If there are no ITT participants with missing GAIS by EI assessment at Month 1, the difference in the responder rate of GAIS by EI assessment between VOLITE XC and control will be tested using a 2-sided Fisher's exact test at the 0.05 level.

If there are ITT participants with missing GAIS by EI assessment at Month 1, missing values will be imputed using the multiple imputation method as described below:

Step 1: The missing data will be imputed by treatment group using the FCS method for ordinal data (via SAS procedure MI with FCS logistic statement along with the FIRTH option). Thirty (30) imputed datasets will be generated, using *seed 20297012* and *nimpute 30* based on the following model:

Month 1 GAIS by EI assessment $\beta 0 + \beta 1 \times$ Age $+ \beta 2 \times$ Sex $+ \beta 3 \times$ Fitzpatrick skin type $+ \beta 4 \times$ Body mass index $+ \beta 5$ Site $+ \beta 6 \times$ Baseline ATNLS

If complete or quasi-complete separation data issues arise from the above procedure, the LIKELIHOOD AUGMENT option will be implemented in the FCS logistic statement to address the issue.

Step 2: After imputation, change from baseline will be calculated, from which the responder status in GAIS by EI assessment at Month 1 of CP will be determined.

Step 3: Each complete (imputed) dataset obtained from Steps 1 and 2 will be analyzed using SAS procedure FREQ where the standard error for the responder rate within each treatment group and between-group difference will be estimated based on the normal approximation.

Step 4: SAS PROC MIANALYZE will be used to generate the final inferences of the between-group difference in the responder rate.

The pooled responder rates for GAIS by EI and the corresponding 2-sided 95% CI will be tabulated separately by treatment group. The pooled between-group difference in GAIS by EI responder rates, and the corresponding 2-sided 95% CI and p-value will be summarized. If the 2-sided p-value for the between-group comparison is ≤ 0.05 and the GAIS by EI responder rate of VOLITE XC group is greater than that of the control group, then VOLITE XC will be considered statistically significantly higher (better) than the control group.

The analysis for change from baseline in the Rasch-transformed score of FACE-Q Appraisal of the Neck within VOLITE XC group at Month 1

The corresponding null and alternative hypotheses are:

$$H_0: \mu_v \leq 0$$

$$H_a: \mu_v > 0$$

where μ_v denotes the mean change from baseline in the Rasch-transformed score of FACE-Q Appraisal of the Neck questionnaire for the VOLITE XC group at Month 1 after the last treatment (initial or touch-up).

If there are no ITT participants within VOLITE XC group with missing Rasch-transformed score of FACE-Q Appraisal of the Neck at Month 1, the within-group change from baseline in the Rasch-transformed score of FACE-Q Appraisal of the Neck at Month 1 will be tested for VOLITE XC using a 1-sided paired t-test at the 0.025 level.

If there are ITT participants within VOLITE XC group with missing Rasch-transformed score of FACE-Q Appraisal of the Neck at Month 1, missing values will be imputed using the multiple imputation method as described below.

Step 1: The missing data will be imputed by treatment group. Thirty (30) imputed datasets will be generated, using *seed 20297013* and *nimpute 30* based on the following model:

Month 1 FACE-Q $\beta 0 + \beta 1 \times$ Age $+ \beta 2 \times$ Sex $+ \beta 3 \times$ Fitzpatrick skin type $+ \beta 4 \times$ Body mass index $+ \beta 5 \times$ Site $+ \beta 6 \times$ Baseline ATNLS $+ \beta 7 \times$ Baseline FACE-Q

If complete or quasi-complete separation data issues arise from the above procedure, the LIKELIHOOD AUGMENT option will be implemented in the FCS regression statement to address the issue.

Step 2: After imputation, change from baseline will be calculated.

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

Step 4: SAS PROC MIANALYZE will be used to generate the final inferences of the within-group change from baseline value for the VOLITE XC group and between-group difference in the change from baseline values.

The pooled change from baseline values in the Rasch-transformed score of FACE-Q Appraisal of the Neck, and the corresponding 2-sided 95% CI and p-value will be summarized for the VOLITE XC group. If the 1-sided p-value is ≤ 0.025, then the Rasch-transformed score in FACE-Q Appraisal of the Neck at Month 1 will be considered statistically significantly greater (better) than that at baseline for the VOLITE XC group.

The analysis for change from baseline in the Rasch-transformed score of FACE-Q Appraisal of the Neck between VOLITE XC and control groups at Month 1

The corresponding null and alternative hypotheses are:

$$H_0: \mu_v \quad \mu_c$$

$$H_a: \mu_v \neq \mu_c$$

where μ_v and μ_c denote the mean change from baseline in the Rasch-transformed score of FACE-Q Appraisal of the Neck questionnaire for the VOLITE XC group at Month 1 after the last treatment (initial or touch-up) and for the control group at Month 1 after randomization, respectively.

If there are no ITT participants with missing Rasch-transformed score of FACE-Q Appraisal of the Neck at Month 1, the change from baseline to Month 1 in the Rasch-transformed score of FACE-Q Appraisal of the Neck for VOLITE XC group will be tested against that of the control group using 2-sample t-test at the 0.05 level.

If there are ITT participants within VOLITE XC group with missing Rasch-transformed score of FACE-Q Appraisal of the Neck at Month 1, missing values will be imputed using the multiple imputation method as described above under “The analysis for change from

baseline in the Rasch-transformed score of FACE-Q Appraisal of the Neck within VOLITE XC group at Month 1".

The pooled change from baseline values in the Rasch-transformed score of FACE-Q Appraisal of the Neck and the corresponding 2-sided 95% CI will be tabulated separately by treatment group. The pooled between-group difference in change from baseline values in the Rasch-transformed score of FACE-Q Appraisal of the Neck, and the corresponding 2-sided 95% CI and p-value will be summarized. If the 2-sided p-value is ≤ 0.05 and the mean change from baseline in Rasch-transformed score of FACE-Q Appraisal of the Neck of VOLITE XC group is greater than that of the control group, then VOLITE XC will be considered statistically significantly greater (better) than the control group.

The summary for responder status based on participant assessments of GAIS for the VOLITE XC and control groups at Month 1

The GAIS responder rate based on the participant assessment at Month 1 and the corresponding 2-sided 95% CI based on the Clopper-Pearson exact binomial method for both VOLITE XC and control groups will be summarized based on observed data. There is no hypothesis testing for the GAIS response rate based on participant assessments at Month 1.

Secondary effectiveness analyses are summarized in Table 8.

Table 8. Secondary Effectiveness Analyses

Endpoint	Description	Missing Data Handling Rule	Methodology
GAIS responder status as assessed by the EI at Month 1	Number (%) of responders by treatment group	Multiple Imputation	Multiple imputation with aggregated statistics for the between-group difference in GAIS by EI responder rate p-value at 0.05 significance level and corresponding 2-sided 95% CI
Change from baseline in Rasch-transformed score of FACE-Q Appraisal of the Neck at Month 1	Mean change from baseline within the treatment group	Multiple Imputation	Multiple imputation with aggregated statistics for the within-group change from baseline of Rasch-transformed score of FACE-Q p-value at 0.025 significance level and corresponding 2-sided 95% CI
Change from baseline in Rasch-transformed score of FACE-Q Appraisal of the Neck at Month 1	Mean change from baseline by treatment group	Multiple Imputation	Multiple imputation with aggregated statistics for the between-group change from baseline of Rasch-transformed score of FACE-Q responder rate p-value at 0.05 significance level and corresponding 2-sided 95% CI
GAIS responder status as assessed by the participant at Month 1	Number (%) of responders by treatment group	AO method	Categorical descriptive statistics

8.5 Other Effectiveness Analyses

The other effectiveness analyses will be performed for all variables listed in Table 9. Analyses on the primary and secondary variables at other visits in CP, TP and RPT will also be performed. All analyses are based on observed data with no imputation for the missing data. All analyses are performed by treatment group and by visit. Additionally, no hypotheses will be tested and, thus, no multiplicity adjustment will be performed.

For CP, the ITT population will be used, including Month 1.

For TP, VT population will be used. VOLITE XC group will include Month 1, 2, 4, 6 visits and control group will include Month 2, 4, 6 visits.

For RTP, the RT population will be used, including Month 1, 2, and 4 visits.

In the responder analysis on FACE-Q Appraisal of the Neck questionnaire, a responder is defined as any participant whose post-baseline Rasch-transformed score is greater than their baseline Rasch-transformed score.

Neck skin hydration measured by the MoistureMeter®D instrument will be converted and analyzed from the originally collected unit of Tissue Dielectric Constant (TDC) to the unit of Percentage of Water Content (PWC) based on the formula below:

$$PWC = \frac{TDC-1}{77.5} * 100\%$$

Neck skin hydration will also be summarized excluding all participants from Site 108 due to measurement error in the use of the instrument.

Allergan Neck Satisfaction Scale (ANSS) Overall Scoring

For ANSS Feel, ANSS Appearance, and ANSS Psychosocial scales, an overall score will be calculated and summarized for each domain, ranging from 0 (worst) to 9 (best) for Feel, 0 (worst) to 15 (best) for Appearance, and 0 (worst) to 18 (best) for Psychosocial with higher values representing better values. If there are any missing items (which

includes selection of more than 1 response for an item, a response of non-applicable (N/A) or skipping a question) in any of the ANSS questionnaire, the overall score for that ANSS questionnaire will be considered missing.

Overall scores for each domain will be calculated as follows:

ANSS Feel Domain:
Step 1: Convert scores to numeric values as "Very dissatisfied" 0, "Dissatisfied" 1, "Satisfied" 2, "Very satisfied" 3.

Step 2: Calculate Overall Score as the sum of questions 1 through 3. Higher scores reflect a better outcome. If any of the components are missing, ANSS Feel Overall Score will be missing.

ANSS Appearance Domain:
Step 1: Convert scores to numeric values as "Very dissatisfied" 0, "Dissatisfied" 1, "Satisfied" 2, "Very satisfied" 3.

Step 2: Calculate Overall Score as the sum of questions 1 through 5. Higher scores reflect a better outcome. If any of the components are missing, ANSS Appearance Overall Score will be missing.

ANSS Psychosocial Domain:
Step 1: Convert scores to numeric values as "Not at all" 0, "Slightly" 1, "Moderately" 2, "Very" 3 for questions 1 through 3 and 7.

Step 2: Convert scores to numeric values as "Never" 0, "Rarely" 1, "Sometimes" 2, "Often" 3 for question 4.

Step 3: Convert scores to numeric values as "Very" 0, "Moderately" 1, "Slightly" 2, "Never" 3 for question 6.

Step 4: Calculate Overall Score as the sum of questions 1 through 4 and questions 6 through 7. Higher scores reflect a better outcome. If any of the components are missing, ANSS Psychosocial Overall Score will be missing.

Table 9. Other Effectiveness Analyses

Endpoint	Description	Analysis Period	Methodology
Responder status based on the EI's live assessment of transverse neck lines using ATNLS	Number (%) of responders	CP, TP, RTP	Categorical descriptive statistics
Responder status based on the EI's assessment of the neck using GAIS	Number (%) of responders	CP, TP, RTP	Categorical descriptive statistics
Responder status based on participant's assessment of the neck using GAIS	Number (%) of responders	CP, TP, RTP	Categorical descriptive statistics
Change from baseline in participant responses to FACE-Q Appraisal of the Neck questionnaire	Summary by treatment group, visit	CP, TP, RTP	Continuous descriptive statistics
Participant responses to FACE-Q Appraisal of the Neck questionnaire	Number (%) by response category	CP, TP, RTP	Categorical descriptive statistics
Percentage of participants showing improvement on the FACE-Q Appraisal of the Neck questionnaire	Number (%) of responders	CP, TP, RTP	Categorical descriptive statistics
Change from baseline in neck skin hydration (PWC)	Summary by treatment group, visit	CP, TP, RTP	Continuous descriptive statistics
Change from baseline in participant responses to questionnaires (ANSS – Feel, ANSS Appearance, ANSS – Psychosocial)	Summary by treatment group, visit	CP, TP, RTP	Continuous descriptive statistics
Participant responses to questionnaires (ANSS – Feel, ANSS – Appearance, ANSS – Psychosocial, ANTOS – Satisfaction with Treatment, ANTOS – Recommend)	Number (%) by response category	CP (except ANTOS), TP, RTP	Categorical descriptive statistics

8.6 Effectiveness Subgroup Analyses

Analysis of the primary effectiveness endpoint, ATNLS responder status at Month 1, will be done separately by the following subgroups descriptively.

- Baseline ATNLS score (moderate and severe)

- Total injection volume (at initial and touch-up treatments combined) (≤ median and > median)
- Injection with/without cannula (needle only and cannula with or without needle)
- Sex
- Age (≤ median and > median)
- Race
- Fitzpatrick skin phototype (I/II, III/IV, and V/VI)
- Investigational site

9.0 Safety Analyses

9.1 General Considerations

- For CP, safety summaries will be performed on the safety population.
- For TP, safety summaries will be performed on the VT population.
- For RTP, safety summaries will be performed on the VRT population.
- Adverse Events (AEs) and medical history will be coded using MedDRA. The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report.

9.2 Adverse Events

Adverse events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) according to the version of the MedDRA coding dictionary used for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each participant for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a participant, the highest severity and level of relationship to investigational product will be reported.

9.2.1 Treatment-Emergent Adverse Events

Treatment-emergent AEs are defined as any AE with the onset on or after the date (and time, if known) of the first study treatment for treatment group or of randomization for control group (participants are grouped as treated). Events where the onset date is the same as the study treatment date are assumed to be treatment-emergent unless the AE onset time is recorded and is earlier than the study treatment administration time. All treatment-emergent AEs will be summarized overall, as well as by primary MedDRA SOC and Preferred Term. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

The number and percentage of participants experiencing treatment-emergent AEs will be summarized.

An AE will be considered a TESAE if it is a TEAE that additionally meets any serious adverse event (SAE) criterion.

9.2.2 Adverse Event Overview

A TEAE will be considered a treatment-related TEAE if the TEAE is deemed related to the procedure or the study device by the TI. Duration, time to onset on/after the most recent treatment, and outcome of treatment-related TEAEs will be summarized by severity. The summary will include incidence rate as well as total number of events.

The summaries will be presented using number and percentage of participants with TEAEs as well as the number of events as described in Table 10.

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.

Any adverse events that occur after obtaining written informed consent but before randomization will be listed, but not summarized.

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

Listings of all AEs, TESAEs, treatment-related TEAEs, TEAEs leading to withdrawal of study treatment, and death will be presented.

Listings for treatment related TEAEs will include severity, time to onset, duration, action taken, and outcome.

Table 10. TEAE Summaries

Endpoint	Description	Timing	Methodology
Overall summary	• Overall summary including the following: o TEAEs o Treatment-related TEAEs o Treatment-related TEAEs at injection site o Treatment-related TEAEs not at injection site o TESAEs o Treatment-related TESAEs o Treatment-related TESAEs at injection site o Treatment-related TESAEs not at injection site o Adverse Events of Special Interest (AESIs) o Treatment-related AESI o TEAEs leading to withdrawal of study treatment o Deaths	CP, TP, RTP	Categorical descriptive statistics, event descriptive statistics
TEAEs	• Overall summary by SOC and PT in descending order • Overall summary by SOC, PT, and maximum severity	CP, TP, RTP	Categorical descriptive statistics, event descriptive statistics
Treatment-related TEAEs	• Overall summary by SOC and PT • Overall summary 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, RTP	Categorical descriptive statistics, event descriptive statistics
TESAEs	Overall summary by SOC and PT if 5 or more participants reported such events.	CP, TP, RTP	Categorical descriptive statistics, event descriptive statistics
AESIs	Overall summary by SOC and PT if 5 or more participants reported such events.	CP, TP, RTP	Categorical descriptive statistics, event descriptive statistics

9.2.3 Treatment-Emergent Adverse Events by SOC and/or PT

As provided in Table 10, TEAEs and treatment-related TEAEs will be tabulated by SOC and PT for CP, TP, and RTP.

If more than one event is coded to the same preferred term for a participant, the participant will be counted only once for that preferred term using the greatest severity for the summary on severity.

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.

9.2.4 Treatment-Emergent SAEs (Including Deaths) and Treatment-Emergent Adverse Events leading to Withdrawal of Study Treatment

TESAEs (including deaths) and TEAEs leading to withdrawal of study treatment will be tabulated by SOC and PT, if there are such AEs in 5 or more participants. TEAEs leading to withdrawal of study treatment are events that prevent participants from receiving further study treatment. Listing will be provided.

9.2.5 Adverse Events of Special Interest

AESI are defined in Appendix B. The incidence of AESI will be summarized separately for CP, TP, and RTP in descending order of PT, if the total number of participants with AESI is at least 5. A listing will also be provided.

9.3 Pregnancy Test

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

9.4 Analysis of Vital Signs

Vital signs (systolic and diastolic blood pressure, heart rate, respiratory rate, and temperature) at baseline (screening) and changes from baseline at each assessment time point will be presented as data listings for each treatment group using the safety population. Weight, height, and BMI at baseline (screening) and changes from baseline at each assessment time point will be presented as data listings for each treatment group using the safety population.

9.5 Injection Site Response

ISRs recorded in participant diaries after each treatment (initial, touch-up, and repeat) will be summarized at the participant level for the VT and VRT populations. Descriptive statistics will be provided for the endpoints listed in Table 11 for the initial treatment and touch-up and repeat treatments by predefined symptoms.

For participants who received repeat treatment, their ISRs reported during the initial treatment and touch-up treatment will be summarized and presented in parallel with ISRs reported during the repeat 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.

A data listing of ISRs recorded in participant diaries and ongoing ISR e-CRF will also be provided.

Table 11. 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, Repeat	Categorical descriptive statistics
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 and ongoing ISR eCRF (i.e., date of last ISR - date of first ISR + 1.)	Initial, Touch-up, Repeat	Categorical and continuous descriptive statistics
Ongoing ISR	ISRs recorded on Ongoing ISR page		Listing

9.6 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 repeat treatments will be summarized. The summary will be performed for VT and VRT populations as described in Table 12.

Table 12. Procedural Pain Analyses

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

9.7 Safety Subgroup Analyses

Safety subgroup analysis on ISRs (symptom incidence) and treatment-related TEAEs (preferred term incidence) during the TP will be performed for the following subgroups:

- Total injection volume (at initial and touch-up treatments combined) (≤ median and > median)
- Injection with/without cannula (needle only and cannula with or without needle)

9.8 Other Safety Analyses

Other safety endpoints include the line change on Snellen visual acuity, confrontational visual fields, ocular motility, and concomitant medication and concurrent procedures.

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 $10 \times [\log10 (d_{baseline}/20) \quad \log10 (d_{follow\text{-}up}/20)]$

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

$d_{follow\text{-}up}$ 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.

Neurological assessment will be listed.

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

10.0 Other Analyses

Not applicable.

11.0 Interim Analyses

No interim analysis is planned for this study.

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 5% level of significance. All CIs will be 2-sided 95% CIs, unless stated otherwise.

The blinded sample size re-estimation will be conducted based on the pooled data without unblinding the randomization schedule. Thus, there is no impact to the overall type I error control.

A gatekeeping procedure will be used for hypothesis testing of the primary endpoint (ATNLS responder status at Month 1) and the secondary effectiveness endpoints (GAIS responder status as assessed by the EI at Month 1, change from baseline in Rasch-transformed score of FACE-Q Appraisal of the Neck at Month 1) following a predefined sequence to control the overall type I error rate at the 0.05 level. Both primary hypotheses must be rejected in order to test the secondary effectiveness endpoints in the sequential order specified below:

Order	Category	Endpoint	Description	Methodology

1	Primary	ATNLS responder status at Month 1 of CP	(1) Number (%) of responders' comparison between VOLITE XC and control groups (2) Number (%) of responders within VOLITE XC group.	1) Multiple imputation with aggregated statistics for the between-group difference in ATNLS responder rate p-value at 0.05 significance level (2) Lower bound of 95% CI (obtained from the multiple imputation method) compared with 50%
2	Secondary	GAIS responder status assessed by the EI at Month 1 of CP	Number (%) of responders' comparison between VOLITE XC and control groups	Multiple imputation with aggregated statistics for the between-group difference in GAIS by EI responder rate p-value at 0.05 significance level
3	Secondary	Change from baseline of Rasch-transformed score of participant responses to FACE-Q Appraisal of the Neck questionnaire at Month 1 of CP	Mean change from baseline within VOLITE XC group	Multiple imputation with aggregated statistics for the within-group change from baseline of Rasch-transformed score of FACE-Q p-value at 0.025 significance level

4	Secondary	Change from baseline of Rasch-transformed score of participant responses to FACE-Q Appraisal of the Neck questionnaire at Month 1 of CP	Mean change from baseline comparison between VOLITE XC and control groups	Multiple imputation with aggregated statistics for the between-group change from baseline of Rasch-transformed score of FACE-Q responder rate p-value at 0.05 significance level

13.0 Version History

SAP Version History Summary			
SAP Version	**Approval Date**	**Change**	**Rationale**
0.1	19 November 2021		Initial version
1.0	27 October 2022	• Updated statistical hypothesis structure, sample size determination, and no-treatment control group study design	Per change in Protocol Amendment 3
2.0	28 August 2023	• Updated primary population from mITT to ITT for effectiveness analyses per FDA comment • Updated sections of sample size determination and blinded sample size re-estimation for enhanced clarity • Added site as a predictor to the multiple imputation model for primary effectiveness analysis • Updated “no-treatment control group” to “control group”, updated poolability analysis per FDA comment	Per change in Protocol Amendment 4 and FDA responses
3.0	08 September 2023	• No changes	New version after Protocol Amendment 4 approval date of 08 September 2023.
4.0	07 June 2024	• Updated description of hypothesis testing, missing data imputation for secondary effectiveness endpoints, and sample size re-estimation • Updated FACE-Q questionnaire name, and updated “overall FACE-Q score” to “Rasch-transformed FACE-Q score”	Per change in Protocol Amendment 5 and FDA responses

5.0	07 October 2024	• Added ANSS Overall Scoring summaries • Updated 97.5% one-sided confidence intervals to 95% two-sided confidence intervals for FACE-Q within VOLITE XC group analysis • Added LIKELIHOOD=AUGMENT option to address separation/quasi-separation data issues • Removed Skin Hydration Summaries in units of TDC • Added site as a predictor to the multiple imputation model for FACE-Q analysis • Added sensitivity analysis for neck skin hydration to exclude Site 108 that had measurement error • Updated presentation of adverse events of repeat treatment participants • Updated language of subjects to participants	Per new scoring algorithm of the ANSS questionnaires and to align and maintain consistency with standards.

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 13. If there are participants who are not treated in accordance with randomization, participants will be tabulated as randomized.

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

Table 13. 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	Categorical descriptive statistics

Appendix B. Definition of Adverse Events of Special Interest

The following AESIs have been identified for the study intervention in this protocol: visual disturbance symptoms associated with vascular occlusion (including, but not limited to, any loss of vision, pain in or around eye, blind spot or shadow in the visual field), myocardial infarction, stroke, pulmonary embolism, pulmonary edema, and pneumothorax.

Appendix C. Changes to Protocol-planned Analysis

An ANSS overall score will be calculated and summarized by visit as well as change from baseline by visit for each domain.

Neck skin hydration will be presented with the exclusion of Site 108 in addition to the summary with all data. Site 108 is excluded due to measurement error in the use of the instrument.

Appendix D. List of Abbreviations

Abbreviation	Definition
3D	3-dimensional
AE	adverse event
AESI	adverse events of special interest
AO	as observed
ATNLS	Allergan Transverse Neck Lines Scale
CI	confidence interval
CP	control period
EI	Evaluating Investigator
FCS	fully conditional specification
GAIS	Global Aesthetic Improvement Scale
ISR	injection site response
ITT	intent-to-treat
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
NRI	non-responder imputation
OPE	observed primary endpoint
PI	Principal Investigator
PP	per-protocol
PT	preferred 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	treated period
VT	VOLITE XC treated
VRT	VOLITE XC repeat treatment
WHO	World Health Organization