Title Page

Protocol Title: A multicenter, evaluator-blind, randomized, parallel-group, controlled study of the safety and effectiveness of JUVÉDERM VOLUX™ XC injectable gel for restoring jawline definition

Protocol Number: V25L-002 Amendment 2

Compound Number: JUVÉDERM VOLUX™ XC Injectable Gel

Sponsor Name: Allergan Sales LLC

Legal Registered Address: 2525 Dupont Drive,
Irvine, CA 92612

Regulatory Agency Identifier Number(s)

ClinicalTrials.gov identifier (NCT number): NCT03712137

Table of Contents

List of Tables

List of Figures

SAP Version History

This SAP for study V25L-002 is based on protocol Amendment 2 (approved version dated 11-Mar-2019). The changes are described below:

SAP Version History Summary			
SAP Version	**Approval Date**	**Change**	**Rationale**
1	March 13, 2019	Not Applicable	Original Version
2 (amendment)	April 28, 2020	Template Changed	Changed the template to keep it consistent among all Filler delayed control studies

1. Introduction

This SAP provides a more technical and detailed elaboration of the statistical analyses of the effectiveness and safety data as outlined and specified in study V25L-002 protocol amendment 2 (approved version dated 11-Mar-2019). Specifications of tables, figures, and data listings are contained in a separate document.

1.1. Objectives and Endpoints

Please refer to Sections 3 and 7 of the protocol.

1.2. Study Design

Study V25L-002 is a multicenter, evaluator-blind, randomized, parallel-group, controlled study of the safety and effectiveness of JUVÉDERM VOLUX™ XC injectable gel for jawline restoration. Jawline restoration of each side of the jawline (left and right) will be assessed independently by an EI from cropped 2D [redacted] profile images using the ALJDS, an assessment of loss of jawline definition measured by a 5-point scale (0=None to 4=Extreme). Participants randomized to JUVÉDERM VOLUX™ XC will receive injectable gel (initial treatment) to restore jawline definition, provide a smooth contour, and achieve an aesthetically pleasing result. A touch-up treatment, if needed, will be given 30 days after initial treatment with an optional maintenance treatment available 12 months after the last treatment (initial or touch-up) received. At the initial and touch-up (if performed) treatments, the TI will determine the appropriate volume to inject to achieve at least a 1-grade improvement on the ALJDS based on clinical experience, but the maximum volume will not exceed 8.0 mL for initial and touch-up treatments combined. Injection volume for the maintenance treatment will not exceed 8.0 mL.

The following treatment groups are defined for this study:

- VOLUX: JUVÉDERM VOLUX™ XC injectable gel
- Control: No-treatment Control

Up to 20 investigational sites will enroll up to 280 participants and follow participants who meet the study criteria. Each site will have a blinded EI and a principal investigator who is responsible for the overall conduct of the study at that site and may also be the TI. The TI performs all study treatments, and the blinded EI evaluates ALJDS, GAIS, and facial sensation measurements.

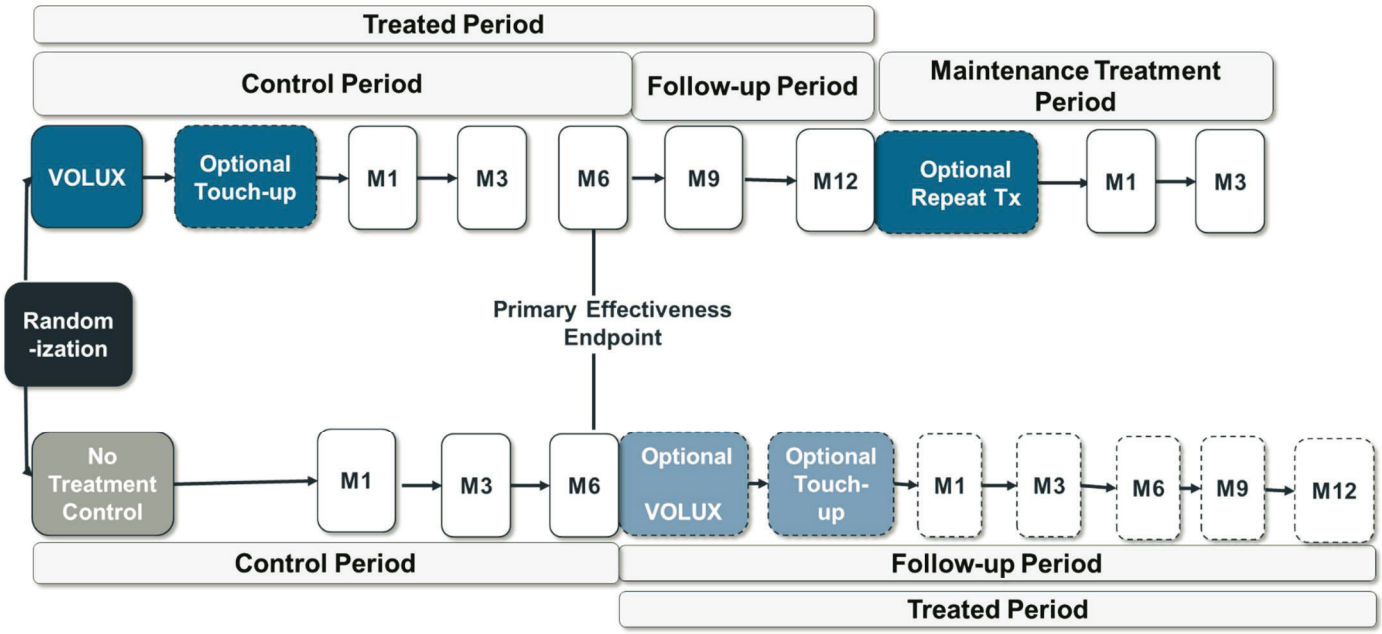
The study design is presented in Figure 1-1. The study will span a total of approximately 26 months: an estimated 6 months for recruitment and 20 months for follow-up. Up to

280 participants will be enrolled, and up to 216 participants will be randomized in a 3:1 ratio to treatment group and control group. For the treatment group, each participant's participation will encompass up to 3 treatments [initial treatment (at Day 1), touch-up treatment (at Day 30) and optional maintenance treatment (at Month 12) and up to 15 months of follow-up after the last treatment (initial or touch-up; 12 months of follow-up for those choosing not to receive optional maintenance treatment at Month 12). For the control group, each participant's participation will encompass a 6-month no-treatment period; for control participants who opt to receive treatment at the completion of the 6-month no-treatment period, the treatment period will be up to 30 days (initial treatment at Day 1, and touch-up at Day 30), and the follow-up period will be up to 6 months for effectiveness and up to 12 months for safety.

Treatment group participants receive initial treatment and optional touch-up treatment at 30 days if the TI judges that an ALJDS score improvement of at least 1 grade has not been achieve. Participants will complete a safety follow-up telephone call 3 days after each treatment (collection of safety data). Follow-up visits will occur at Day 14 after each treatment and at Months 1, 3, 6, 9, and 12 after the last treatment (initial or touch-up). Participants exit the study at Month 12 or receive maintenance treatment and are followed at Days 3 and 14 and at Months 1 and 3 after maintenance treatment.

Control group participants are followed for 6 months after randomization (the 'no-treatment control' period) and then either exit the study or receive optional treatment (initial and touch-up, if needed) in the delayed treatment period. During the 'no-treatment control' period follow-up visits are at Months 1, 3 and 6. After the Month 6 procedures are complete, control group participants will be offered to enter the delayed-treatment period. During the delayed-treatment period, subjects will receive optional treatment (initial with optional touch-up) after which they will be followed for effectiveness for 6 months and for safety for 12 months after the last treatment (initial or touch-up). Control subjects who receive treatment will have the same follow-up visit schedule as the treatment group for 6 months after the last treatment (initial or with touch-up).

Figure 1-1 Study Design



2. Statistical Hypotheses

The primary effectiveness variable is restoration of jawline definition based on the ALJDS evaluated by EIs. The primary effectiveness endpoint is the proportion of ALJDS responders at Month 6. A responder is defined as a participant who shows at least a 1-point improvement from baseline (pretreatment) on the ALJDS on both sides of the jaw.

Month 6 refers to 6 months after the last treatment (initial or touch-up) for participants in the treatment group and 6 months after randomization for participants in the control group.

The study device will be determined to be clinically effective if at least 50% of participants in the treatment group are observed to be responders on ALJDS at Month 6 and if the responder rate for the treatment group is statistically superior to the responder rater for the control group at Month 6. The following hypotheses will be tested:

$$H_o: P_v < 50\%$$
$$H_a: P_v \geq 50\%$$

and

$$H_o: P_v \leq P_c$$
$$H_a: P_v > P_c$$

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

The primary effectiveness analysis will be performed based the mITT population using multiple imputation as described in Section 5.3.2. Missing values will be imputed using multiple imputation method. The pooled responder rates and 2-sided 95% CI for the responder rates in the treatment group and in the control group respectively, the pooled between-group difference in the responder rate and the corresponding 2-sided p-value will be calculated, by pooling the results from the 30 imputed data sets using PROC MIANALYZE with normal approximation in SAS.

The first hypothesis will be tested using the pooled 2-sided 95% CI for the responder rate in the treatment group. The null hypothesis will be rejected if the lower bound of the 95% CI is greater than or equal to 50%.

The second hypothesis will be tested using the pooled between-group difference in the responder rate and the corresponding p-value for the between-group difference. The null hypothesis will be

rejected if the 2-sided p-value is < 0.05 and the responder rate is greater for the treatment group than the control group.

In addition, the 95% CI on the proportion of responders based on the pooled estimate from the 30 imputed data sets will be provided for each treatment group.

3. Sample Size Determination

Sample size is determined to provide adequate power to demonstrate that the product is effective as well as safe. A sample size of 120 participants in the treatment group and 40 participants in the control group will provide > 99% power to detect a difference of at least 40% in the responder rates between the 2 groups. This calculation is based on a 1-sided Fisher's exact test at the 2.5% level. The treatment group is assumed to have at least an 80% responder rate and the control group is assumed to have at most a 40% responder rate. Assuming a screen fail rate of 25% and a drop-out rate of 15% between randomization and the Month 6 visit, up to 280 participants will be enrolled, with at least 144 participants and at most 162 participants randomized to the treatment group and at least 48 participants and at most 54 participants randomized to the control group (3:1 ratio).

This sample size is considered adequate to determine the safety profile of the product in the jawline area. Assuming a screen fail rate of 25%, a 5% drop-out rate between randomization and initial treatment for the treatment group, and a 15% drop-out rate between randomization and optional treatment for the control group, it is estimated that at least 176 subjects will be treated in the study. With 176 treated subjects, there is at least 92% power to detect an AE related to treatment with an incidence rate of 1.5% in the population using a 1-sided exact binomial test at the 2.5% level.

The commercial software, PASS (2008, Version 8.0.13), was used for the power calculation. The sample size calculation used an inequality test for 2 proportions to demonstrate that the treatment group is superior to the control group. The safety calculation used an inequality test for 1 proportion.

4. Populations for Analysis

The 5 analysis populations will consist of participants as defined in Table 4-1.

Table 4-1 Analysis Populations

Population	Definition	Study Treatment Assignment
mITT	All randomized participants with non-missing baseline on the ALJDS scale on both sides of the jaw	As randomized
Observed Primary Endpoint	All mITT participants who have Month 6 assessment on the ALJDS scale on both sides of the jaw	As randomized
Safety	All randomized participants who receive study treatment (VOLUX or control)	As treated[a]
VOLUX Treated	! Participants randomized and received VOLUX treatment at the beginning of the Control Period[b], or ! Participants randomized and received optional VOLUX treatment after the Control Period	As treated
VOLUX Maintenance Treatment	Participants in the VT population who received VOLUX maintenance treatment	As treated

[a] If a randomized control participant does not receive the optional treatment, the actual received treatment will be assigned as control treatment.
[b] Control Period is defined in Section 5.1.

5. Statistical Analyses

5.1. General Considerations

- The primary analysis will be performed after the database is locked and randomization schedule is released.
- The Control Period (CP), Treated Period (TP) and Maintenance Treatment Period (MP) are defined in Table 5-1. Note, for participants randomized to the treatment group TP is period between initial treatment and maintenance treatment, which includes CP plus the follow-up period. For participants randomized to the control group TP is the follow-up period once the participants receive optional treatment. Follow-up Period (FP) starts on the end date of CP and ends on the same day as TP.

Table 5-1 Definition of the Control, Treatment and Maintenance Treatment Periods by Randomization Group

Randomization Group	Period	Treatment Group Label	Start Date	End Date
Treatment	CP	VOLUX	Initial treatment date	The end date is the date of the Month 6 assessment: ! If multiple assessments fall into the Month 6 window, the date of the assessment included in the Month 6 analysis will be used as end date. ! For participants who exit before Month 6, the end date is study exit date. ! For participants with a missing assessment at Month 6, the end date is the target day for Month 6 (181 days from preceding treatment).
	TP	VOLUX	Initial treatment date	Date of maintenance treatment or the study exit date, whichever is earlier
	MP	VOLUX	Maintenance treatment date	Study exit date
Control	CP	Control	Randomization date	For participants who do not receive optional treatment: ! If multiple assessments fall into the Month 6 window, the date of the assessment included in the Month 6 analysis will be used as end date. ! For participants who exit before Month 6, the end date is study exit date. ! For participants with missing assessment at Month 6, the target day for Month 6 (181 days from randomization date) will be the end date. ! For participants who receive optional treatment: the end date is the date of when the optional treatment is received.
	TP	VOLUX-Treated Control	Optional initial treatment date	Study exit date

- For CP, effectiveness analyses and baseline characteristics will be performed on the mITT population using the "as-randomized" assignment.
- For TP, effectiveness and safety summaries will be performed on the VT population using as-treated grouping.

- For MP, effectiveness and safety summaries will be performed on the VMT population.
- For applicable effectiveness and safety endpoints, baseline is defined as the last non-missing value on or before the latter of randomization date or first treatment date in the control period, unless otherwise specified.
- Descriptive statistics for continuous variables include the sample size (N), number of non-missing observations (n), mean, standard deviation, median, 1st and 3rd quartiles, minimum, and maximum.For continuous data, a 2-sample t-test or Wilcoxon rank test will be used for between group comparisons as appropriate, unless stated otherwise. Specified that 95% CI for continuous variables will be based on the standard normal distribution in Section 5.1.
- Summary statistics for categorical variables include the sample size (N1), frequency count, and percentage. Ordinal categorical variables will be analyzed using the Cochran-Mantel-Haenszel test. Nominal variables will be analyzed using Pearson's chi-squared test or Fisher's exact text as appropriate. The 95% CI will be computed using the exact binomial method.
- The level of significance used for all statistical tests will be 0.05, 2-sided, unless stated otherwise.
- The change from baseline values will be computed as the value for the post baseline visit minus the baseline value, unless otherwise indicated.
- In the event if any data collected was after exit, it will not be included in tables but will only be presented in the listings.
- All statistical analysis will be performed using SAS version 9.3 or higher.
- MedDRA 22.1 or newer will be used to code AEs and medical history.
- WHO Drug Dictionary 201703 or newer will be used to code medications.

5.2. Participant Dispositions

The number of participants in each of the 4 study populations (mITT, Safety, VT, and VMT) will be summarized by treatment group.

The number of participants screened will be summarized. Screen-failure participants (ie, participants who are screened but are not randomized) and the associated reasons for failure as recorded in the eCRF will be listed for the all screened participants.

Study disposition will be summarized for all screened participants overall and by treatment group for the following:

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

- Number of participants who continued after CP
- Number of participants who completed FP
- Number of participants who discontinued during FP
- Reasons for discontinuation during FP

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

A participant is considered to have completed the study if he/she has completed the study exit visit. Treatment group participants who elect not to receive maintenance treatment are considered completers if they complete the Month 12 visit. Control group participants who do not elect to receive optional treatment are considered completers if they complete the Month 6 visit. The end of the study is defined as the date of the last visit of the last participant in the study.

5.3. Primary Effectiveness Endpoint Analysis

5.3.1. Definition of Endpoint

The primary effectiveness endpoint is the ALJDS responder status at Month 6 in CP based on EI's assessment. A responder is defined as a participant who achieves at least 1-point improvement (reduction) from baseline on the photonumeric ALJDS (described in Table 5-2) on both sides of the jaw.

Table 5-2 Allergan Loss of Jawline Definition Scale

Score	Grade	Description
0	None	[redacted]
1	Mild	
2	Moderate	
3	Severe	
4	Extreme	

5.3.2. Main Analytical Approach

Two hypotheses regarding the proportion of ALJDS responders at Month 6 based on EI's assessment will be tested.

1) H_o: $P_v < 50\%$
   H_a: $P_v \geq 50\%$

2) H_o: $P_v \leq P_c$
   H_a: $P_v > P_c$

   P_v denotes the responder rate for the treatment group at Month 6 after the preceding treatment and P_c denotes the responder rate for the control group at Month 6 after randomization.

The 2 hypotheses will be tested using the multiple imputation method as follows:

Step 1: The missing data will be imputed by each side of the jaw and treatment group using the fully conditional specification method with seed = 3575309 and nimpute = 30 based on the following model:

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

Step 2: After imputation, the changes from baseline will be calculated, from which the responder status in ALJDS at Month 6 of CP will be determined.

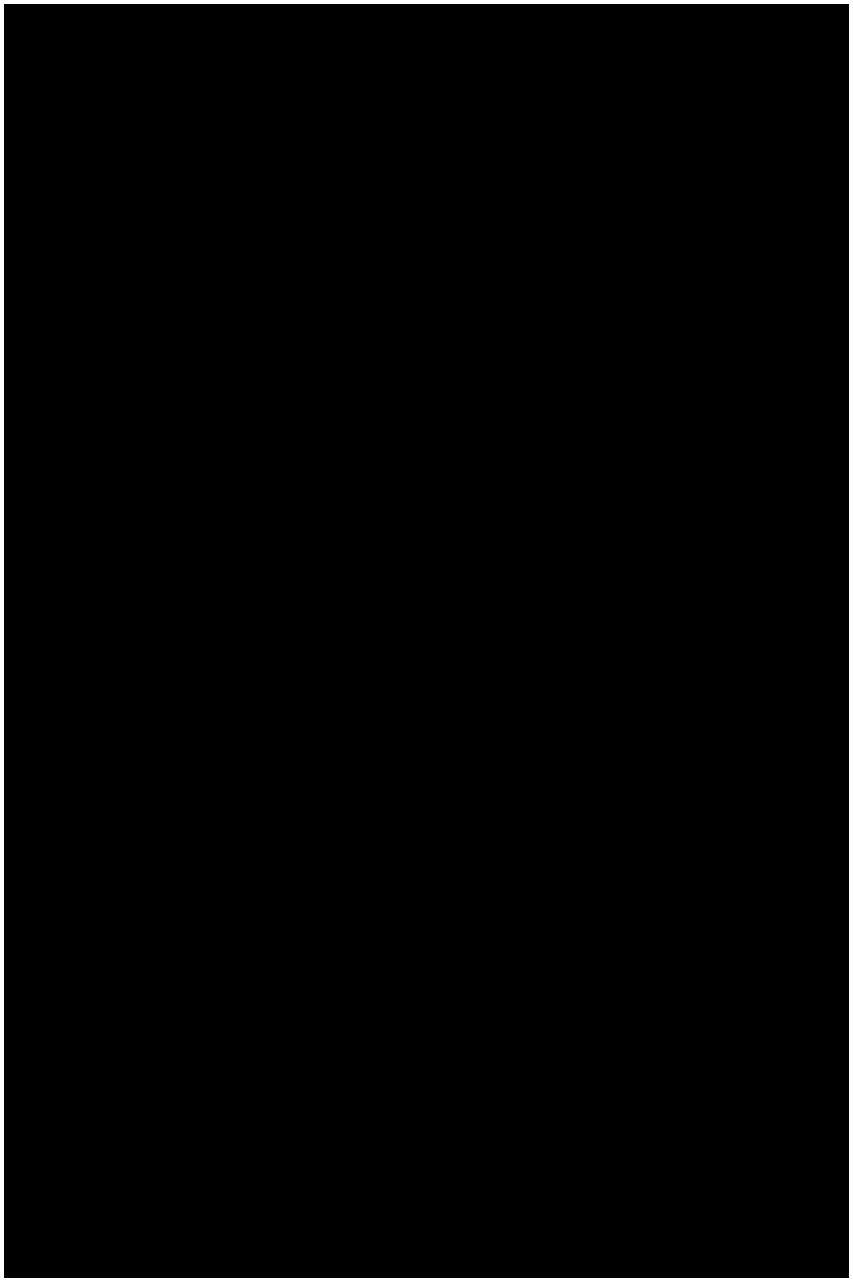
Step 3: Each complete dataset obtained from Steps 1 and 2 will be analyzed using SAS procedure FREQ

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

The number and proportion of participants who achieved at least 1-point improvement from baseline at Month 6 during CP will be tabulated by treatment group. The pooled responder rates by treatment group and the associated p-value and 95% CI estimates based on the normal approximation will be presented. For the first hypothesis, the null hypothesis will be rejected is the lower bound of the 95% CI on the responder rate at Month 6 for VOLUX is greater than or equal to 50%. For the second hypothesis, if the 2-sided p-value is less than 0.05, then the ALJDS responder rate at Month 6 in JUVÉDERM VOLUX XC is statistically significantly different

from that in the no-treatment control group during the control period. Furthermore, if the responder rate is greater for JUVÉDERM VOLUX XC than the no-treatment control group, then VOLUX will be considered superior to the no-treatment control group.

The detailed SAS codes are as follows:



[redacted]

5.3.3. Sensitivity Analyses

Two sensitivity analyses of the primary effectiveness endpoint will also be performed based on the mITT population.

A summary of sensitivity analyses is presented in Table 5-3.

Table 5-3 Missing Data Handling Rules for the Sensitivity Analyses of the Primary Endpoint

Endpoint	Sensitivity Analysis No.	Missing Data Handling Rule	Population	Methodology
ALJDS responder rate at Month 6	1	Use observed data only	Observed Primary Endpoint	Fisher's exact test
ALJDS responder rate at Month 6	2	Multiple Imputation	mITT excluding participants who were inadvertently unblinded to the EI or who did not have a baseline ALJDS score of 2 or 3	Multiple Imputation

5.3.4. Supplementary Analyses

Descriptive statistics for the primary endpoint will be summarized for the mITT population during the control period (ie, Month 1, 3, and 6 after study treatment for the treatment group and after randomization for the control group). Additionally, descriptive statistics for the primary endpoint will be summarized for the VT and VMT populations, which includes all visits after study treatment (ie, Month 1, 3, 6, 9 and 12 for the VT population and Month 1 and 3 after maintenance treatment for the VMT).

5.4. Secondary Effectiveness Endpoints Analysis

5.4.1. Secondary Effectiveness Endpoints

The secondary effectiveness endpoints are (1) GAIS responder at Month 6 as assessed by the EI, (2) GAIS responder at Month 6 as assessed by the participant, and (3) change from baseline in

Rasch transformed score of FACE-Q Satisfaction with Lower Face and Jawline at Month 6 during CP.

A GAIS responder is defined as a participant who shows improvement from baseline in the overall aesthetic assessment in jawline area, a rating of improved or much improved on the GAIS scale (Table 5-4).

The GAIS responder rates as assessed by the EI and by the participant at Month 6 will be analyzed using Fisher's exact test.

Table 5-4 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

Each participant's responses to the 5 items on the FACE-Q Satisfaction with Lower Face and Jawline questionnaire will be summed and converted to a scale score that ranges from 0 to 100 (higher score indicates more satisfaction) using the algorithm developed by the FACE-Q scale developers (see Section 6.5.1). The analysis will be conducted using observed data (eg, no imputation for missing baseline or Month 6 FACE-Q scale scores).

5.4.2. Main Analytical Approach

For the secondary effectiveness endpoints of GAIS responder status at Month 6 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. These analyses will be performed on the mITT population.

No multiplicity adjustments will be made for the 2 GAIS responder rates since these endpoints are descriptive in nature.

For the secondary effectiveness endpoint of change from baseline in FACE-Q Satisfaction with Lower Face and Jawline overall score at Month 6 for the treatment group in the mITT population, a 2-sided paired t-test at the 5% level will be used to determine if the mean overall satisfaction score at Month 6 is statistically greater than that at baseline. This test will be

performed only if the co-primary endpoints are met. For items with missing data (which includes more than 1 response was selected, not applicable was selected, or a question was skipped), insert the mean of the completed items into the total sum score. If less than 50% of the questions have been answered (ie, number of missing items is 2 or more), then the overall score will be missing. The summary score including the imputation of missing items is rounded to the nearest integer and converted to the Rasch transformed score using the conversion table in Section 6.5.1.

Secondary effectiveness analyses are summarized in Table 5-5.

Table 5-5 Secondary Effectiveness Analyses

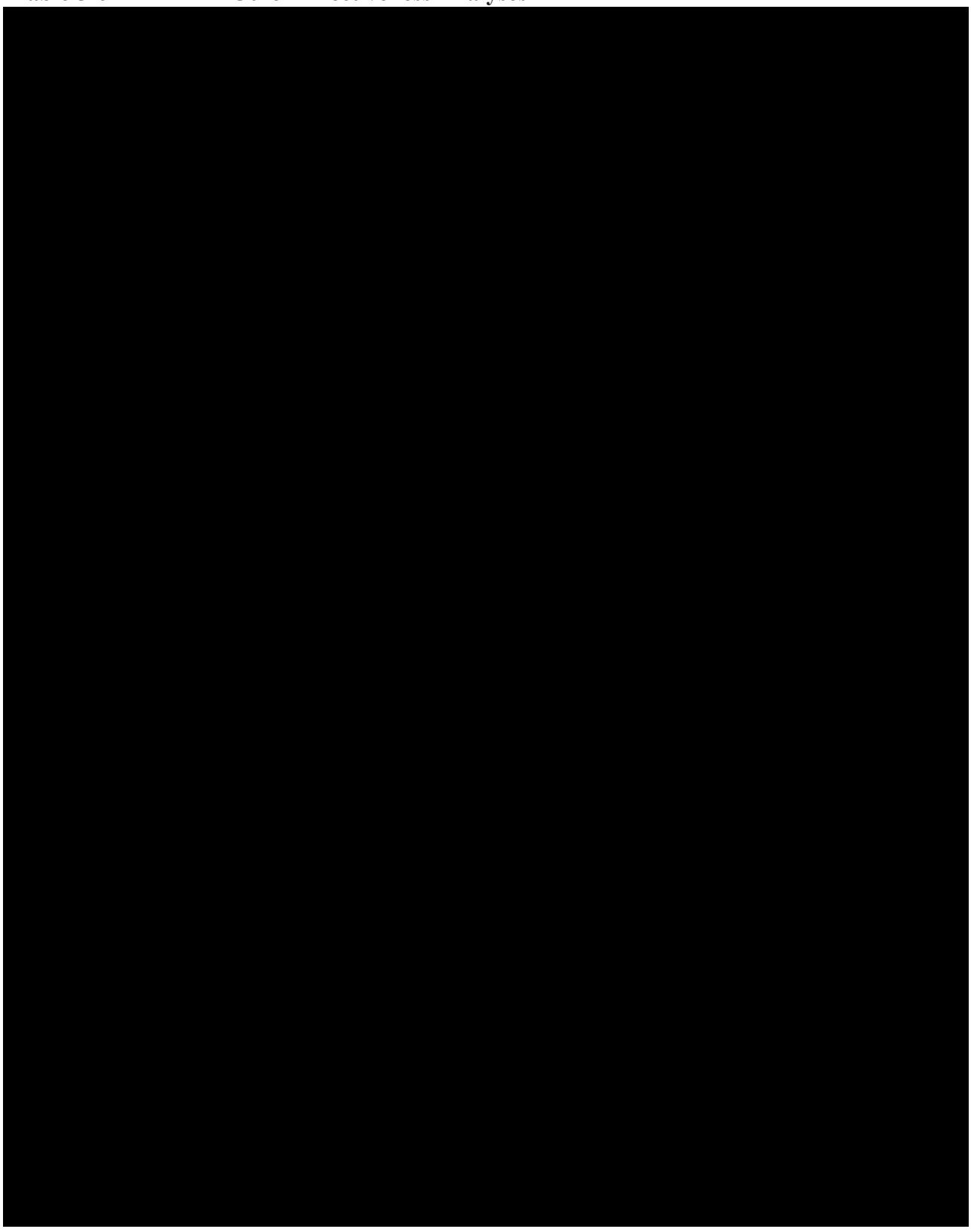
Endpoint	Description	Missing Data Handling Rule	Methodology
GAIS responder status as assessed by the EI at Month 6	Number (%) of responders by treatment group	Observed data is used; no imputation is needed	Categorical descriptive statistics
GAIS responder status as assessed by the participant at Month 6	Number (%) of responders by treatment group	Observed data is used; no imputation is needed	Categorical descriptive statistics
Mean change from baseline in Rasch transformed score of FACE-Q satisfaction with lower face and jawline at Month 6	Mean change from baseline within the treatment group	No imputation. Only observed data will be used in the analysis.	2-sided paired t-test

5.5. Other Effectiveness Endpoints Analysis

[redacted]

[redacted]

Table 5-6 Other Effectiveness Analyses



5.6. Safety Analyses

5.6.1. Extent of Exposure

Treatment exposure-related variables will be summarized for the VT and VMT populations.

Study treatment exposure will be measured by volume injected at each treatment (Table 5-7).

Table 5-7 Exposure to Study Treatment

Endpoint	Description	Timing	Methodology
Volume injected ! Chin ! Right and left pre-jowl area ! Right and left post-jowl area ! Right and left marionette lines ! Right and left angle of the mandible ! Right pre-jowl + post-jowl + marionette + angle of mandible ! Left pre-jowl + post-jowl + marionette + angle of mandible ! Total	Summary by treatment group, treatment (initial, touch-up, maintenance, initial and touch-up combined), and side of jawline	Initial, Touch-up, Maintenance, Initial and Touch-up combined	Continuous descriptive statistics

5.6.2. Administration of Study Treatment

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 5-8 will be summarized for the VT population by treatment group for both sides of the jaw combined at the initial and the touch-up treatments, and for the VMT population for both sides of the jaw combined at the maintenance treatment.

Table 5-8 Administration of Study Treatment

Endpoint	Description	Timing	Methodology
Pre-treatment Anesthesia type ! Ice ! Topical ! Local anesthesia	Summary by treatment group and treatment (initial, touch-up or maintenance)	Initial, Touch-up, Maintenance	Categorical descriptive statistics
Pre-treatment anesthesia duration (minutes) ! Anesthesia duration is computed as injection time minus start of anesthesia administration time. Summarize by treatment group, treatment, and anesthesia type.	Summary by treatment group and treatment (initial, touch-up or maintenance)	Initial, Touch-up, Maintenance	Continuous descriptive statistics
Treatment administration ! Planes of injection (chin, right and left pre-jowl area, right and left post-jowl area, right and left marionette lines, right and left angle of the mandible) ! Injection technique (chin, right and left pre-jowl area, right and left post-jowl area, right and left marionette lines, right and left angle of the mandible) ! Route of administration: needle only or cannula (chin, right and left pre-jowl area, right and left post-jowl area, right and left marionette lines, right and left angle of the mandible) ! Number of needles used ! Device/Needle problem or malfunction	Summary by treatment group and treatment (initial, touch-up or maintenance)	Initial, Touch-up, Maintenance	Categorical descriptive statistics and Continuous descriptive statistics
Treatment Characteristics ! Injection ease ! Product Moldability	Summary by treatment group and treatment (initial or touch-up)	Initial and Touch-up	Categorical descriptive statistics and Continuous descriptive statistics

5.6.3. Adverse Events

An AE will be considered a TEAE if the AE began or worsened (increased in severity or became serious) on or after the date (and time, if known) of the initial treatment for the treatment group and on or after the date of randomization for the control group. An AE will be considered a TESAE if it is a TEAE that additionally meets any SAE criterion.

A TEAE will be considered a treatment-related AE if the TEAE is deemed related to the procedure or the study device by the TI.

An AESI is defined as any vision-related TEAE. The incidence of AESI will be summarized separately for CP, TP, and RP in descending order of PT, if the total number of participants with AESI is at least 5.

Overall summary will be provided on a per-participant level for TEAEs, treatment-related TEAEs, TESAEs, deaths, and TEAEs leading to study intervention discontinuation. If more than one event is coded to the same PT for the same participant, the participant will be counted only once for that PT using the greatest severity for the summarizations by severity and most related occurrence for the summarizations by relationship to study intervention. If the severity of a TEAE is missing, the maximum severity will be assigned to the event for the summarization by 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 summarization. The value will be displayed as missing in the data listing.

TEAEs will be summarized by treatment group for CP using the Safety population. TEAEs will also be summarized for TP and RP using the VT and VRT populations, respectively.

The number and percentage of participants with TEAEs and the events in the following AE categories will be summarized for the above populations as described in Table 5-10. For participants who received maintenance treatment, their AEs reported during the three months in the initial treatment period will be summarized and presented in parallel with AEs reported during the maintenance treatment period.

Summary tables will be provided for participants with TESAEs and participants with TEAEs leading to discontinuation if these occurred in 5 or more participants. Listings of all TEAEs, AESIs, TESAEs, TEAEs leading to discontinuation, and death will be presented.

AEs that occur after obtaining written informed consent but before treatment 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.

Table 5-9 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 TEAE ! Deaths	CP, TP, MP	Categorical descriptive statistics, event descriptive statistics
TEAEs	! Overall summary and by SOC and PT ! Overall summary and by SOC, PT, and maximum severity	CP, TP, MP	Categorical descriptive statistics, event descriptive statistics
Treatment-related TEAEs	! Overall summary and by SOC and PT ! Overall summary and by SOC, PT, and maximum severity ! Overall summary by maximum severity, time to onset, duration, action taken, and outcome	CP, TP, MP	Categorical descriptive statistics, event descriptive statistics
TESAEs	Overall summary and by SOC and PT	CP, TP, MP	Categorical descriptive statistics, event descriptive statistics
AEs leading to discontinuation	Overall summary and by SOC and PT	CP, TP, MP	Categorical descriptive statistics, event descriptive statistics

5.6.4. Injection Site Responses

ISRs recorded in participant diaries after initial and touch-up treatments will be summarized for the VT population, and ISRs recorded after maintenance treatment will be summarized for the VMT population. The diary date and time will be compared with the treatment injection date and time to determine the corresponding treatment (ie, initial, touch-up, or maintenance) for each diary for the purposes of analysis. If there are multiple diary entries for any day, then the worse severity and maximum duration will be included in the summary for the corresponding symptoms analyzed. Descriptive statistics will be provided for the endpoints listed in Table 5-10 for the initial treatment, touch-up treatment and maintenance treatment periods by predefined symptoms.

Table 5-10 Injection Site Responses Analyses

Endpoint	Description	Timing	Methodology
ISR severity	Maximum reported severity.	Initial, Touch-up, Maintenance	Categorical descriptive statistics
ISR duration	Duration from first instance of the symptom to the last instance of the symptom within the treatment period, where last instance means no further symptoms until the end of the 30-day diary period. Duration is derived as date of last ISR minus date of first ISR plus one, on the participant level, regardless of the side of the jaw.	Initial, Touch-up, Maintenance	Categorical descriptive statistics

Other symptoms of ISR entered in the e-diary by participants will be coded into medical terminologies by the medical safety physician. Other ISRs with "No," "None," or texts with a similar meaning will be coded as "N/A." If other ISRs are coded as one of the prespecified symptoms (ie, redness, pain after injection, tenderness to touch, firmness, lumps/bumps, bruising, itching, and discoloration), then the participant will be counted only once based on the maximum severity and duration across the prespecified symptom entries and the coded terms. Other symptoms of ISR that are coded into terms beyond the list of prespecified symptoms will be summarized under "Other" symptoms in the summary tables for ISR. If a participant reported more than one non-prespecified terms of "Other ISR," the participant will be counted once in the summary of "Other" symptoms for that treatment. The free texts entered by the participant, and the coded terms for "Other" symptoms will be included in the data listing.

If the incidence of any particular category of "Other" ISR is greater than 5%, then the severity and duration analyses described in Table 5-10 will be performed for that specific category of the "Other" ISR.

ISR day will be derived as ISR date – most recent treatment date + 1.

5.6.5. 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 each treatment will be summarized. The summary will be performed by treatment group for TP and MP as described in Table 5-11.

Table 5-11 Procedural Pain Analyses

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

5.6.6. Other Safety Endpoints

Other safety endpoints include the line change on Snellen visual acuity, confrontational visual fields, ocular motility, EI assessment of facial sensation, pronunciation video recording, the global score of jaw functional limitation (see Section 6.5.3), and concomitant medications and concurrent procedures.

The other safety endpoints will be summarized by treatment groups, for VT and VMT where applicable, respectively, except for concomitant medications and concurrent procedures, where listings will be provided for the Safety Population.

Line change based on Snellen visual acuity using the formula below, confrontation visual fields, and ocular motility assessments evaluated by the TI (or designee) will be summarized by descriptive statistics. Baseline in the below formula is the pre-treatment value prior to the most recent treatment.

Line change = $10 \times [\log10 (d_{baseline}/20) - \log10 (d_{follow\text{-}up}/20)]$
where $d_{baseline}$ = denominator of the Snellen equivalent unit at baseline,
$d_{follow\text{-}up}$ = denominator of the Snellen equivalent unit at follow-up visit

Facial sensation is based on the 2-point discrimination test and a light touch test. The 2-point discrimination is tested at 4 locations (anterior to the jowl and posterior to the jowl on each side of the jawline), and the number and percent of participants who are able to feel the 2 points will be summarized by distance, location, and treatment group. The light touch is tested at the same 4 locations as the discrimination test, and the smallest filament size that the participant felt the stimulus will also be summarized by location and treatment group.

Safety findings from the pronunciation video recordings will be summarized for actual and change from baseline values by visit and by treatment group, for VT and VMT where applicable, respectively. Values of "No Data/Inaudible Response" will be considered as missing data, and no imputation will be done.

The WHO Drug Dictionary Enhanced 201703 or newer will be used to classify prior and concomitant medications by therapeutic class and drug name. Prior medication is defined as any medication taken before the start date of the study treatment for treatment group, or before the randomization date for control group. Concomitant medication is defined as any medication

taken on or after the start date of study treatment for treatment group or on or after the randomization date for control group.

Actual and change from baseline values for overall (including all items) and the 3 subscales of mastication, mobility, and verbal and non-verbal communication from the Jaw Functional Limitation Scale will be summarized using descriptive statistics by treatment group for the VT and VMT populations. The overall score will be computed as the mean of the available items. Details on the algorithm can be found in Section 6.5.3. Baseline for the control group during TP is defined as the last non-missing value prior to receiving optional treatment.

5.6.7. Vital Signs

Vital sign measurements, including blood pressure (systolic and diastolic, while participant is seated), pulse rate, respiratory rate and temperature will be listed for the Safety Population.

5.6.8. Pregnancy Test Analyses

Urine pregnancy test is taken at screening, prior to every treatment (initial, touch-up, and maintenance), and study exit. Participants with a positive result for the Safety Population throughout the study period will be presented as a listing.

5.7. Other Analyses

5.7.1. Subgroup analyses

Analysis of the primary effectiveness endpoint, ALJDS responder status at Month 6, will be performed on the mITT population during the CP for the following subgroups:

- Investigational sites by treatment group.
- Total volume injected (total injection volume at initial and touch-up treatments combined): ≤ median and > median. This analysis will be done for treatment group only.
- Route of injection (ie, needle or cannula). The primary effectiveness endpoint at Month 6 will be presented by route of injection (i.e., needle or cannula). A participant is defined to be in the needle group if only needle is used at both initial and touch-up treatments. A participant is defined to be in the cannula group if either only cannula is used or cannula in combination with needle at either initial or touch-up treatments. A homogeneity test will be done to compare the primary effectiveness endpoint at Month 6 between the needle and cannula groups at alpha = 0.15. This analysis will be done only for the treatment group.

- Baseline ALJDS score. If the baseline ALJDS score is different for the left and right jaws then the worse score is used as baseline score.

No multiplicity adjustment will be performed.

5.7.2. Impact of COVID-19

All COVID-19 related data collected on the eCRFs will be presented in listings.

5.8. Interim Analyses

No interim analysis is planned for this study.

6. Supporting Documentation

6.1. Appendix 1: List of Abbreviations

AE	adverse event
AESI	adverse events of special interest
ALJDS	Allergan Loss of Jawline Definition Scale
CI	confidence interval
COVID-19	coronavirus disease 2019
CP	control period
eCRF	electronic case report form
EI	Evaluating Investigator
GAIS	Global Aesthetic Improvement Scale
ISR	injection site response
MedDRA	medical dictionary for regulatory activities
mITT	modified intent-to-treat
MP	maintenance treatment period
PT	preferred term
SAE	serious adverse event
SAP	statistical analysis plan
SOC	system organ class
TEAE	treatment-emergent adverse event
TESAE	treatment-emergent serious adverse event
TI	Treating Investigator
TP	treated period
VMT	VOLUX maintenance treatment population
VT	VOLUX treated population
WHO	World Health Organization

6.2. Appendix 2: List of Changes

6.2.1. Changes to Protocol-Planned Analyses

The following changes to the analyses specified in the statistical sections of Study V25L-002 Amendment 2 (Approved version dated 11-March-2019) have been implemented in this SAP:

1. Four analysis populations [mITT, Safety, Observed Primary Endpoint, VT and VMT] are defined instead of the 2 analysis populations (ITT and Safety) stated in the protocol [Section 4. Populations for Analysis].
2. The primary effectiveness endpoint will be analyzed using the mITT population instead of the ITT population as stated in the protocol [Section 5.3 Primary Effectiveness Endpoint Analysis].
3. Multiple imputation will be used as the primary method for the analysis of the primary endpoint instead of using the last observation carried forward. More specifically, multiple imputation with 30 imputed datasets will be applied to participants with missing Month 6 ALJDS values, generated using the following model (performed by treatment group) [Section 5.3 Primary Effectiveness Endpoint Analysis]:

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

4. The analysis of the effectiveness and/or safety variables will be summarized by the VT and VMT populations in TP and MP, respectively. In CP, the analysis of the effectiveness and/or safety variables will be summarized by the mITT and Safety populations [Section 5.3.4 Supplementary Analyses and 5.6.3 Adverse Events].
5. Baseline for the primary effectiveness endpoint is defined as the last non-missing value on or before the latter of randomization date or first treatment date in the control period [Section 5.1 General Considerations]. In the protocol, Baseline is defined as the last assessment by EI prior to randomization.

6.2.2. SAP Version History

The following are the changes to version 2 of SAP:

1. Updated the criteria for receiving touch-up treatment based on the protocol in Section 1.2.
2. Updated the number of imputed datasets used for the primary effectiveness endpoint analysis from 5 to 30 datasets in Sections 2 and 5.3.2 to reduce variability and increase efficiency of estimates. Changed the seed in Step 1 to keep it consistent with what was used in the lines of SAS code. Added the "likelihood=augment" option in the PROC MI SAS code in case the data has either complete separation pattern or quasi-complete separation pattern.
3. The naming of the analysis population for the sensitivity analysis of primary endpoint, ALJDS responder status at Month 6 was changed from "Completer Population" to "Observed Primary Endpoint" in Sections 4, 5.3.3, and 6.2.1. The definition of this population refers to the completion of the primary timepoint visit (ie., Month 6 of CP). This naming change is aimed to avoid potential confusion with study completion.
4. Updated the definition of TP for the control group to "follow-up period once the participants receive the optional treatment" and starts in the end date of CP in Section 5.1.
5. The definition of "baseline" for the effectiveness and safety variables in Sections 5.1 and 6.2.1 was changed to "the last non-missing assessment prior to the latter of randomization or first treatment in the control period" due to possibility of participants who were mis-treated (ie., either randomized to VOLUX but was not treated or randomized to Control but received VOLUX).
6. Specified that 95% CI for continuous variables will be based on the standard normal distribution in Section 5.1.
7. Added that the exact binomial method will be used to compute 95% CI for categorical variables in Section 5.1.
8. Added a bullet regarding how if any data is collected after exit then it will be summarized in the tables and listings in Section 5.1.
9. The wording of "Rasch transformed score" were specified for the analyses of FACE-Q Satisfaction with Lower Face and Jawline and FACE-Q Appraisal Lines: Marionette in Sections 5.4.1, 5.4.2, and 5.5 and to enhance clarity and consistency with Appendix 6.5.
10. The definition of change from baseline was moved from Section 5.4.1 to Section 5.1

11. Added clarification in Section 5.4.2 that analyses will be performed on the mITT population.
12. The imputation for the missing items of FACE-Q Satisfaction with Lower Face and Jawline and FACE-Q Appraisal Lines: Marionette were added to Section 5.4.2 and 5.5, respectively.
13. Added the definition of baseline for the control group for FACE-Q Satisfaction with Lower Face and Jawline in Section 5.5 since this assessment is not done on the control group during the control period.
14. Added clarification on the subset of participants on whom the analyses of FACE-Q Appraisal Lines: Marionette will be performed in Section 5.5. Additionally, the definition of responder on FACE-Q Appraisal Lines: Marionette was added to this section.
15. The wording of "responder status" was added to the "number (%) of responder" analysis for the endpoints of ALJDS, GAIS (EI and participant), FACE-Q Satisfaction with Lower Face and Jawline, and FACE-Q Appraisal Lines: Marionette in Table 5-6 for clarity.
16. The analysis of ISR in Section 5.6.4 was revised to include: (1) ISR summary by maximum duration instead of simply by duration and (2) analysis of "other symptoms" based on the coded terms assigned by the Medical Safety Physician.
17. Added clarification regarding analysis for 2-point discrimination and pronunciation video as well as definition of baseline in the control group during TP for Jaw Functional Limitation Scale in Section 5.6.6.
18. Subgroup analysis of the primary effectiveness endpoint by baseline ALJDS was added in Section 5.7.1.
19. A data listing to tabulate the COVID-19 related data collected during the study is added to Section 5.7.2.
20. Since no interim analysis regarding the primary and secondary endpoints was performed the text regarding the same was deleted in Section 5.8.
21. A summary for baseline ALJDS at jaw level as well as subject level was added in Section 6.3.2.
22. Changed the population for medical history listing from mITT population to all randomized participants in Section 6.3.4.

23. Removed the summary of prior and concomitant medications and changed the population for prior and concomitant medications listing from mITT population to all randomized participants in Section 6.3.5.

24. Added details regarding analysis windows for safety assessments in Section 6.4.1.2.

6.3. Appendix 3: Supporting Study Information

6.3.1. Demographics

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

6.3.2. Baseline and Disease Characteristics

Baseline characteristics will be summarized in total and by treatment group for the mITT population as follows:

- Weight (kg), height (cm), and body mass index (kg/m^2)
- Sun exposure (hours per day)
- Fitzpatrick Skin Phototype (Each phototype, and by phototype groups, ie, I and II, III and IV, V and VI)
- Smoking history (status, amount, tobacco product, frequency, and duration in months)
- Baseline ALJDS score at jaw level as well as subject level. For the subject level summary, if the right and the left jaws have different ALJDS score then the worse score is will be used for the summary

6.3.3. 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 participants as described in Table 6-1.

Listing of significant protocol deviations will be provided.

Table 6-1 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

6.3.4. 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 MedDRA, version 22.1 or newer. Listing will be provided for all randomized participants.

6.3.5. Prior/Concomitant Medications

Medications will be coded using the WHO Drug Dictionary, version March 2017 or newer. Listing of prior and concomitant medications will be provided for all randomized participants.

Table 6-2 Prior and Concomitant Medications

Endpoint	Description	Timing	Methodology
Prior medications	Medications taken before the study treatment for the treatment group or before randomization for the control group	Screening	Listing
Concomitant medications	Medications taken during the corresponding period	CP, TP, MP	Listing

6.4. Data Handling Convention

6.4.1. Analysis Window

6.4.1.1. Effectiveness

All scheduled and unscheduled visits with complete dates will be assigned visit windows and used for analysis (Table 6-3, Table 6-4, Table 6-5, and Table 6-6). However, if a scheduled visit with incomplete date is available with no other visits with complete date during the visit window, then the visit with incomplete date may be used for analysis, and the nominal visit will be used as the visit window (eg, if the only assessment collected for Month 3 visit window is on a Month 3 eCRF with incomplete date, then those data will be used for the Month 3 assessments).

For the treatment group, the effectiveness day is calculated relative to the last treatment date (either the initial treatment if no touch-up is performed, or the touch-up treatment, or the maintenance treatment date). If the analysis date is on or after the last treatment date, then the effectiveness day = analysis date - the last treatment date + 1. If the analysis date is before the last treatment date, then the effectiveness day = analysis date - the last treatment date.

For the control group during the no-treatment control period, the effectiveness day is calculated relative to the randomization date. If the analysis date is on or after the randomization date, then

effectiveness day = analysis date - randomization date + 1. If the analysis date is before the randomization date, then effectiveness day = analysis date - randomization date.

For the control group after receiving optional treatment, the effectiveness day is calculated relative to the last treatment date (either the initial treatment, if no touch-up is performed, or the touch-up treatment). If the analysis is on or after the last treatment date, then effectiveness day = analysis date - last treatment date + 1.

If there are multiple visits occurring within a single visit window with relevant data, the visit closest to the target day will be used in the analysis of the corresponding visit windows. If 2 visits are equal distance to the target day and are the same type of visit, then the later visit will be used.

Analysis visit windows for effectiveness endpoints are defined as follows:

If analysis date ≥ reference date, then

Analysis day = analysis date – reference date + 1

If analysis date < reference date, then

Analysis day = analysis date – reference date

Table 6-3 Effectiveness Analysis Visit Definitions for Treatment Group during the Treated Period

Scheduled Visit for Effectiveness Measure	Target Day of the Visit	Analysis Visit Window
Screening[a]	N/A	Screening visit; No more than 30 days prior to randomization
Baseline	N/A	Randomization/Initial Treatment visit
Touch-up Treatment[b]	Day 31 Post Initial Treatment	Day of the Touch-up Treatment
Last Treatment[c]	Day 1 Post Last Treatment	Day 1
Month 1	Day 31 Post Last Treatment	Day 2 to Day 61
Month 3	Day 91 Post Last Treatment	Day 62 to Day 136
Month 6	Day 181 Post Last Treatment	Day 137 to Day 226
Month 9	Day 271 Post Last Treatment	Day 227 to Day 316
Month 12	Day 361 Post Last Treatment	Day 317 to Day of Maintenance Treatment[d] or Day of Study Exit, whichever is earlier

[a] Participants may have screening and randomization on the same day. In such cases, only randomization visit is relevant.
[b] Not all participants will receive touch-up treatment.
[c] If touch-up is not performed then reference date is the date of initial treatment, otherwise reference date is the date of touch-up treatment.
[d] Not all participants will receive maintenance treatment.

Table 6-4 Effectiveness Analysis Visit Definitions for Control Group during the Control Period

Scheduled Visit for Effectiveness Measure	Target Day of the Visit	Analysis Visit Window with Touch-up
Randomization Day[a]	Day 1	N/A
Month 1	Day 31	Day 2 to Day 61
Month 3	Day 91	Day 62 to Day 136
Month 6	Day 181	Day 137 to Day of Optional Treatment or Day of Study Exit, whichever is earlier

[a] Participants may have screening and randomization on the same day. In such cases, only randomization visit is relevant. Randomization date is the reference date.

Table 6-5 Effectiveness Analysis Visit Definitions for Control Group during the Treated Period

Scheduled Visit for Effectiveness Measure	Target Day of the Visit	Analysis Visit Window
Optional Treatment	N/A	Initial Treatment visit
Touch-up Treatment[a]	Day 31 Post Initial Treatment	Day of the Touch-up Treatment
Last Treatment[b]	Day 1 Post Last Treatment	Day 1
Month 1	Day 31 Post Last Treatment	Day 2 to Day 61
Month 3	Day 91 Post Last Treatment	Day 62 to Day 136
Month 6	Day 181 Post Last Treatment	≥ Day 137

[a] Not all participants will receive touch-up treatment.
[b] If touch-up is not performed then reference date is the date of initial treatment, otherwise reference date is the date of touch-up treatment.

Table 6-6 Effectiveness Analysis Visit Definitions for Treatment Group during the Maintenance Period

Scheduled Visit for Effectiveness Measure	Target Day of the Visit	Analysis Visit Window
Maintenance Treatment[a]	Day 1 Maintenance Treatment	Day of Maintenance Treatment
Month 1 after Maintenance Treatment	Day 31 Post Maintenance Treatment	Day 2 to 61 Post Maintenance Treatment
Month 3 after Maintenance Treatment	Day 91 Post Maintenance Treatment	≥ Day 62 post Maintenance Treatment

[a] Not all participants will receive maintenance treatment. Date of maintenance treatment is the reference date.

6.4.1.2. Safety

No analysis windows are required for TEAEs, ISRs, Snellen visual acuity, confrontational visual field, and ocular motility assessments. Nominal visits will be used for summarizing Snellen visual acuity, confrontational visual field, and ocular motility assessments. Pronunciation videos

and facial sensation assessments will be summarized using analysis visit windows in Table 6-3 and Table 6-6 for the treatment group during the TP and MP, respectively, and in Table 6-4 and Table 6-5 for the control group during the CP and TP, respectively. Additionally, Jaw Functional Limitation Scale will be summarized using the analysis visit windows in Table 6-7 for the treatment group during the TP and MP and in Table 6-8 for the control group during the TP.

Table 6-7 Jaw Functional Limitation Scale Visit Windows for Treatment Group during the Treated and Maintenance Periods

Scheduled Visit for Safety Measure	Target Day of the Visit	Analysis Visit Window with Touch-up
Screening[a]	N/A	Screening visit; No more than 30 days prior to randomization
Baseline	N/A	Randomization/Initial Treatment visit
Day 14 After Initial Treatment	Day 15 Post Randomization/Initial Treatment	Day 2 to Day 23 of Randomization/Initial Treatment Period
Touch-up Treatment[b]	Day of Touch-up Treatment	Day 24 of Randomization/Initial Treatment Period to Touch-up Treatment Date
Day 14 After Touch-up Treatment	Day 15 Post Touch-up Treatment	Day 2 to Day 23 of Touch-up Treatment Period
Last Treatment[c]	Day of Last Treatment	Day of Last Treatment (Initial or Touch-up)
Month 1	Day 31 Post Last Treatment	Day 24 to Day 61
Month 3	Day 91 Post Last Treatment	Day 62 to Day 136
Month 6	Day 181 Post Last Treatment	Day 137 to Day 226
Month 9	Day 271 Post Last Treatment	Day 227 to Day 316
Month 12	Day 361 Post Last Treatment	Day 317 to Maintenance Treatment Date[d] or Study Exit Date, whichever is earlier
Maintenance Treatment[d]	Day 1 Maintenance Treatment	Day of Maintenance Treatment
Day 14 Maintenance Treatment	Day 15 Post Maintenance Treatment	Day 2 to Day 23 Post Maintenance Treatment
Month 1 after Maintenance Treatment	Day 31 Post Maintenance Treatment	Day 24 to Day 61 Post Maintenance Treatment
Month 3 after Maintenance Treatment	Day 91 Post Maintenance Treatment	≥ Day 62 Post Maintenance Treatment

[a] Participants may have screening and randomization on the same day. In such cases, only randomization visit is relevant.
[b] Not all participants will receive touch-up treatment.
[c] Initial treatment if touch-up is not performed, otherwise touch-up treatment.
[d] Not all participants will receive maintenance treatment.

Table 6-8 Jaw Functional Limitation Scale Visit Windows for Control Group during the Treated Period

Scheduled Visit for Safety Measure	Target Day of the Visit	Analysis Visit Window with Touch-up
Optional Treatment	N/A	Initial Treatment visit
Day 14 After Initial Treatment	Day 15 Post Initial Treatment	Day 2 to Day 23 of Initial Treatment Period
Touch-up Treatment[a]	Day of Touch-up Treatment	Day 24 of Initial Treatment Period to Touch-up Treatment Date
Day 14 After Touch-up Treatment	Day 15 Post Touch-up Treatment	Day 2 to Day 23 of Touch-up Treatment Period
Last Treatment[b]	Day of Last Treatment	Day of Last Treatment (Initial or Touch-up)
Month 1	Day 31 Post Last Treatment	Day 24 to Day 61
Month 3	Day 91 Post Last Treatment	Day 62 to Day 136
Month 6	Day 181 Post Last Treatment	Day ≥ 137

[a] Not all participants will receive touch-up treatment.
[b] Initial treatment if touch-up is not performed, otherwise touch-up treatment.

6.4.2. Missing Date Imputation

Missing date will only be imputed for TEAEs. Dates may be imputed with year, month, and day values under certain scenarios described in Table 6-9.

Table 6-9 Imputation Scenarios

Scenario	Complete			Imputable
	Year	Month	Day	
1	Yes	Yes	Yes	Complete
2	Yes	Yes	—	Yes
3	Yes	—	Yes	No[a]
4	Yes	—	—	Yes
5	—	Yes	Yes	No[a]
6	—	Yes	—	No[a]
7	—	—	Yes	No[a]
8	—	—	—	Yes

[a] Not allowed per database design.

6.4.2.1. Missing/Incomplete AE Start Date

For Scenario 2 (ie, if day is missing but month and year are available), the imputed day will be the first day of the month or the initial treatment date (randomization date for control group) if they have the same month and year, whichever is later (because TEAE onset is not expected prior to administration of study treatment);

For Scenario 4 (ie, if day and month are missing but year is available), the imputed day and month will be 01 Jan or the initial treatment date (randomization date for control group) if they have the same year, whichever is later.

For Scenario 8 (ie, if day, month, and year are all missing), the imputed start date is the initial treatment date for treatment group and the randomization date for control group.

6.4.2.2. Missing/Incomplete AE End Date

For Scenario 2 (ie, if day is missing but month and year are available), then the imputed day will be the last day of the month or the study exit date if they have the same month and year, whichever is earlier.

For Scenario 4, (ie, if day and month are missing but year is available), the imputed day and month will be 31 Dec or the study exit date if they have the same year, whichever is earlier.

For Scenario 8 (ie, if day, month, and year are all missing), the imputed date will be the study exit date.

6.5. Scales and Scoring Algorithms

6.5.1. FACE-Q™ – Satisfaction with Lower Face and Jawline Scale

For each question, circle only one answer. With your lower face in mind (lower cheeks and jawline), in the past week, how satisfied or dissatisfied have you been with:

	Very Dissatisfied	Somewhat Dissatisfied	Somewhat Satisfied	Very Satisfied
a. How prominent your jawline looks?	1	2	3	4
b. How sculpted (well-defined) your jawline looks?	1	2	3	4
c. How your jawline looks in profile (side view)?	1	2	3	4
d. How nice your lower face looks?	1	2	3	4
e. How smooth your lower face looks (i.e. no jowls or folds of fatty skin)?	1	2	3	4

Note to Investigators: This scale can be used independently of the other scales. REDCap Data Dictionary files (http://projectredcap.org/) and SPSS syntax files are available.

Psychometric Paper: Klassen AF, Cano SJ, Scott AM, Pusic AL. Measuring outcomes that matter to facelift patients: development and validation of FACE-Q Appearance Appraisal Scales and Adverse Effects Checklist for the lower face and neck. Plast Reconstr Surg. 2014;133(1):21-30.

FACE-Q™ - SATISFACTION WITH LOWER FACE & JAWLINE CONVERSION TABLE

Instructions: Higher scores reflect a better outcome. If missing data is less than 50% of the scale's items, insert the mean of the completed items. Use the Conversion Table below to convert the raw scale summed score into a score from 0 (worst) to 100 (best).

SUM SCORE	EQUIVALENT RASCH TRANSFORMED SCORE (0-100)
5	0
6	9
7	16
8	22
9	28
10	34
11	40
12	46
13	52
14	59
15	66
16	72
17	78
18	84
19	92
20	100

FACE-Q™ - SATISFACTION WITH LOWER FACE & JAWLINE INTERPRETATION TABLE

	Very Dissatisfied	Somewhat Dissatisfied	Somewhat Satisfied	Very Satisfied
a. How prominent your jawline looks?	0-14	14-39	39-81	81-100
b. How sculpted (well-defined) your jawline looks?	0-19	19-48	48-78	78-100
c. How your jawline looks in profile (side view)?	0-22	22-48	48-81	81-100
d. How nice your lower face looks?	0-19	19-54	54-81	81-100
e. How smooth your lower face looks (i.e. no jowls or folds of fatty skin)?	0-25	25-56	56-83	83-100

FACE-Q™ Interpretation Tables:

Prior psychometric evidence justified the unweighted sum score across items in the FACE-Q. The estimation of person measures linearized the raw scores allowing for the application of parametric statistics requiring metric scales. The scoring table (based on person measures) was based on weighted likelihood estimation (WLE) [1] given a person's raw score and item parameters, the calibration of which was based on the unrestricted Rasch model for polytomous responses [2].

For complete data, a conversion table was produced that indicates how sum scores are converted to measures expressed in the original Rasch logit format. These values range, in principle, from minus to plus infinity. However, since WLE allows for a finite estimate for extreme response patterns, the actual range of measures is finite. With a metric scale format, linear transformations are permissible. Accordingly, as a final step, the logit measures were transformed to 0-to-100 for convenience and to aid interpretation and reporting of results by moving the scale range origin (i.e., 0) to 50 and adding 50.

The Conversion Table provided above shows the conversion of the raw sum score into a 0-to-100 score. The Interpretation Table shows the implied range of scores for each response option (for each and every item/response option). This table is itself based directly on the threshold plots produced through the Rasch analysis [3, 4].

After running a statistical analysis of clinical trial data, users of this table can take mean transformed FACE-Q scores (or even individual scores) and cross reference the numbers in the table to ascertain better interpretation.

References

1. Warm T. Weighted likelihood estimation of ability in item response theory. Psychometrika 1989; 54:427-450.
2. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978; 43:561-573.
3. Andrich D. Rating scales and Rasch measurement. Expert Rev Pharmacoeconomics Outcomes Res 2011; 11:571-585.
4. RUMM 2030 [computer program]. Perth, WA: RUMM Laboratory Pty Ltd, 1997-2016.

6.5.2. FACE-Q™ – Appraisal Lines: Marionette

For each question, circle only one answer. With the deep lines that run downward from the corner of your lips to your chin in mind (marionette lines), in the past week, how much have you been bothered by:

	Not at all	A little	Moderately	Extremely
a. How angry your marionette lines make you look?	1	2	3	4
b. How sad your marionette lines make you look?	1	2	3	4
c. How tired your marionette lines make you look?	1	2	3	4
d. How your marionette lines look when your face is relaxed (still)?	1	2	3	4
e. How deep your marionette lines look?	1	2	3	4
f. How old your marionette lines make you look?	1	2	3	4
g. How noticeable your marionette lines are?	1	2	3	4

Note to Investigators: This scale can be used independently of the other scales. REDCap Data Dictionary files (http://projectredcap.org/) and SPSS syntax files are available.

Psychometric Paper: Klassen AF, Cano SJ, Schwitzer J, Baker SB, Carruthers A, Carruthers J, Chapas A, Pusic AL. Development and Psychometric Validation of the FACE-Q Skin, Lips and Facial Rhytides Appearance Scales and Adverse Effect Checklists for Cosmetic Procedures. JAMA Dermatol. 2016 Apr 1;152(4):443-51.

FACE-Q™ - APPRAISAL OF LINES: MARIONETTE CONVERSION TABLE

Instructions: Ensure the data are rescored as follows: "Not at all" = 4; "A little" = 3; "Moderately" = 2; "Extremely" = 1. Higher scores reflect a better outcome. If missing data is less than 50% of the scale's items, insert the mean of the completed items. 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)
7	0
8	7
9	13
10	17
11	21
12	25
13	30
14	35
15	40
16	45
17	48
18	52
19	56
20	60
21	64
22	70
23	75
24	79
25	83
26	88
27	93
28	100

FACE-Q™ - APPRAISAL OF LINES: MARIONETTE INTERPRETATION TABLE

	Extremely	Moderately	A little	Not at all
a. How angry your marionette lines make you look?	0-11	11-44	44-69	69-100
b. How sad your marionette lines make you look?	0-14	14-46	46-77	77-100
c. How tired your marionette lines make you look?	0-25	25-50	50-79	79-100
d. How your marionette lines look when your face is relaxed (still)?	0-19	19-52	52-84	84-100
e. How deep your marionette lines look?	0-19	19-52	52-85	85-100
f. How old your marionette lines make you look?	0-25	25-55	55-80	80-100
g. How noticeable your marionette lines are?	0-21	21-52	52-89	89-100

FACE-Q™ Interpretation Tables:

Prior psychometric evidence justified the unweighted sum score across items in the FACE-Q. The estimation of person measures linearized the raw scores allowing for the application of parametric statistics requiring metric scales. The scoring table (based on person measures) was based on weighted likelihood estimation (WLE) [1] given a person's raw score and item parameters, the calibration of which was based on the unrestricted Rasch model for polytomous responses [2].

For complete data, a conversion table was produced that indicates how sum scores are converted to measures expressed in the original Rasch logit format. These values range, in principle, from minus to plus infinity. However, since WLE allows for a finite estimate for extreme response patterns, the actual range of measures is finite. With a metric scale format, linear transformations are permissible. Accordingly, as a final step, the logit measures were transformed to 0-to-100 for convenience and to aid interpretation and reporting of results by moving the scale range origin (i.e., 0) to 50 and adding 50.

The Conversion Table provided above shows the conversion of the raw sum score into a 0-to-100 score. The Interpretation Table shows the implied range of scores for each response option (for each and every item/response option). This table is itself based directly on the threshold plots produced through the Rasch analysis [3, 4].

After running a statistical analysis of clinical trial data, users of this table can take mean transformed FACE-Q scores (or even individual scores) and cross reference the numbers in the table to ascertain better interpretation.

References

5. Warm T. Weighted likelihood estimation of ability in item response theory. Psychometrika 1989; 54:427-450.
6. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978; 43:561-573.
7. Andrich D. Rating scales and Rasch measurement. Expert Rev Pharmacoeconomics Outcomes Res 2011; 11:571-585.
8. RUMM 2030 [computer program]. Perth, WA: RUMM Laboratory Pty Ltd, 1997-2016.

6.5.3. Jaw Functional Limitation Scale 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 2 settings due to the very high reliability of the global score, whether derived from the full instrument or from the short instrument.

Scoring

From either the short form (all items) or the long form (items 1, 3, 6, 10, 11, 12, 13, and 19), 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., et al. (2008). The Jaw Functional Limitation Scale: Development, reliability, and validity of 8-item and 20-item versions. *Journal of Orofacial Pain* **22**: 219-230.

Ohrbach, R., et al. (2011). "Clinical findings and pain symptoms as potential risk factors for chronic TMD: Descriptive data and empirically identified domains from the OPPERA case-control study." Journal of Pain 12(11, Supplement 3): T27-T45.

7. References

Guidance for industry, patient-reported outcome measures: use in medical product development to support labeling claims, December 2009. https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf. Accessed January 30, 2018.

Klassen AF, Cano SJ, Schwitzer J, Baker SB, Carruthers A, Carruthers J, et al. Development and psychometric validation of the FACE-Q skin, lips and facial rhytides appearance scales and adverse effect checklists for cosmetic procedures. JAMA Dermatol. 2016;152(4):443-51.

Klassen AF, Cano SJ, Scott AM, Pusic AL. Measuring outcomes that matter to facelift patients: development and validation of FACE-Q Appearance Appraisal Scales and Adverse Effects Checklist for the lower face and neck. Plast Reconstr Surg. 2014;133(1):21-30.

Ohrbach R, Larsson P, List T. The Jaw Functional Limitation Scale: development, reliability, and validity of 8-item and 20-item versions. Journal of Orofacial Pain. 2008;22:219.