I5Q-MC-CGAL Statistical Analysis Plan Version 5

A Phase 3 Randomized, Double-Blind, Placebo-Controlled Study of LY2951742 in Patients with Episodic Cluster Headache

NCT02397473

Approval Date: 05-Apr-2018

1. Statistical Analysis Plan for Protocol I5Q-MC-CGAL: A Phase 3 Randomized, Double-Blind, Placebo-Controlled Study of LY2951742 in Patients with Episodic Cluster Headache

Galcanezumab (LY2951742)

Study CGAL is a Phase 3 multi-center, outpatient, randomized, double-blind, placebo-controlled study of galcanezumab 300 mg in the prevention of episodic cluster headache. The study consists of 4 study phases (SP): SP I (screening/washout), SP II (pre-randomization diary), SP III (randomized, double-blind, placebo-controlled treatment), and SP IV (post-treatment follow-up).

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol I5Q-MC-CGAL
Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 18 December 2014
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly: 11 February 2015
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly: 13 September 2016
Statistical Analysis Plan Version 4 electronically signed and approved by Lilly : January 31 2018
Statistical Analysis Plan Version 5 electronically signed and approved by Lilly on date provided below.

Approval Date: 05-Apr-2018 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Appendix **Page**

3. Revision History

Statistical Analysis Plan (SAP) Version 1 was approved on 18 December 2014.

Statistical Analysis Plan Version 2 was approved prior to first patient visit and any unblinding. The overall changes and rationale for the changes incorporated in Version 2 are as follows:

- The electronic patient reported outcome (ePRO) diary will now collect the average duration and average pain for the time period rather than for each attack. Thus the derivation for mean severity and mean duration of cluster headache attack were updated.
- The approach for missing data was updated for each weekly interval.
    - 1) If there are less than or equal to 3 days with nonmissing answer to cluster headache attack frequency in the weekly interval; or 2) the primary efficacy compliance rate is less than or equal to 50%, then the weekly interval will be considered missing;
    - Otherwise, 1) if there are greater than or equal to 4 days with nonmissing answer to cluster headache attack frequency in the weekly interval; and 2) the primary efficacy compliance rate is greater than 50%, then the average number of cluster headache attacks across the nonmissing days will be used to impute the missing days.
- The algorithm for pooling of sites was updated.
- The primary endpoint point estimate was updated to use the unadjusted estimate, and the median unbiased estimate will be used for sensitivity.
- ePRO diary compliance was updated to calculate both ePRO diary primary efficacy compliance rate and overall ePRO diary compliance rate.
- Addition of analysis for change from baseline in total weekly dose of sumatriptan Sc, sumatriptan nasal spray, and zolmitriptan nasal spray separately as well as combined.

Statistical Analysis Plan Version 3 was approved prior to IA1. The changes incorporated in Version 3 are as follows:

- Posttreatment follow-up phase safety analyses will have only 1 baseline.
- A section on protocol violations to be identified was added.
- Sensitivity analyses were updated, to be consistent with other Phase 3 studies of LY2951742.
- Infections section will only deal with upper respiratory tract infections; analyses were modified to be consistent with other Phase 3 studies of LY2951742.
- For Columbia-Suicide Severity Rating Scale (C-SSRS), 1 bullet was split into 2 to enhance readability, and baseline definition for improvement from baseline analysis was clarified.

- Criteria for sustained elevation in diastolic blood pressure was changed to be consistent with the single time point analysis.
- An additional criteria threshold for QTc increase was added.
- Analyses of elevations in hepatic laboratory tests were clarified, and an additional subset was added.
- Immunogenicity analyses were updated to be consistent with other Phase 3 studies of LY2951742.
- An additional subgroup analysis category, for age, was added.
- Some minor corrections and clarifications were made.

Statistical Analysis Plan Version 4 was approved after the last patient was randomized and prior to interim analysis (the first and final assessment of primary efficacy endpoint after all patients completed double-blind phase, which is the first unblinding to study team). Enrollment in the trial was terminated (due to enrollment infeasibility) prior to reaching the sample size target for the originally planned interim analysis. Therefore, the planned interim analysis for sample size re-estimation did not occur.

In the SAP Version 4, the updates were made mainly for incorporating the recent learnings from migraine data or for consistency across the galcanezumab program. The changes incorporated in Version 4 of the SAP are summarized as follows:

- LY2951742 was replaced by galcanezumab in the body of the SAP.
- Consistent with the primary endpoint and analysis methodology for the pivotal migraine studies, the primary endpoint was updated to be the overall treatment effect over the Weeks 1 to 3 during the double-blind treatment phase, rather than the treatment effect at the single time point, Week 3.
- In Section 5.4.1.2, the exploratory endpoints for severity and duration of cluster headache attack pain and for the abortive medications were updated to clarify the research questions and the derivations were modified correspondingly.
- In Section 5.5.1.1 and 5.5.8.3, it was clarified that, for other secondary and exploratory efficacy measures that are not derived from cluster headache frequency, the baseline average daily cluster headache attack frequency category variable is included in the statistical analysis models.
- The list of analyses for other secondary and exploratory efficacy variables were updated in Table CGAL.5.3. Last observation carried forward (LOCF) analysis for some exploratory variables was removed.
- Since no partially completed diary can be submitted, the ePRO diary primary efficacy compliance and overall ePRO diary compliance are combined into 1 diary compliance calculation in Section 5.5.6.

- In Section 5.5.1 and Table CGAL.5.2, safety population and modal treatment description for SP III were added for safety analyses since it is more appropriate to present safety results by the actual treatments patients received.
- Terminologies and identification criteria were updated for adverse event of special interest (AESIs) for the consistency across the galcanezumab program.
- In Section 5.5.9.1.3, detailed baseline and postbaseline definition for vital signs and weight were added. The patient populations for analysis that do not satisfy treatment-emergent definition were removed from Table CGAL.5.4.
- In Section 5.5.9.1.4, the parameter of large clinical trial population based QT correction (QTcLCTPB) was removed for electrocardiogram (ECG) analysis. The detailed baseline and postbaseline definitions for ECG were added.
- In Section 5.5.9.2, for continuous safety measures, Box-whisker plots with summary tables for SP III replaced LOCF and repeated measures analysis.
- Section 5.5.9.1.6 of Immunogenicity was updated to clarify definitions and modify analyses to focus on evaluation of the incidence of baseline anti-drug antibodies (ADA) and treatment-emergent ADA.
- Subgroup analysis for safety endpoints were removed due to the small size of the study. A few subgroup variables for the efficacy endpoint were removed due to the small size in subgroups.
- Section 5.6, Interim Analysis is updated to explain the changes in the interim analysis plan.
- Since the originally planned interim analysis for sample size re-estimation will not be conducted due to enrollment infeasibility, all languages and methodology descriptions related to sample size increase are removed.
- In Section 5.8, reports to be generated were updated to reflect that analyses from all SPs specified in this SAP will be performed at the interim analysis instead of only performing analyses for SP III. However, the analyses conducted for SP III will be deemed final since all patients will complete SP III at the interim analysis. The analyses using data from SP IV will be rerun and updated when the completed data are available at the final database lock.
- An appendix of important protocol deviations was added.
- Other minor corrections, modifications, and clarifications were made.

Statistical Analysis Plan Version 5 has been approved prior to the interim analysis (the first and final assessment of primary efficacy endpoint after all patients completed double-blind phase, which is the first unblinding to study team). There is no modification to the primary analysis methodologies for the primary, key secondary, and other secondary efficacy endpoints. The changes incorporated in Version 5 are summarized as follows:

- In Section 5.4.1.2, the exploratory endpoint for cluster headache attack duration was modified from "average weekly cluster headache attack minutes per attack for the remaining cluster headache days" to "weekly total cluster headache attack duration." An exploratory responder endpoint for the weekly total cluster headache attack duration that is defined as 30% or greater reduction is also added.
- In Section 5.5.9.1.2, removed the requirement of needing at least 4 events occurred in at least one treatment to display p-value.
- In Table CGAL.5.4, added additional patient populations for analysis of treatment-emergent, potentially clinically significant changes and sustained elevation in vital signs.
- In Section 5.5.10, additional subgroup variables were added for subgroup analysis.

CCI

- In the table of Description of Important Protocol Deviations in Appendix 1, updated the data source of the Important Protocol Deviations (IPDs) to only display the final data source for the IPD analysis. Two new IPDs were added.

4. Study Objectives

4.1. Primary Objective

The primary objective is to assess the efficacy of galcanezumab 300 mg every 30 days compared with placebo in reducing the frequency of weekly cluster headache attacks in patients with episodic cluster headache. The primary outcome measure will be the weekly cluster headache attack frequency. The primary endpoint will be the overall mean change from baseline in weekly cluster headache attack frequency across Weeks 1 to 3 with galcanezumab compared with placebo. Baseline is defined as the last 7 days in the eligibility report (prerandomization diary phase).

4.2. Secondary Objectives

4.2.1. Gated Objective

To assess the efficacy of galcanezumab compared with placebo in the proportion of patients meeting response at Week 3. For this analysis, response is defined as a reduction from baseline of 50% or greater in the weekly cluster headache attack frequency.

4.2.2. Other Secondary Objectives

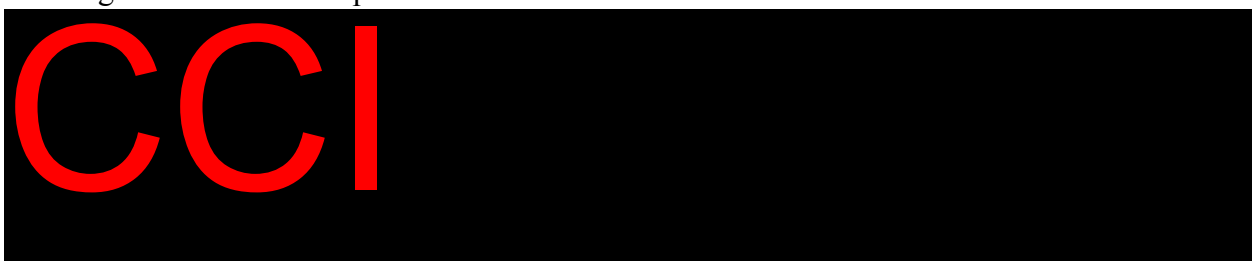
- To assess whether galcanezumab is superior to placebo on the following:
    - The proportion of patients with a 50% or greater reduction in the weekly number of cluster headache attacks from baseline for each weekly interval through Week 8
    - The proportion of patients with a 30% or greater reduction in the weekly number of cluster headache attacks from baseline for each weekly interval through Week 8
    - Mean change in the weekly cluster headache attack frequency from baseline for each weekly interval through Week 8
    - Proportion of patients reporting a score of 1 (“very much better”) or 2 (“much better”) on the Patient Global Impression of Improvement (PGI-I) at Week 4 and Week 8.
- To compare the safety and tolerability of galcanezumab with placebo in patients with episodic cluster headache using the following measures:
    - spontaneously reported treatment-emergent adverse events (TEAEs)
    - serious adverse events (SAEs)
    - discontinuation rates
    - suicidal ideation and behaviors assessed by solicited questioning using the C-SSRS

- To assess the development and consequences of ADA to galcanezumab in patients exposed to galcanezumab; to provide samples for subsequent evaluation of neutralizing ADA (NAb).
- To evaluate the pharmacokinetics of galcanezumab.

4.3. Exploratory Objectives

To assess whether galcanezumab is superior to placebo as measured by the following:

- Mean change in the weekly number of times an abortive medication was taken from baseline for each weekly interval through Week 8 comparing galcanezumab with placebo.
- Change in percentage of times using oxygen from baseline for each weekly interval through Week 8 comparing galcanezumab with placebo.
- Change in percentage of times using triptan from baseline for each weekly interval through Week 8 comparing galcanezumab with placebo.
- Change in percentage of times of using acetaminophen/paracetamol or nonsteroidal anti-inflammatory drugs (NSAIDs) from baseline for each weekly interval through Week 8 comparing galcanezumab with placebo.
- The proportion of patients with a 75% or greater reduction in the weekly number of cluster headache attacks from baseline for each weekly interval through Week 8 comparing galcanezumab with placebo.
- The proportion of patients with a 100% reduction in the weekly number of cluster headache attacks from baseline for each weekly interval through Week 8 comparing galcanezumab with placebo.
- Mean change from baseline in the cluster headache attack average weekly pain severity (based on 5-point scale) from baseline through Week 8 comparing galcanezumab with placebo.



5. A Priori Statistical Methods

5.1. Study Design

Study CGAL is a Phase 3 multi-center, outpatient, randomized, double-blind, placebo-controlled study of LY2951742 300 mg in the prevention of episodic cluster headache. The study has 4 study phases (SP): SP I (screening/washout), SP II (pre-randomization diary), SP III (randomized, double-blind, placebo-controlled treatment), and SP IV (post-treatment follow-up). Patients who discontinue the study during the double-blind treatment phase should enter the post-treatment follow-up phase.

5.2. Determination of Sample Size

The study is planned to have a minimum of approximately 162 patients randomized 1:1 to placebo or galcanezumab 300 mg with the opportunity to increase the final sample size at an interim analysis if indicated in order to maintain a well-powered study. To preserve blinding, details of the sample size and power calculations are omitted from this SAP and are provided in a separate document to the Ethical Review Board (ERB).

5.3. Randomization and Treatment Assignment

At Visit 3, eligible patients will be randomized in a 1:1 ratio to double-blind placebo or galcanezumab 300 mg (GMB300mg), respectively. To achieve marginal balance of treatment assignments for the factors of gender, average daily attack frequency (≤4 attacks per day, >4 attacks per day) and investigative site, randomization will be conducted with a dynamic allocation (minimization) method (Pocock and Simon 1975) with target probability of 0.8. Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive web-response system (IWRS).

5.4. Endpoints

5.4.1. Efficacy Endpoint

5.4.1.1. Cluster Headache Attack Primary Endpoint

Patient-Rated Daily Electronic Patient-Reported Outcome (ePRO) Diary: Patients will be asked to record the number of cluster headache attacks in their daily ePRO diary during SP II and SP III, which is used to derive the primary efficacy endpoint. Information regarding abortive medication use, cluster headache attack duration on average, and cluster headache attack pain severity on average will also be recorded. Pain severity will be rated using a 5-point pain scale, where 0=no pain, 1=mild pain, 2=moderate pain, 3=severe pain, and 4=very severe pain (The Sumatriptan Cluster Headache Study Group 1991). Patients should record all cluster attacks regardless of attack duration.

5.4.1.2. Derived Variables for Cluster Headache Attacks

In Study CGAL, for primary measure of cluster headache attacks, the daily data for each patient (including last 7 days in the eligibility report [prerandomization diary phase], 8 weeks of daily

data during treatment) will be converted into 9 roughly 7-calendar day intervals: the baseline 7-day interval, Week 1, 2, 3, 4, 5, 6, 7, and 8. Each day, the patient may report zero, 1, or multiple cluster headache attacks. Any ePRO diary data reported beyond the protocol defined collection period will not be used for statistical analysis.

The approach to split the postbaseline data into weekly intervals is done as follows:

- Firstly, postbaseline daily data will be split into Weeks 1 to 4 versus Weeks 5 to 8 using 1st and 2nd injection date. All data ≥1st injection date and <2nd injection date will be considered as Weeks 1 to 4; all data ≥2nd injection date and < treatment phase disposition date will be considered as Weeks 5 to 8. If 2nd injection date is missing, then all the data before treatment phase disposition date will be put into Weeks 1 to 4.
- Secondly, data within Weeks 1, 2, and 3 will be determined using calendar days. In other words, the 1st injection date will be considered as Day 1, then Days 1 to 7 will be Week 1; Days 8 to 14 will be Week 2; Days 15 to 21 will be Week 3. Week 4 will include all the data from Day 22 to the date before the 2nd injection (or before the treatment phase disposition date if second injection date is missing).
- Thirdly, if the 2nd injection date is not missing, the data within Weeks 5, 6, 7, and 8 will be determined using calendar days. In other words, the 2nd injection date will be considered as Day 1, then Days 1 to 7 will be Week 5; Days 8 to 14 will be Week 6; Days 15 to 21 will be Week 7. Week 8 will include all the data from Day 22 to the date before the treatment phase disposition date.

For each weekly interval, the following missing data imputation method will be used:

- 1) if there are less than or equal to 3 days with nonmissing answer to cluster headache attack frequency in the weekly interval; or 2) the diary compliance rate is less than or equal to 50%, then the weekly interval will be considered missing;
- Otherwise, 1) if there are greater than or equal to 4 days with nonmissing answer to cluster headache attack frequency in the weekly interval; and 2) the diary compliance rate is greater than 50%, then the average number of cluster headache attacks across the nonmissing days will be used to impute the missing days. Furthermore, the total cluster headache attack frequency during the weekly interval will be calculated as the average number of cluster headache attacks across nonmissing days times the actual number of calendar days within each weekly interval.

The same missing data imputation approach will also be applied to secondary and exploratory efficacy measures that are derived from ePRO data.

Then to estimate a weekly outcome of the total frequency for an efficacy measure from ePRO diary, the data within each week will be adjusted to a 7-day interval by multiplying $\frac{7}{x}$, where "x" is the actual number of calendar days within each weekly interval. Lastly, the change from baseline to Weeks 1, 2, 3, 4, 5, 6, 7, and 8 will be derived.

An example of missing data imputation is described below in Table CGAL.5.1.

Table CGAL.5.1. Example of Missing Data Imputation Outcome

	Example 1			Example 2		
	Number of Calendar Days	Number of Days with Nonmissing Answer to Cluster Headache Attack Frequency	Missing Data Imputation	Number of Calendar Days	Number of Days with Nonmissing Answer to Cluster Headache Attack Frequency	Missing Data Imputation
Week 1	7	7	*a	7	7	*a
Week 2	7	4	*b	7	4	*b
Week 3	7	3	*c and *d	7	3	*c and *d
Week 4	13	6	*c	13	8	*b
Week 5	7	7	*a	7	7	*a
Week 6	7	7	*a	7	7	*a
Week 7	7	7	*a	7	7	*a
Week 8	5	3	*d	3	3	*d

*a No imputation.

*b The average number of cluster headache attacks across the non-missing days will be used to impute the missing days.

*c Set to missing (diary compliance ≤50%).

*d Set to missing (number of days with nonmissing answer to cluster headache attack frequency ≤3).

Gated secondary, other secondary, and exploratory efficacy measures will be derived for each patient for each 7-day interval as follows:

- A 30%, 50%, 75%, and 100% responder is defined as any patient who has a ≥30%, ≥50%, ≥75%, and =100% reduction in the number of cluster headache attacks in a 7-day interval relative to baseline interval. For 30%, 50%, 75%, and 100% responder definition, percentage reduction from baseline will be calculated as follows:

$$\frac{100 \times (-1)\text{X(weekly of cluster headache attacks at week X} - \text{weekly of cluster headache attacks at baseline Interval)}}{\text{weekly \# of cluster headache attacks at baseline interval}}$$

- Change from baseline for the remaining cluster headache attack days:
  - Change from baseline in the cluster headache attack average weekly pain severity for the remaining cluster headache attack days will be derived at each weekly interval through Week 8. For the calculation of mean severity of cluster headache attacks, severity has 5 categories: 0=no pain, 1=mild pain, 2=moderate pain, 3=severe pain, and 4=very severe pain. The mean severity for the remaining cluster headache attack days for each interval will be calculated as follows:

$$\frac{\text{Sum of average cluster headache severity per day during the interval}}{\text{\# of days with cluster headache attack during the interval}}$$

If there is zero cluster headache attack within the weekly interval, then the mean severity of cluster headache attack for that interval will be considered not applicable hence missing at the interval for analyses purpose.

- Change from baseline in weekly total cluster headache attack duration will be calculated for each weekly interval. Average duration of cluster headache attacks during a 24-hour period was asked in the ePRO diary. Patients were instructed to round up to the next duration selection with following choices: 15 minutes, 30 minutes, 1 hour, 2 hours, 3 hours, >3 hours. If the duration is >3 hours, then 4 hours will be imputed for the calculation of the total cluster headache attack duration. The total cluster headache attack duration for each interval will be calculated as the summation of the average duration of cluster headache attack multiplied by the number of cluster headache attacks in the day during the interval. If the total duration is more than 24 hours for a day, it will be set to 24 hours.
- The proportion of patients with a 30% or greater reduction from baseline in the weekly total cluster headache attack duration will be calculated for each weekly interval.
- Change from baseline in weekly number of times of using oxygen as abortive medication at each interval will be calculated.
- Change from baseline in weekly number of times of using oral triptan, sumatriptan nasal spray or zolmitriptan nasal spray as abortive medication at each interval will be calculated.
- Change from baseline in weekly number of times of using sumatriptan Sc as abortive medication at each interval will be calculated.
- Change from baseline in weekly number of times of using acetaminophen/paracetamol or NSAIDs as abortive medication at each interval will be derived.
- Change from baseline in number of times using oxygen as abortive medication per cluster headache attack at each interval will be derived. The endpoint at each interval will be calculated as follows:

$$\frac{\text{Total number of times using of oxygen during the interval}}{\text{\# of cluster headache attack during the interval}}$$

- Change from baseline in number of times using oral triptan, sumatriptan nasal spray, or zolmitriptan nasal spray as abortive medication per cluster headache attack at each interval will be derived. The endpoint at each interval will be calculated as follows:

$$\frac{\text{Total number of times using the specified types of triptan during the interval}}{\text{\# of cluster headache attack during the interval}}$$

- Change from baseline in number of times using sumatriptan Sc as abortive medication per cluster headache attack at each interval will be derived.
- Change from baseline in number of times using acetaminophen/paracetamol or NSAIDs as abortive medication per cluster headache attack at each interval will be derived. The endpoint at each interval will be calculated as follows:

$$\frac{\text{Total number of times using of acetaminophen/paracetamol or NSAIDs during the interval}}{\text{\# of cluster headache attack during the interval}}$$

- Change from baseline in total weekly dose for oral triptan, sumatriptan nasal spray, and zolmitriptan nasal spray combined will be derived. Total weekly dose will be calculated as follows:

$$\frac{\text{Sum of doses of oral triptan, sumatriptan nasal spray, and zolmitriptan nasal spray during the interval} * 7}{\text{\# of nonmissing diary days during the interval}}$$

- Change from baseline in total weekly dose for sumatriptan Sc, oral triptan, sumatriptan nasal spray, and zolmitriptan nasal spray will be derived separately. Total weekly dose, respectively, will be calculated as follows:

$$\frac{\text{Sum of doses of sumatriptan Sc during the interval} * 7}{\text{\# of nonmissing diary days during the interval}}$$

$$\frac{\text{Sum of doses of oral triptan during the interval} * 7}{\text{\# of nonmissing diary days during the interval}}$$

$$\frac{\text{Sum of doses of sumatriptan nasal spray during the interval} * 7}{\text{\# of nonmissing diary days during the interval}}$$

$$\frac{\text{Sum of doses of zolmitriptan nasal sprayduring the interval} * 7}{\text{\# of nonmissing diary days during the interval}}$$

5.4.1.3. Patient Global Impression of Improvement Endpoint

The PGI-I requests patients to mark the box that best describes their cluster headache condition since they started taking this medicine. The options in the displayed boxes are represented on a 7-point scale, with 1=very much better and 7=very much worse (Guy 1976).

The patient-reported PGI-I information will be captured at office visits. If the PGI-I collection date is greater than 10 days from the visit date, the record will not be used for analysis.

5.4.2. Safety Endpoints

Safety endpoints consist of the incidences of TEAEs, SAEs, and discontinuations due to adverse events (AEs), vital signs (blood pressure, pulse, and body temperature), weight, suicidal ideation and behaviors assessed by solicited questioning using the C-SSRS, ECGs, and laboratory measures (chemistry, hematology, and urinalysis).

5.4.3. Immunogenicity Endpoints

Immunogenicity endpoints consist of the incidences of antibodies to LY2951742 (ADA). An additional endpoint is the incidence of NAb present in those trial participants with ADA.

CCI

5.4.5. Pharmacokinetic Assessment

Pharmacokinetic assessment will be summarized in the pharmacokinetic/pharmacodynamic (PK/PD) SAP.

5.5. Statistical Analyses

The protocol for this study was approved on 18 December 2014. Protocol amendment (a) for this study was approved on 12 February 2015. Protocol amendment (b) for this study was approved on 22 December 2015. Protocol amendment (c) for this study was approved on 10 February 2017. The SAP Version 4 supersedes the statistical plans described in the protocol and previous versions of the SAP.

5.5.1. General Considerations

General aspects of statistical analyses are described below.

Unless otherwise specified, efficacy analyses during SP III will be conducted on an **intent-to-treat (ITT) population**, which include all patients who are randomized and receive at least 1 dose of study drug. Patients in the ITT population will be analyzed according to the treatment group that they were randomized to. Safety analyses during SP III will be conducted on the **safety population,** which also includes all patients who are randomized and receive at least 1 dose of study drug. However, patients will be analyzed by actual study treatment received most often (modal treatment) during the double-blind treatment phase. Modal treatment will be the same as randomized treatment except in some cases of incorrect treatment administration. When mean change from baseline is assessed, the patient will be included in the analysis only if he or she has a baseline and a postbaseline measurement.

Unless otherwise specified, for analyses of posttreatment phase, the **posttreatment population** will be used. **Posttreatment population** will be defined as all patients who entered the post-treatment phase (SP IV) as indicated by entering any posttreatment visit.

Statistical analysis will be carried out for the 8-week treatment phase (SP III), the 16-week post treatment phase (SP IV) as well as the 8-week treatment and 16-week post treatment phases combined (SP III/IV) as listed in Table CGAL.5.2.

Safety analyses (Section 5.5.9) in SP III and SP IV and analyses for exposure will be conducted based on the modal treatment group patients have received (placebo or GMB300 mg). For determining modal treatment, if there are 2 modes, then the modal treatment group will be GMB300 mg.

Unless otherwise specified, for the analyses in SP IV alone, no statistical comparisons between any treatment groups will be conducted.

Since cluster headache attack information collected through ePROs will only be collected for SP III, analyses of ePRO data will be conducted at SP III only.

Treatment effects will be evaluated based on a 2-sided significance level of 0.05 for all the other efficacy and safety analyses. The 95% confidence intervals for the difference in least-square means (LSMeans) between treatment groups will be presented. Adjustments for multiple comparisons for the analyses corresponding to the primary and gated secondary objectives are described in the sections on the primary and secondary efficacy analyses below. There will be no adjustments for multiplicity for analyses of other data.

A repeated measures analysis refers to a restricted maximum likelihood (REML)-based, mixed-effects repeated measures (MMRM) analysis using all the longitudinal observations at each postbaseline visit/week.

Categorical comparisons between treatment groups for safety measures will be performed using Fisher's exact tests, where appropriate.

Any change to the data analysis methods described in the protocol will require an amendment ONLY if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol, and the justification for making the changes, will be described in the SAP and/or in the clinical study report (CSR).

Additional exploratory analyses of the data will be conducted as deemed appropriate.

Statistical analysis of this study will be the responsibility of Eli Lilly and Company (Lilly) or designee. SAS® software will be used to perform most or all statistical analyses.

5.5.1.1. Adjustments for Covariates

The repeated measures models will include the fixed, categorical effects of treatment, gender, pooled investigative site, visit/week, and treatment-by-visit/week interaction, as well as the continuous, fixed covariates of baseline value. Rules for pooling of investigative sites are described in Section 5.5.1.3. Note, in repeated measures analysis, visit will be used for measures collected at visit interval, while week will be used for all the ePRO data.

When an ANOVA model is used to analyze a continuous efficacy variable, the model will contain the main effects of treatment, gender and pooled investigative site, and appropriate baseline value included as a covariate.

The categorical, pseudo-likelihood-based repeated measures models for the visit wise/week wise binary outcomes of response will include the fixed, categorical effects of treatment, gender, visit/week, and treatment-by-visit/week interaction, as well as the continuous, fixed covariate of baseline value. Pooled investigative site was not included in the model in order to increase the likelihood of convergence.

With the exception of efficacy analyses on cluster headache attack frequency or categorical analysis of response rate (such as 50% response rate) derived from cluster headache attack frequency where the continuous value of baseline weekly cluster headache frequency will be used as covariate, all other efficacy analyses will include baseline average daily cluster headache

attack frequency category (≤4 vs. >4) as a covariate in the MMRM, GLIMMIX, and ANOVA model.

5.5.1.2. Handling of Dropouts or Missing Data

Repeated measures analyses will be used as the statistical approach for handling missing data. The model parameters are simultaneously estimated using restricted likelihood estimation incorporating all of the observed data. Estimates have been shown to be unbiased when the missing data are missing at random and when there is ignorable non-random missing data (Mallinckrodt 2008). Missing at random (MAR) assumption will be evaluated using sensitivity analyses as defined in Section 5.5.11.

Approaches for Handling Missing Data for Derivation of Cluster Headache Attacks Derived from ePRO per 7-Day Interval

In Study CGAL, to derive the number of cluster headache attacks per 7-day interval, the daily data for each patient (including the last 7 days in the eligibility report [pre-randomization diary phase] and the 8 weeks of daily data during treatment) will be converted into 9 roughly 7-calendar day intervals: the baseline 7-day period, Weeks 1, 2, 3, 4, 5, 6, 7, and 8. Each day, the patient may have zero, 1, or multiple cluster headache attacks. For each weekly interval, the following missing data imputation method will be used:

- 1) if there are less than or equal to 3 days with nonmissing answer to cluster headache attack frequency in the weekly interval; or 2) the diary compliance rate is less than or equal to 50%, then the weekly interval will be considered missing;
- Otherwise, 1) if there are greater than or equal to 4 days with nonmissing answer to cluster headache attack frequency in the weekly interval; and 2) the diary compliance rate is greater than 50%, then the average number of cluster headache attacks across the non-missing days will be used to impute the missing days.

For detailed example about missing data imputation, please see Section 5.4.1.2.

Then the change from baseline to Weeks 1, 2, 3, 4, 5, 6, 7, and 8 will be derived.

The same approach will also be applied to secondary and exploratory efficacy measures that derived from ePRO data.

5.5.1.3. Multicenter Studies

The following investigative site pooling method will be used:

All investigative sites with fewer than 2 randomized patients per each treatment group with non-missing cluster headache attacks during baseline interval and at least 1 postbaseline value will be pooled together within each country and considered a single site for analyses. If this results in a pooled site still having fewer than 2 randomized patients per each treatment group, the pooled site will also be pooled with the next smallest site in that country, determined to be the site with the smallest number of randomized patients, or if more than 1 site meets that criterion, the smallest site with the lowest investigator number. If this results in a pooled site still having

fewer than 2 patients randomized to each treatment arm, these sites will be pooled together with the next smallest site in the geographic region. Two geographic regions are defined, including US and Canada combined, as well as Europe. If this still results in a site having fewer than 2 patients randomized to each treatment, then these sites will be pooled together with the next smallest site in the whole study.

All analyses will use pooled investigative sites. The actual investigative site numbers will be included in the listings.

5.5.1.4. Multiple Comparisons/Multiplicity

The primary efficacy analysis will be the overall treatment effect of GMB300 mg every 30 days versus placebo over Weeks 1 to 3 using a MMRM analysis, which is equivalent to the average of the MMRM-estimated weekly treatment effect over the 3-week period for change in weekly cluster headache attack frequency from baseline. The Type I error rate will be controlled at a 1-sided 0.025 level for the primary efficacy analysis.

A fixed sequential gatekeeper method will be utilized for testing secondary hypotheses to be eligible for inclusion in the proposed label. Specific details of the testing of the secondary gatekeeper objectives are provided in Section 5.5.8.2.

5.5.1.5. Analysis Populations

Three analysis populations, including the ITT population, the safety population, and the posttreatment population, are defined in Section 5.5.1.

5.5.1.6. Baseline and Postbaseline Definition

Table CGAL.5.2 describes the rules for determining the patient population and baseline and postbaseline observations for each study phase and type of analysis. When “last of Visit x-x” is used in the table, the last nonmissing observation obtained in the visit interval will be used.

Table CGAL.5.2. Patient Population with Baseline and Postbaseline Definitions by Study Phase and Type of Analysis

Study Phase/Analysis	Patient Population	Baseline Observation	Postbaseline Observation(s)
Study Phase III			
Continuous secondary efficacy analyses (Repeated measures)	Patients in ITT population with a baseline and at least 1 postbaseline observation	Visit 3	All Visits 4-7
TEAEs	Safety population	All Visits 1-3 before dosing	Visit 3 after dosing through Visit 7
SAEs, Discontinuations due to AEs	Safety population	NA	Visit 3 after dosing through Visit 7
C-SSRS categorical analyses	Patients in safety population with a baseline and at least one postbaseline C-SSRS assessment	Recent history: All Visits 1-3 excluding lifetime[a] All prior history: Visits 1- 3 including lifetime[a]	All Visits 3.01-7
TE abnormal laboratory values	Patients in safety population with normal laboratory values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least 1 postbaseline observation	All Visits 1-3	All Visits 3.01-7
TE immunogenicity	Patients in safety population who are evaluable for TE ADA	Visit 3	All Visits 3.01-7
TE changes in vital signs and weight, ECG parameters	Patients in safety population with a baseline and at least 1 postbaseline observation	Last nonmissing value from Visits 1-3 for BP, pulse, and ECG All Visits 1-3 for weight and temperature	Visits 3.01-7

Patient Population with Baseline and Postbaseline Definitions by Study Phase and Type of Analysis

Study Phase/Analysis	Patient Population	Baseline Observation	Postbaseline Observation(s)
Continuous safety analysis of vital signs, weight, laboratory, and ECG parameters (Box-whisker plot)	Safety population	Last non-missing value from Visits 1-3	Visits 4-7
Study Phase III and IV Combined			
TE immunogenicity	Patients in safety population who are evaluable for TE ADA	Visit 3	All Visits 3.01–9
Study Phase IV			
Post TEAEs	Posttreatment population	All Visits 1-7	All Visits 7.01-9
SAEs, Discontinuations due to AEs	Posttreatment population	NA	All Visits 7.01-9
C-SSRS categorical analyses	Patients in posttreatment population with a baseline and at least 1 postbaseline C-SSRS assessment	Recent history: All Visits 1-7 excluding lifetime[a] All prior history: Visits 1-7 including lifetime[a]	All Visits 7.01-9
Post TE abnormal laboratory values	Patients in posttreatment population with normal laboratory values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least 1 postbaseline observation	All Visits 1-7	All Visits 7.01-9

Patient Population with Baseline and Postbaseline Definitions by Study Phase and Type of Analysis

Study Phase/Analysis	Patient Population	Baseline Observation	Postbaseline Observation(s)
Post TE changes in vital signs and weight, ECG parameters	Patients in posttreatment population and with a baseline and at least 1 postbaseline observation	Last nonmissing value from Visits 1-7 for BP, pulse and ECG All Visits 1-7 for weight and temperature	7.01-9

Abbreviations: ADA = anti-drug antibody; AE = adverse event; BP = blood pressure; C-SSRS = Columbia Suicide Severity Rating Scale; ECG = electrocardiogram; ITT = intent-to-treat; NA = not applicable; SAE = serious adverse event; TE = treatment emergent; TEAE = treatment-emergent adverse event.

Note: Visit 3.01 indicates the first unscheduled visit occurring after Visit 3 and prior to Visit 4.

[a] Lifetime is captured in the C-SSRS Visit 1 case report form.

5.5.2. Patient Disposition

The number and percentage of ITT patients who complete the study or discontinue early will be tabulated for all treatment groups for SP III and SP IV both overall and by visit. Reasons for discontinuation will be compared between treatment groups using Fisher's exact test for SP III with the ITT population. Descriptive statistics only will be presented for the treatment groups in SP IV. For patients who were randomized without drug injection, reasons for early discontinuation will be provided in a listing.

Patient allocation by investigator will be summarized for SP III for all ITT patients.

Patient allocation by investigator will also be listed for all SPs.

5.5.3. Important Protocol Deviations

Important protocol deviations that potentially compromise the data integrity and patients' safety will be summarized by treatment for the ITT population.

Section 7 (appendix) lists the categories, subcategories, study-specific terms of important protocol deviations, and source of identification. Per study team's discretion, for non-programmable protocol deviation, additional categories and subcategories other than the ones on Section 7 can always be added into the final nonprogrammable protocol deviations list as deemed necessary.

Tables and listings of subjects with important protocol deviations will be provided for the ITT population.

5.5.4. Patient Characteristics

The following patient characteristics at baseline will be summarized by treatment group for all ITT patients:

- demographic (age, gender, race, ethnicity, country, region, height, weight, body mass index)
- baseline disease characteristics:
- number of weekly cluster headache attacks
- average severity of cluster headache pain for the cluster headache attack days
- weekly total cluster headache attack duration
- weekly number of times using oxygen
- weekly number of times using oral triptan, sumatriptan nasal spray, or zolmitriptan nasal spray
- weekly number of times using sumatriptan Sc
- weekly number of times using acetaminophen/paracetamol or NSAIDs
- number of times using oxygen per cluster headache attack

- number of times using oral triptan, sumatriptan nasal spray, or zolmitriptan nasal spray per cluster headache attack
- number of times of using sumatriptan Sc per cluster headache attack
- number of times using acetaminophen/paracetamol or NSAIDs per cluster headache attack
- total weekly dose for oral triptan, sumatriptan nasal spray, and zolmitriptan nasal spray combined
- total weekly dose for sumatriptan Sc
- total weekly dose for oral triptan
- total weekly dose for sumatriptan nasal spray
- total weekly dose for zolmitriptan nasal spray
- prior cluster headache history in last 7 days prior to Visit 1
- baseline alcohol, tobacco, caffeine, and nicotine consumption
- medical history and pre-existing conditions

Comparisons between treatment groups will be performed using Fisher's exact tests for categorical data and analysis of variance (ANOVA) with treatment and pooled investigative site as independent variables in the model for continuous data.

Medical history and pre-existing conditions will be summarized by descending frequency of preferred term (PT) within system organ class (SOC), and by descending frequency of PT respectively, and comparison between treatment groups will be performed using Fisher's exact test. Medical history is defined as illness(es) that ended prior to the signing of informed consent. Pre-existing conditions are medical events ongoing at the time of informed consent.

5.5.5. Exposure to Investigational Product

Patients will receive the investigational product (IP) at the following planned time points:

- Week 0 (Visit 3)
- Week 4 (Visit 5)

The following information will be recorded on the electronic case report form (eCRF) for each dose:

- confirmation that the patient received the IP
- date and time of administration

5.5.5.1. Duration of Exposure

From the information recorded on the eCRF, the following will be derived:

- Duration of exposure in days calculated as Treatment End Date (disposition date in SP III) – First date IP administered + 1
- Number and percentage of patients with 1 full dose or 2 full doses administered

Comparisons between treatments using safety population for duration of exposure will be performed using an ANOVA with treatment and pooled investigative site in the model.

The number of full doses will also be summarized.

5.5.5.2. Treatment Compliance

Treatment compliance will be calculated as follows

(number of full doses received)/(Number of intended full doses)*100

Note, full dose means that patients have to receive all 3 injections. For patients that are early discontinued, number of intended full doses will only include scheduled doses prior to discontinuation. Comparisons between treatments in the ITT population for treatment compliance will be performed using an ANOVA with treatment and pooled investigative site in the model.

5.5.6. Electronic Patient-Reported Outcome Diary Compliance

Electronic patient-reported outcome diary compliance at each weekly interval (including baseline, Weeks 1, 2, 3, 4, 5, 6, 7, and 8) will be calculated. Diary compliance at each interval is calculated as follows:

$$\frac{\text{Actual number of diary entry days in the interval} * 100}{\text{Expected number of diary entry days in the interval}}$$

The diary entry can only be saved and submitted after all the required ePRO questions are answered, so the actual number of diary entry days represents the total number of days with non-missing answer to all the required cluster headache attack ePRO questions.

The expected number of diary entry days is calculated as the (last calendar date - the first calendar date in each interval + 1).

Comparisons between diary compliance for each interval separately will be performed using an ANOVA with treatment and pooled investigative site in the model.

Compliance will also be listed by weekly interval for each patient.

5.5.7. Concomitant Therapy

The proportion of patients who received concomitant medication collected from eCRF will be summarized by PT separately for all ITT patients for both SP III and SP IV. Abortive medications for cluster headache attack collected through ePRO diary will be summarized separately by PT for all ITT patients for SP III. If there are different PTs for salt forms of an abortive medication, these PTs will be combined for the medication in the summary. Concomitant therapies for SP III are those that stopped during SP III or continued in SP III. If medication started and stopped on the same day of first injection, it will still be considered as concomitant medication for SP III. If a medication started before the first day of injection but

stopped on the same day of injection, then it will not be counted as concomitant medication for SP III. Concomitant therapies for SP IV are those that either started, stopped, or continued in SP IV.

Treatment group comparisons will be done using Fisher's exact test for SP III with the ITT population. Descriptive statistics only will be presented for the treatment groups in SP IV with posttreatment population.

5.5.8. Efficacy Analyses

5.5.8.1. Primary Outcome and Methodology

The primary analysis will be conducted by a REML-based, MMRM analysis using all the longitudinal observations from Weeks 1 to 3. The analysis of the primary outcome will be the main effect of treatment between GMB 300 mg and placebo across Weeks 1 to 3 of the treatment phase from a repeated measures analysis on mean change from baseline in the weekly attack frequency. This provides the average treatment effect over the 3-week period. Baseline is defined as the last 7 days in the eligibility report (prerandomization diary phase). In addition to the primary endpoint results, the mean profiles for GMB300mg and placebo over the 3-week period will also be reported from the MMRM.

The MMRM model for the primary analysis only uses the first 3 weeks of double-blind treatment phase data to avoid impact of the data from later weeks when patients may start to get into the remission period.

The model for the primary analysis will include the fixed, categorical effects of treatment, gender, pooled investigative site, week, and treatment-by-week interaction, as well as the continuous, fixed covariates of baseline value. An unstructured covariance structure will be used to model the within-patient errors. The Kenward-Roger (Kenward and Roger 1997) approximation will be used to estimate denominator degrees of freedom. If the model does not converge with both the Hessian and the G matrix being positive definite under the default fitting algorithm used by PROC MIXED, the Fishers' scoring algorithm will be implemented by specifying the SCORING option in SAS®. If the model still fails to converge, the model will be fit using covariance matrices of the following order specified by a decreasing number of covariance parameters until convergence is met:

- heterogeneous Toeplitz
- heterogeneous first-order autoregressive
- Toeplitz
- first-order autoregressive

If necessary, both the default and the scoring fitting algorithms will be used in the prespecified order before proceeding to the next covariance structure in the sequence. For models where the unstructured covariance matrix is not utilized, the sandwich estimator (Diggle et al. 1994) will be used to estimate the standard errors of the fixed effects parameters. The sandwich estimator is

implemented by specifying the EMPIRICAL option in SAS®. When the sandwich estimator is utilized, the Kenward-Roger approximation for denominator degrees of freedom cannot be used. Instead, the denominator degrees of freedom will be partitioned into between-subject and within-subject portions by the DDFM=BETWITHIN option in SAS®. SAS® PROC MIXED will be used to perform the analysis.

5.5.8.2. Gated Secondary

The gated secondary outcome, 50% response, is the proportion of patients meeting the response criteria at Week 3. A nonresponder imputation for missing values will be used. Specifically, all patients who discontinue study treatment at any time prior to Week 3, for any reason, will be considered a nonresponder.

Treatment differences in the proportions of patients meeting 50% response definition at Week 3 will be determined using Koch's Nonparametric Randomization-Based Analysis of Covariance method (Koch et al. 1998). This method will adjust for pooled investigative site by including it as a stratification variable and will also adjust for the continuous baseline value and gender. A SAS/IML macro (NParCov3) (Zink and Koch 2012) will be used for the calculation. The options with this SAS/IML macro are specified in the example SAS code below.

%NPARCOV3(outcomes =[response]

covars = [baseline] [gender]

trtgrps = [treatment]

strata = [PINVID]

hypoth = NULL

transform = NONE

combine = FIRST

c = 1

dsnin = [input]

dsnout = [output])

In this method, the option of "hypoth=NULL" indicates that the variance covariance structure will be calculated under the assumption that the means and covariance matrices of the treatment groups are equal and therefore computes a single covariance matrix for each stratum. The option of "combine=FIRST" indicates that the covariate adjustment will be performed after a weighted average of treatment group differences across pooled investigative sites to account for the possibility of small numbers of patients at some sites. The option of "c=1" indicates the use of Mantel-Haenszel weights for each pooled investigative site. The option of "transform=NONE" indicates that there is not a transformation of the data.

The analysis result of the secondary gatekeeper objective will be evaluated if the placebo versus GMB300mg comparison is significant for the primary efficacy analysis at a one sided α=0.025 significance level.

5.5.8.3. Additional Secondary and Exploratory Efficacy Analyses

Table CGAL.5.3 summarizes all the planned additional secondary efficacy analyses for SP III and SP IV.

For the continuous additional secondary and exploratory efficacy measures, the change from baseline to each postbaseline period will be estimated for each treatment from repeated measures analyses as described in Section 5.5.8.2. The treatment comparison at each week and overall across 8 weeks will be provided. As discussed in Section 5.5.1.1, for the efficacy measures that are not derived from cluster headache frequency, the baseline average daily cluster headache attack frequency category (≤4 vs. >4) will be added as a covariate in the MMRM model.

For the categorical additional secondary and exploratory efficacy measures including 30%, 50%, 75% response, and 100% response, the percentage of patients meeting response criteria at each period will be estimated for each treatment from a categorical, pseudo-likelihood-based repeated measures analysis of visit wise binary outcomes indicating whether patients meet response criteria. The treatment comparison at each week and overall across 8 weeks will be provided. This analysis will be implemented using the GLIMMIX procedure in SAS to compare treatments and include the fixed, categorical effects of treatment, gender, visit/week, and treatment-by-visit/week interaction, as well as the continuous, fixed covariate of baseline value. An unstructured covariance structure will be used to model the within-patient errors (denoted by TYPE=CHOL in the RANDOM statement). The Newton-Raphson method with ridging will be used for nonlinear optimization (denoted by including NLOPTIONS TECH=NRRIDG). The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. If the model does not converge, the Fishers scoring algorithm will be utilized by the SCORING option in SAS. If the model still fails to converge, the model will be fit using covariance matrices in the following order specified by a decreasing number of covariance parameters until convergence is met: heterogeneous Toeplitz, heterogeneous autoregressive, Toeplitz, and autoregressive. If necessary, both fitting algorithms will be used in the prespecified order before proceeding to the next covariance structure in the sequence. For models where the unstructured covariance matrix is not utilized, the sandwich estimator (Diggle et al. 1994) will be used to estimate the standard errors of the fixed effects parameters. The sandwich estimator is utilized by the EMPIRICAL option in SAS. When the sandwich estimator is utilized, the Kenward-Roger approximation for denominator degrees of freedom cannot be used. Instead, the denominator degrees of freedom will be partitioned into between-subject and within-subject portions by the DDFM=BETWITHIN option in SAS. As discussed in Section 5.5.1.1, for the efficacy measures that are not derived from cluster headache frequency, the baseline average daily cluster headache attack frequency category (≤4 vs. >4) will be added as a covariate in the GLIMMIX model.

Table CGAL.5.3. Other Secondary and Exploratory Efficacy Variables and Their Derivation

Study Phase III	Study Phase IV	Efficacy Variable	Analyses
Change from baseline to each 7-day interval (Weeks 1, 2, 3, 4, 5, 6, 7, and 8)	No planned analysis	1. Weekly cluster headache attack frequency 2. Weekly average cluster headache attack pain severity in the remaining cluster headache attack days 3. Weekly total cluster headache attack duration 4. Weekly number of times using oxygen 5. Weekly number of times using oral triptan, sumatriptan nasal spray, or zolmitriptan nasal spray 6. Weekly number of times using sumatriptan Sc 7. Weekly number of times using acetaminophen/paracetamol or NSAIDs 8. Number of times using oxygen per cluster headache attack 9. Number of times using oral triptan, sumatriptan nasal spray, or zolmitriptan nasal spray per cluster headache attack 10. Number of times using sumatriptan Sc per cluster headache attack 11. Number of times using acetaminophen/paracetamol or NSAIDs per cluster headache attack 12. Total weekly dose of oral sumatriptan, sumatriptan nasal spray and zolmitriptan nasal spray combined 13. Total weekly dose of sumatriptan Sc 14. Total weekly dose of oral sumatriptan 15. Total weekly dose of sumatriptan nasal spray 16. Total weekly dose of zolmitriptan nasal spray	Variables will be analyzed by a repeated measures analysis using a model as described in Section 5.5.8.1 and Section 5.5.8.3.

Other Secondary and Exploratory Efficacy Variables and Their Derivation

Study Phase III	Study Phase IV	Efficacy Variable	Analyses
Value at each visit (Visits 5, 7, corresponding to Month 1, 2)	Value at each visit (Visit 9, corresponding to Month 6)	1. PGI-I Score	For SP III, the variable will be analyzed by a repeated measures analysis using a model as described in Section5.5.8.3. without baseline covariate. For SP IV, the variable will be analyzed by an ANOVA model as described in Section 5.5.1.1.
Categorical variables at each 7 day period (Weeks 1, 2, 3, 4, 5, 6, 7, and 8)	**No planned analyses**	1. 30% response 2. 50% response 3. 75% response 4. 100% response 5. 30% reduction in weekly total cluster headache attack duration	For all variables, the visit wise percentages of patients meeting criteria will be compared between treatments using a categorical, repeated measures analysis described in Section 5.5.8.3.
Categorical variables at each visit (Visits 5, 7, corresponding to Month 1, 2)	**Categorical variables at each visit (Visit 9, corresponding to Month 6)**	1. Proportion of patients reporting a score of 1 (“very much improved”) or 2 (“much improved”) on Patient Global Impression of Improvement (PGI-I)	For SP III, the visit wise percentages of patients meeting criteria will be compared between treatments using a categorical, repeated measures analysis described in this Section 5.5.8.3 but without baseline value covariate. For SP IV, the Koch’s Nonparametric Randomization Based Analysis of Covariance method as described in Section 5.5.8.2 will be used, but adding the baseline average daily cluster headache attack frequency category (≤4 vs. >4) and removing baseline value covariate.

Abbreviations: ANOVA = analysis of variance; NSAIDs = nonsteroidal anti-inflammatory drugs; SP = study phase.

5.5.9. Safety Analyses

The safety analyses will be conducted for SP III in safety population and SP IV in posttreatment population.

The safety and tolerability of treatment will be assessed by summarizing the following:

- adverse events
  - treatment-emergent adverse events
    - by PT
    - by PT nested within SOC
    - by maximum severity
  - treatment-emergent adverse events by PT nested within SOC
  - adverse events leading to discontinuation by PT nested within SOC
  - adverse events of special interest
- suicide-related thoughts and behaviors by CSSRS
- vital signs and weight
- laboratory measurements
- electrocardiograms
- antibodies (ADA and Nab)

The baseline and postbaseline for all safety measures are described in Table CGAL.5.2 unless specified otherwise. For SAEs, only events with a start date during the postbaseline phase will be accounted for the corresponding study phase analysis.

5.5.9.1. Categorical Safety Variables

Unless specified otherwise, the categorical safety analyses will include both scheduled and unscheduled visits and be conducted for SP III and SP IV separately.

Comparisons between treatment groups for all categorical safety measures will be made using Fisher's exact test for SP III with the safety population. Descriptive statistics only will be presented for the treatment groups in SP IV with post-treatment population.

5.5.9.1.1. Adverse Events

Treatment-emergent AEs are defined as the reported AEs that first occurred or worsened during the postbaseline phase compared with baseline phase. For events occurring on the day of first administration of study drug, the CRF-collected flag will be used to determine whether the event was pretreatment versus posttreatment. For each TEAE, the severity level of the event (mild, moderate, or severe) will be determined by patient or physician opinion. The Medical Dictionary for Regulatory Activities (MedDRA) Lowest Level Term (LLT) will be used in the treatment-emergent computation. For each LLT, the maximum severity at baseline will be used as the baseline severity. If the maximum severity during postbaseline is greater than the maximum baseline severity, the event is considered to be treatment emergent for the specific postbaseline period. For events with a missing severity during the baseline period, it will be treated as "mild" in severity; for events with a missing severity during the postbaseline period, it will be treated as

"severe" for TEAE computation. For each patient and TEAE, the maximum severity for the MedDRA level being displayed (PT, High Level Term, or SOC) is the maximum postbaseline severity observed from all associated LLTs mapping to that MedDRA level.

For events that are gender specific, the denominator and computation of the percentage will be gender adjusted.

5.5.9.1.1.1. Potential Hypersensitivity Events

Potential hypersensitivity events will be defined using the following terms (standard MedDRA query [SMQ]):

- Broad and narrow terms in the Anaphylactic reaction SMQ (20000021)
- Broad and narrow terms in the Angioedema SMQ (20000024)
- Broad and narrow terms in the Hypersensitivity SMQ(20000214)

A listing of patients having an event identified from these analyses will be medically reviewed to determine if the terms identified represent events likely hypersensitivity in nature. Listings should include information on timing of event relative to latest dose of study drug administration, the event term from this query, other AEs for the patient and timing, any abnormal laboratory findings, concomitant medication, medical history, and pre-existing conditions. Only those that are judged medically to be events likely hypersensitivity in nature will be included in the final tables.

The number and percentage of patients with potential and/or likely TEAEs will be summarized by treatment groups using MedDRA PT nested within the SMQ. Events will be ordered by decreasing frequency within the SMQ. The number and percentage of patients with likely hypersensitivity SAEs and AEs resulting in study drug discontinuation will be presented by treatment groups using MedDRA PT and ordered by decreasing frequency by PT.

The number and percentage of patients with likely hypersensitivity TEAEs by maximum severity will be summarized by treatment groups using MedDRA PT.

The number and percentage of patients with likely hypersensitivity TEAEs by timing will be summarized using MedDRA PT. Events will be ordered by decreasing frequency of PT. Note the timing of the likely hypersensitivity events is collected through eCRF and categorized into the following 4 categories:

1. Immediate: occurs within minutes (<60 minutes) from study drug administration
2. Acute Reaction: occurs from 1 up to 6 hours from study drug administration
3. Delayed Reaction: occurs from >6 hours through 14 days from study drug administration, which will be split into 2 categories: on the same day of injection and after the day of injection
4. Reaction >14 days

5.5.9.1.1.2. Adverse Events Related to Injection Sites

Adverse events related to injection sites will be defined using terms from the MedDRA High Level Term injection site reactions.

The number and percentage of patients with TEAEs related to injection sites, SAEs related to injection sites, and AEs related to injection sites resulting in study drug discontinuation will be summarized using MedDRA PT. Events will be ordered by decreasing frequency of PT term.

The number and percentage of patients with TEAEs related to injection sites by maximum severity will be summarized by treatment groups using MedDRA PT. For each patient and injection site related event, the maximum severity for the MedDRA level being displayed (PT) is the maximum postbaseline severity observed from all associated LLTs mapping to that MedDRA level.

The number and percentage of patients with TEAEs related to injection sites by timing will be summarized using MedDRA PTs ordered by decreasing frequency. Note the timing of AEs related to injection sites is collected through eCRF and categorized into the same categories as for hypersensitivity events.

5.5.9.1.1.3. Upper Respiratory Tract Infections

Upper respiratory tract infections will be defined using all the PTs from the 2 High Level Terms of "upper respiratory tract infections" and "upper respiratory tract infections NEC" as defined in MedDRA. The number and percentage of patients with TEAEs of upper respiratory tract infections will be summarized by treatment group using MedDRA PTs. Events will be ordered by decreasing frequency in the galcanezumab 300 mg group.

The number and percentage of patients with TEAEs of upper respiratory tract infections by maximum severity will be summarized by treatment group using MedDRA PTs. For each patient and upper respiratory tract infection event, the maximum severity for the MedDRA level being displayed (PT) is the maximum postbaseline severity observed from all associated LLTs mapping to that MedDRA level.

By-subject listings of treatment-emergent upper respiratory tract infections and upper respiratory tract infections leading to study drug discontinuation will be provided.

5.5.9.1.2. Suicide-Related Thoughts and Behaviors

Postbaseline suicidal ideation, suicidal behavior, and self-injurious behavior without suicidal intent occurring during SP III, based on the C-SSRS, will be summarized by treatment. In particular, for each of the following events, the number and percent of patients with the event will be enumerated by treatment: completed suicide, nonfatal suicide attempt, interrupted attempt, aborted attempt, preparatory acts or behavior, active suicidal ideation with specific plan and intent, active suicidal ideation with some intent to act without specific plan, active suicidal ideation with any methods (no plan) without intent to act, nonspecific active suicidal thoughts, wish to be dead, and self-injurious behavior without suicidal intent. These measures will also be summarized for SP IV.

In addition, the number and percent of patients who experienced at least 1 of various composite measures during SP III and SP IV separately will be presented and compared. These include suicidal behavior (completed suicide, nonfatal suicidal attempts, interrupted attempts, aborted attempts, and preparatory acts or behavior), suicidal ideation (active suicidal ideation with specific plan and intent, active suicidal ideation with some intent to act without specific plan, active suicidal ideation with any methods [no plan] without intent to act, nonspecific active suicidal thoughts, and wish to be dead), and suicidal ideation or behavior.

The number and percent of patients who experienced at least 1 of various comparative measures during treatment will be presented and compared for SP III and SP IV. These include treatment-emergent suicidal ideation compared to recent history, treatment-emergent serious suicidal ideation compared to recent history, emergence of serious suicidal ideation compared to recent history, improvement in suicidal ideation at endpoint compared to baseline, and emergence of suicidal behavior compared to all prior history.

Specifically, the following outcomes are C-SSRS categories and have binary responses (yes/no). The categories have been re-ordered from the actual scale to facilitate the definitions of the composite and comparative endpoints, and to enable clarity in the presentation of the results.

Category 1 – Wish to be Dead

Category 2 – Nonspecific Active Suicidal Thoughts

Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act

Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan

Category 5 – Active Suicidal Ideation with Specific Plan and Intent

Category 6 – Preparatory Acts or Behavior

Category 7 – Aborted Attempt

Category 8 – Interrupted Attempt

Category 9 – Actual Attempt (Non-Fatal)

Category 10 – Completed Suicide

Self-injurious behavior without suicidal intent is also a C-SSRS outcome (although not suicide related) and has a binary response (yes/no).

Composite endpoints based on the above categories are defined below.

- Suicidal ideation: A “yes” answer at any time during treatment to any 1 of the 5 suicidal ideation questions (Categories 1 to 5) on the C-SSRS.
- Suicidal behavior: A “yes” answer at any time during treatment to any 1 of the 5 suicidal behavior questions (Categories 6 to 10) on the C-SSRS.
- Suicidal ideation or behavior: A “yes” answer at any time during treatment to any 1 of the 10 suicidal ideation and behavior questions (Categories 1 to 10) on the C-SSRS.

The following outcome is a numerical score derived from the C-SSRS categories. The score is created at each assessment for each patient and is used for determining treatment emergence.

- Suicidal Ideation Score: The maximum suicidal ideation category (1 to 5 on the C-SSRS) present at the assessment. Assign a score of 0 if no ideation is present.

For SP III and SP IV only, comparative endpoints of interest are defined below. "Treatment emergence" is used for outcomes that include events that first emerge or worsen. "Emergence" is used for outcomes that include events that first emerge.

- Treatment-emergent suicidal ideation compared to recent history:
  An increase in the maximum suicidal ideation score during treatment (Visits 3.01 to 7 for SP III; Visits 7.01 to 9 for SP IV) from the maximum suicidal ideation category during the screening and lead-in periods (C-SSRS scales taken at Visits 1 to 3 excluding "lifetime" for SP III; C-SSRS scales taken at Visits 1 to 7 excluding "lifetime" for SP IV).
- Treatment-emergent suicidal ideation compared to all prior history:
  An increase in the maximum suicidal ideation score during treatment (Visits 3.01 to 7 for SP III; Visits 7.01 to 9 for SP IV) from the maximum suicidal ideation category prior to treatment (C-SSRS scales taken at Visits 1 to 3 including "lifetime" for SP III; C-SSRS scales taken at Visits 1 to 7 including "lifetime" for SP IV).
- Treatment-emergent serious suicidal ideation compared to recent history:
  An increase in the maximum suicidal ideation score to 4 or 5 on the C-SSRS during treatment (Visits 3.01 to 7 for SP III; Visits 7.01 to 9 for SP IV) from not having serious suicidal ideation (scores of 0 to 3) during the screening and lead-in periods (C-SSRS scales taken at Visits 1 to 3 excluding "lifetime" for SP III; C-SSRS scales taken at Visits 1 to 7 excluding "lifetime" for SP IV).
- Treatment-emergent serious suicidal ideation compared to all prior history:
  An increase in the maximum suicidal ideation score to 4 or 5 on the C-SSRS during treatment (Visits 3.01 to 7 for SP III; Visits 7.01 to 9 for SP IV) from not having serious suicidal ideation (scores of 0 to 3) prior to treatment (C-SSRS scales taken at Visits 1 to 3 including "lifetime" for SP III; C-SSRS scales taken at Visits 1 to 7 including "lifetime" for SP IV). Emergence of serious suicidal ideation compared to recent history:
- An increase in the maximum suicidal ideation score to 4 or 5 on the C-SSRS during treatment (Visits 3.01 to 7 for SP III; Visits 7.01 to 9 for SP IV) from no suicidal ideation (scores of 0) during the screening and lead-in periods (C-SSRS scales taken at Visits 1 to 3 excluding "lifetime" for SP III; C-SSRS scales taken at Visits 1 to 7 excluding "lifetime" for SP IV). Recent history excludes "lifetime" scores from the Baseline C-SSRS scale or Baseline/Screening C-SSRS scale.
- Emergence of serious suicidal ideation compared to all prior history:
  An increase in the maximum suicidal ideation score to 4 or 5 on the C-SSRS during treatment (Visits 3.01 to 7 for SP III; Visits 7.01 to 9 for SP IV) from no suicidal ideation (scores of 0) prior to treatment (C-SSRS scales taken at Visits 1 to 3 including "lifetime" for SP III; C-SSRS scales taken at Visits 1 to 7 including "lifetime" for SP IV).

- Improvement in suicidal ideation at endpoint compared to baseline:
  A decrease in suicidal ideation score at endpoint (the last measurement during treatment; Visits 3.01 to 7 for SP III; Visits 7.01 to 9 for SP IV) from the baseline measurement (the measurement taken just prior to that study phase (last nonmissing value taken at Visit 2 to Visit 3 for SP III; last nonmissing value taken at Visit 3.01 to Visit 7 for SP IV).
- Emergence of suicidal behavior compared to all prior history:
  The occurrence of suicidal behavior (Categories 6 to 10) during treatment (Visits 3.01 to 7 for SP III; Visits 7.01 to 9 for SP IV) from not having suicidal behavior (Categories 6 to 10) prior to treatment (Visits 1 to 3 including "lifetime" for SP III; C-SSRS scales taken at Visits 1 to 7 including "lifetime" for SP IV).

Patients who discontinued from the study with no postbaseline C-SSRS value will be considered unevaluable for analyses of suicide-related events. Only evaluable patients will be considered in the analyses. Fisher's exact test will be used for treatment comparisons in SP III.

5.5.9.1.3. Vital Signs and Weight

Vital signs collected during the study include systolic and diastolic blood pressure (SBP/DBP), pulse, and temperature. Blood pressure and pulse measurements will be taken when the patient is in a sitting position. Three measurements of sitting blood pressure and pulse will be collected at approximately 30- to 60-second intervals at every visit and the 3 sitting blood pressure measurements and 3 pulse values will be averaged and used as the value for that visit.

Table CGAL.5.4 displays the criteria used to define treatment emergent, potentially clinically significant changes and sustained elevation in vital signs and weight. The last column of the table displays the patient populations for each analysis based on baseline categories. The number and percent of patients meeting these criteria will be summarized. Treatment group comparisons will be performed using Fisher's exact test for SP III.

The criteria to identify patients with treatment-emergent abnormal changes generally consist of 2 parts: an absolute threshold and a change from baseline amount.

- The absolute threshold in the criteria is based on 1) minimum postbaseline when the direction is low and 2) maximum postbaseline when the direction is high.
- The change from baseline amount in the criteria is 1) decrease from baseline (defined below and in Table CGAL.5.2) to minimum postbaseline when the direction is low; 2) increase from baseline (defined below and in Table CGAL.5.2) to maximum postbaseline when the direction is high.

The baseline for SBP, DBP, and pulse is defined as the last nonmissing value during the baseline period (See Table CGAL.5.2). To be exact,

- The baseline for SBP, DBP, and pulse is defined as the last nonmissing value before randomization. The rationale for using the last available value in the baseline period is to minimize the potential confound of discontinuing or dose stabilization of medications that modulate BP and pulse during the screening phase (which is early in the baseline period).

This baseline definition for SBP, DBP, and pulse applies to all analyses (both continuous and categorical).

The baseline and postbaseline values for temperature and weight are defined below (also in Table CGAL.5.2):

- For continuous analyses of temperature and weight, last nonmissing baseline during the baseline period will be used as the baseline value.
- For the analyses of categorical changes of interest in temperature and weight:
  - The baseline is defined as the minimum value during baseline period when the direction is low.
  - The baseline is defined as the maximum value during the baseline period when the direction is high.

Table CGAL.5.4. Criteria for Treatment-Emergent, Potentially Clinical Significant and Categorical Changes and Sustained Elevation in Vital Signs and Weight

Parameter	Direction	Criteria	Patients Population Defined by Baseline Categories
Systolic BP (mm Hg) (sitting)	Low	≤90 and decrease ≥20	>90; ≤90; All patients
	High	≥140 and increase ≥20	<140; ≥ 140; All patients
	PCS High	≥180 and increase ≥20	<180; ≥ 180; All patients
	Sustained Elevation	≥140 and increase ≥20 at 2 consecutive visits	<140; ≥ 140; All patients
Diastolic BP (mm Hg) (sitting)	Low	≤50 and decrease ≥10	>50; ≤50; All patients
	High	≥90 and increase ≥10	<90; ≥ 90; All patients
	PCS High	≥105 and increase ≥15	<105; ≥ 105; All patients
	Sustained Elevation	≥90 and increase ≥10 at 2 consecutive visits	<90; ≥ 90; All patients
Systolic BP or Diastolic BP (mm Hg) (sitting)	Sustained Elevation	Meeting criteria for systolic BP for 2 consecutive visits or meeting criteria for diastolic BP for 2 consecutive visits or both	All patients
Pulse (bpm) (sitting)	Low	<50 and decrease ≥15	≥50; <50; All patients
	High	>100 and increase ≥15	≤100; >100; All patients
	Sustained Elevation	>100 and increase ≥15 at 2 consecutive visits	≤100; >100; All patients
Weight (kg)	Low	(Loss) decrease ≥7%	All patients
	High	(Gain) increase ≥7%	All patients

Criteria for Treatment-Emergent, Potentially Clinical Significant and Categorical Changes and Sustained Elevation in Vital Signs and Weight

Parameter	Direction	Criteria	Patients Population Defined by Baseline Categories
Temperature (° F)	Low	<96° F and decrease ≥2° F	≥ 96° F
	High	≥101° F and increase ≥2° F	<101° F

Abbreviations: BP = blood pressure; PCS= Potentially Clinically Significant; mm Hg = millimeters of mercury; bpm = beats per minute; kg = kilograms; ° F = degrees Fahrenheit.

5.5.9.1.4. Electrocardiogram Intervals and Heart Rate

Analyses of corrected QT (QTc) interval and QTcF (msec) will be calculated with Fridericia's formula as $QT/RR^{1/3}$. For the QTc calculations, the unit for QT is milliseconds and the unit for RR is seconds. For patients with QRS ≥120 milliseconds at any time during the study, the QT and QTc interval will be excluded from the analyses. A listing of ECG data for patients with QRS ≥120 milliseconds at any time during the study will be provided.

The baseline for ECG is defined as the last nonmissing baseline value during the baseline period. To be exact,

- The baseline for ECG is defined as the last nonmissing value before randomization. The rationale for using the last available value in the baseline period is to minimize the potential confound of discontinuing or dose stabilization of medications that modulate ECG during the screening phase (which is early in the baseline period).

This baseline definition for ECG applies to all analyses (both continuous and categorical, quantitative and qualitative).

The baseline and postbaseline values are summarized in Table CGAL.5.2.

The number and percent of patients meeting criteria for treatment-emergent abnormalities in ECG intervals (PR, QRS, and QTcF) and heart rate at any time during study will be summarized. Treatment group comparisons will be performed using Fisher's exact test for SP III.

Table CGAL.5.5 displays the criteria for treatment-emergent changes in ECG intervals and heart rate.

- For treatment emergent low analyses: Patients with normal or high values at baseline (no low values) will be included.
- For treatment emergent high analyses: Patients with normal or low values at baseline (no high values) will be included.
- For treatment emergent increase analyses: Patients with a baseline and at least 1 postbaseline result will be included.

Table CGAL.5.5. Criteria for Treatment-Emergent Changes in ECG Intervals and Heart Rate

Parameter	Direction	Criteria	
Heart Rate (bpm)	Low	<50 and decrease ≥15	
	High	>100 and increase ≥15	
PR Interval (msec)	Low	<120	
	High	≥220	
QRS Interval (msec)	Low	<60	
	High	≥120	
QTcF (msec)	Low	Males: <330	Females: <340
	High	Males: >450	Females: >470
	PCS High	>500 msec	
	Increase	Increase >30 msec	
		Increase >60 msec	
		Increase >75 msec	

Abbreviations: bpm = beats per minute; ECG = electrocardiogram; PCS = Potentially Clinically Significant; QTcF = Fridericia's corrected QT interval.

In addition, qualitative ECG abnormalities will be evaluated that will include summaries of 11 ECG categories (Axis, Rhythm, Conduction, Ischemia, Infarction, Injury, Morphology, U-waves, T-waves, ST Segment, and Other Abnormalities) of qualitative findings at any time postbaseline. A category is a collection of possible descriptions (findings) of 1 qualitative aspect of an ECG. A category name is the name of the qualitative aspect of the ECG (for example, Rhythm, Conduction, Morphology, Ischemia, and so forth). A finding is 1 of the possible specific descriptions (for example, Sinus Bradycardia, Acute Septal Infarction) within a category.

A shift table summary of qualitative ECGs at any time will be produced, to assess shifts from baseline normal to postbaseline abnormal for the overall ECG and for each of the 11 finding categories mentioned above.

The summaries of the 11 ECG categories will exclude ECGs with any of the following: overall ECG could not be evaluated by the cardiologist, lead reversals or <9 leads, nonmatching demographic data, and those suggesting patient identification errors.

5.5.9.1.5. Laboratory Tests

The incidence rates of patients with treatment-emergent abnormal, high, or low laboratory values for each laboratory test based on Covance reference ranges at any time postbaseline will be summarized. The baseline and postbaseline definitions are summarized in Table CGAL.5.2. The treatment comparisons will be assessed using Fisher's exact tests for SP III.

Patients will be defined as having a treatment-emergent low value if they have all normal or high values at baseline, followed by a value below the lower reference limit at any postbaseline visit. Patients with all normal or high values at baseline (no low values) will be included in the analysis of treatment-emergent low laboratory values. Patients will be defined as having a treatment-emergent high value if they have all normal or low values at baseline, followed by a value above the upper reference limit at any postbaseline visit. Patients with all normal or low values at baseline (no high values) will be included in the analysis of treatment-emergent high laboratory values.

For analytes simply classified as normal or abnormal, patients will be defined as having a treatment-emergent abnormal value if they have all normal values at baseline, followed by an abnormal value at any postbaseline visit. Patients with all normal values at baseline will be included in the analysis of treatment-emergent abnormal laboratory values.

The incidence of patients with the following elevations in hepatic laboratory tests at any time postbaseline will also be summarized and comparison between treatment groups for SP III using Fisher's exact test.

- The percentages of patients with an alanine aminotransferase (ALT) or aspartate aminotransferase (AST) measurement greater than or equal to 3 times (3×), 5 times (5×), and 10 times (10×) the Covance upper limit of normal (ULN) during the treatment period will be summarized for all patients with a postbaseline value.

- The percentages of patients with an alkaline phosphatase (ALP) greater than or equal to 2 times (2×) the Covance ULN during the treatment period will be summarized for all patients with a postbaseline value.
- The percentages of patients with a total bilirubin (TBIL) measurement greater than or equal to 2 times (2×) ULN during the treatment period will be summarized for all patients with a postbaseline value.

The analysis of elevation in ALT, AST, ALP, and TBIL will contain 3 subsets:

- patients whose nonmissing maximum baseline value is less than or equal to 1× ULN for ALT, AST, ALP, and TBIL.
- patients whose nonmissing maximum baseline value is greater than 1× ULN for ALT, AST, ALP, and TBIL, and at the same time less than or equal to 2× ULN for ALT and AST, 1.5× ULN for ALP and TBIL.
- patients whose nonmissing maximum baseline value is greater than 2× ULN for ALT and AST, 1.5× ULN for ALP and TBIL.

A listing of patients who had met any following criteria postbaseline will be provided over all study phases: ALT≥3× ULN, or AST ≥3× ULN, or ALP ≥2× ULN, or TBIL≥2× ULN.

5.5.9.1.6. Immunogenicity

In the immunogenicity assay process, each sample is potentially examined multiple times, according to a hierarchical procedure, to produce a sample ADA assay result and potentially a sample NAb assay result. The cut points used, the drug tolerance of an assay, and the possible values of titers are operating characteristics of the assay.

It can be the case that the presence of high concentrations of galcanezumab will affect the measurements of the presence of ADA or NAb, and conversely high levels of ADA or NAb may affect the measurement of galcanezumab concentration. Thus, a GMB drug concentration, assessed from a sample drawn at the same time as the ADA sample, plays a key role in clinical interpretation of a sample when the laboratory result is Not Detected.

5.5.9.1.6.1. Definitions of Sample Anti-drug Antibody Status

Table CGAL.5.6 and Table CGAL.5.7 list sample ADA assay results and clinical interpretation of the sample results.

Table CGAL.5.6. Sample ADA Assay Results

Sample Laboratory Result	Explanation
Detected	ADA are detected and confirmed.
Not Detected	The raw result as reported from the laboratory indicates not detected. The clinical interpretation of such results depends other factors (see below).
No Test, QNS, etc.	Sample exists but was unevaluable by the assay

Abbreviation: ADA = anti-drug antibody.

Table CGAL.5.7. Sample Clinical ADA Interpretation Results

Sample Laboratory Result	Explanation
ADA Present	ADA assay result detected.
ADA Not Present	ADA assay result is Not Detected and simultaneous drug concentration is at a level that has been demonstrated to not interfere in the ADA detection method (i.e., drug concentration is below the assay's drug tolerance level). For patients receiving placebo, drug concentration is not assessed and is assumed to be below the assay's drug tolerance level.
ADA Inconclusive	ADA assay result is Not Detected but drug concentration in the sample is at a level that can cause interference in the ADA detection method.
ADA Not Detected with Drug Concentration Not Available	If drug concentration analysis was planned but result is not available for a treatment-period sample, a Not Detected sample will be declared ADA Not Detected with Drug Concentration Not Available. In the computation of Patient ADA status (see below, Section5.5.9.1.6.2), these samples will be considered ADA Not Present, on the basis of prior knowledge that the drug tolerance level of the ADA assay is high relative to the expected drug concentration levels.
ADA Missing	ADA sample not drawn, QNS, not tested, etc., causing there to be no laboratory result reported or the result is reported as “no test”.

Abbreviation: ADA = anti-drug antibody.

Parallel terminology applies for NAb Detected, NAb Not Detected, NAb Present, NAb Not Present, NAb Inconclusive, NAb Not Detected with Drug Concentration Not Available, and NAb Missing. Anti-drug antibody and NAb are distinct assays and have different assay-operating characteristics.

5.5.9.1.6.2. Definitions of Patient Anti-drug Antibody Status

Patient evaluable for TE ADA: A patient is evaluable for TE ADA if the patient has a nonmissing baseline ADA result and at least 1 nonmissing postbaseline.

TE ADA positive (TE ADA+) patient: A patient who is evaluable for TE ADA is TE ADA+ if either of the following holds:

- **Treatment-induced:** The patient has baseline status of ADA Not Present and at least one postbaseline status of ADA Present with titer ≥ 20 (ie, 2*MRD where for this ADA assay the MRD, the minimum required dilution of the ADA assay, is 10).
- **Treatment-boosted:** The patient has baseline and postbaseline status of ADA Present, with the postbaseline titer being 2 dilutions (4-fold) greater than the baseline titer. That is, the patient has baseline status of ADA Present, with titer 1:B, and at least 1 postbaseline status of ADA Present, with titer 1:P, with P/B ≥4.

TE ADA Inconclusive patient: A patient who is evaluable for TE ADA is TE ADA Inconclusive if ≥20% of the patient's postbaseline samples, drawn predose, are ADA Inconclusive and the patient is not otherwise TE ADA+.

TE ADA negative (TE ADA-) patient: A patient who is evaluable for TE ADA is TE ADA- when the patient is not TE ADA+ and the patient is not TE ADA Inconclusive.

5.5.9.1.6.3. Analyses to Be Performed

To evaluate the changes in immunogenicity data (Anti-galcanezumab [ADA and NAb]) after treatment, the number and proportion of patients who are TE ADA+ will be tabulated where proportions are relative to the number of patients who are TE ADA evaluable as defined in Section 5.5.9.1.6.2). The baseline and postbaseline definitions for each analysis period is shown in Table CGAL.5.2. In detail the following statistical analyses for each immunogenicity analyte (ADA and NAb) are planned:

- the incidence of TE ADA will be summarized as follows:
  - for safety population during double-blind treatment phase and compared between treatment arms using Fisher's exact test
  - for the ADA follow-up cohort during double-blind treatment phase, the ADA follow-up cohort is defined as patients in the safety population with ADA assessment during post-treatment phase
  - for the ADA follow-up cohort during double-blind treatment phase and post-treatment phase
- shift from baseline to maximum postbaseline ADA titers for the galcanezumab-treated patients during double-blind treatment phase
- summary of time to first TE ADA+ titer during double-blind phase

The following descriptive listings will also be provided:

- listing of patients with TE ADA at any time during study, NAb Status will also be displayed
- listing of patients with inconclusive ADA or inconclusive NAb at any time

- listing of patients with ADA present at any time or TE hypersensitivity events or TEAEs related to injection sites.

5.5.9.2. Continuous Safety Measures

Analyses of continuous safety data will be conducted on patients who have a baseline and at least 1 postbaseline observation for SP III.

For all the continuous safety measures (including planned laboratory measures, vital signs and weight, ECG intervals and heart rate), box-whisker plots with summary statistic tables for absolute value and change from baseline at scheduled visit and at endpoint (defined as the final postbaseline value) will be provided for SP III. The change from baseline results will be compared between treatment arms using the analysis of covariance (ANCOVA) model with treatment, pooled investigative site and baseline value in the model.

5.5.10. Subgroup Analyses

Subgroup analyses will be performed for primary efficacy measure (change from baseline on weekly number of cluster headache attack) only for the ITT patients in the SP III.
Table CGAL.5.8 provides definitions for each subgroup variable. Subgroup variables are usually selected if they are potentially prognostic or predictive. A subgroup variable is prognostic if values of the subgroup variable predict the change in efficacy measures regardless of the treatment group assignment. A subgroup variable is predictive if values of the subgroup variable predict heterogeneous treatment effect. Demographic subgroup variables (sex, racial origin, ethnicity, age, and region) may neither be prognostic nor predictive, but they are standard subgroup variables needed for regulatory submission. Baseline average daily number of cluster headache attack category and sex were included in the dynamic allocation randomization algorithm and are considered possibly prognostic.CCI [redacted]
The purpose of the analyses for these subgroup variables is to assess the consistency of treatment effects across the different values of each subgroup variable.

Table CGAL.5.9 summarizes the subgroup analyses to be conducted, using those subgroup variables presented in Table CGAL.5.8.

Table CGAL.5.8. Definition of Subgroup Variables

Subgroup Variable	Categories
Sex	Male, female
Racial origin (combine those with less than 10%)	American Indian/Alaskan Native Asian Black/African American Native Hawaiian/Pacific Islander White Multiple
Ethnicity	Hispanic or Latino Not Hispanic or Latino
Age	<40 or ≥40

Definition of Subgroup Variables

Subgroup Variable	Categories
Baseline average daily number of cluster headache attack category	1) ≤4 attack per day, >4 attack per day 2) ≤3 attacks per day, >3 attacks per day 3) ≤2 attacks per day, >2 attacks per day
Region	Europe, North America (US and Canada)
CCI [redacted]	

Table CGAL.5.9. Subgroup Analyses

Outcome Variable	Subgroup Variables	Analysis
EFFICACY VARIABLES		
1. Change from baseline to each postbaseline weekly interval up to Week 3 in the SP III for: Number of cluster headache attacks	Sex Racial origin Ethnicity Age Baseline average daily number of cluster headache attack category Region CCI [redacted]	Repeated measures analysis using the model described in Section 5.5.8.1 with additional terms for subgroup, subgroup-by-treatment, subgroup-by-week, and subgroup-by-treatment-by-week interactions added to the base model.

Abbreviations: CCI [redacted] SP = study phase.

For the subgroup variable of race, all the categories that have less than 10% of the patients in the study will be combined in the analysis. CCI [redacted] For subgroup analyses, the subgroup-by-treatment and subgroup-by-treatment-by-visit/week interactions will be tested at a 2-sided 0.05 significance level. Treatment group differences will be evaluated within each category of the subgroup variable.

The subgroup analysis for change from baseline to each weekly interval up to Week 3 in number of cluster headache attacks will be conducted with repeated measures analysis. The same MMRM model as described in Section 5.5.8.1 will be used with additional terms of subgroup, subgroup-by-treatment, subgroup-by-week, and subgroup-by-treatment-by-week interactions added. In this analysis, the p-value for the subgroup-by-treatment, subgroup-by-week, and subgroup-by-treatment-by-week interactions will be reported.

For subgroup analysis, the LSMean and LSMean change estimate as well as the treatment comparisons within each subgroup will be analyzed with the data within that specific subgroup only.

5.5.11. Sensitivity Analysis

For all sensitivity analyses for the primary efficacy endpoint, the diary data up to Week 3 will be used.

Dynamic Allocation (Minimization) Assumption

A permutation test will be performed as a sensitivity analysis of the primary MMRM analysis to confirm the results of the asymptotic inference. The key features of the permutation test that will be employed are as follows:

- The patients' baseline covariates, responses, and enrollment order will be considered fixed.
- The sharp null hypothesis will be assumed, i.e., responses to galcanezumab and placebo will be assumed exactly equal.
- The exact minimization algorithm and exact site pooling algorithm will be used to generate the null distribution of the primary test statistic from the MMRM analysis.
- The p-value based on the generated null distribution (i.e., permutation test p-value) will be obtained by comparing the observed test statistic value to the percentiles of the generated null distribution.

Explicitly, the p-value is derived from the permuted distribution of test statistics as follows. If the total number of permutations is m, and b of these permutations have a test statistic greater than or equal to the observed test statistic, z, then the permutation p-value, p^p is,

$$p^p = \frac{b+1}{m+1}$$

where m equals 100 000. As discussed in Phipson and Smith (2010), this is an upper bound on the estimated p-value. This method is used to generate the approximate null distribution. Note that the described permutation p-value calculation should be conducted such that a positive value for the test statistic should indicate a favorable treatment effect galcanezumab 300 mg relative to placebo.

Missing Data Assumption

Sensitivity analyses will be performed to assess the robustness of the primary analysis conclusions to deviations from MAR assumption. The approach for these analyses is to vary the assumptions of missing data for the primary analysis in a systematic way. Basically, the method will be to predict the missing outcomes and then add a value (denoted as Δ_A) to the predictions in the active treatment group and another value (denoted as Δ_P) to the predictions in the placebo treatment group, consistent with the sensitivity approach suggested in Permutt (2015). This procedure will be repeated multiple times for different values of (Δ_A, Δ_P) using the following steps:

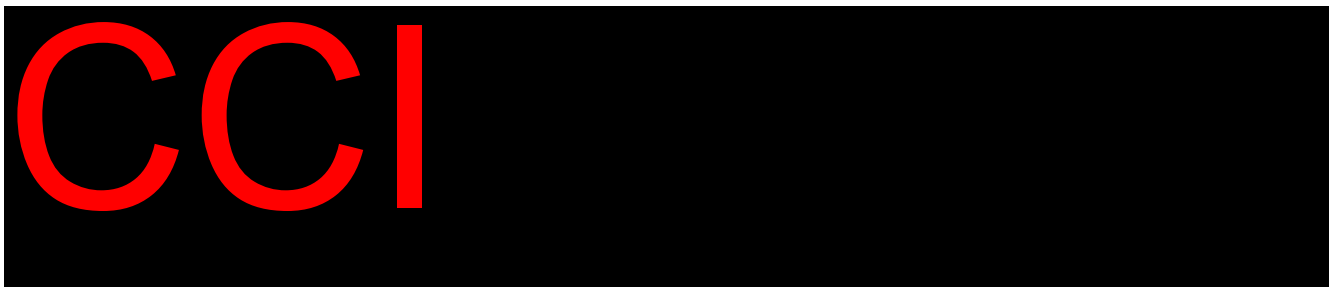
1. Predict the missing outcomes for each treatment via multiple imputation based on observed primary endpoint and baseline values. Such imputation will be carried out using a Markov Chain Monte Carlo method with a Jeffreys prior via SAS® PROC MI. Thirty (30) such imputations will be created.
2. Add Δ_A to the imputed values for patients taking active treatment and Δ_P to the imputed values for patients taking placebo.
3. Conduct the primary analysis separately for each of the 30 imputations.
4. Combine the results of these analyses using Rubin's combining rules, as implemented in SAS® PROC MI ANALYZE.

The above steps will be repeated multiple times for different values of (Δ_A, Δ_P) with Δ_P ranging from (0, twice the absolute value of the mean value seen for placebo in the primary analysis) and Δ_A ranging from (Δ_P, Δ_P + absolute value of the mean treatment difference seen within the primary analysis). For example, if the mean change from baseline for placebo is -3.6 and the corresponding treatment difference is -1.5, then Δ_P would range from (0, 7.2) and Δ_A would range from (Δ_P, Δ_P + 1.5).

Normality Assumption

To assess the robustness of the MMRM results to deviations from normality assumption, a sensitivity analysis for raw number of cluster headache attacks (total number of cluster headache attacks for each interval without imputing missing value and without normalization to 7-day period) will be conducted with a repeated measures negative binomial regression model fitted with SAS PROC GLIMMIX. The model will include treatment, gender, pooled investigative site, weekly time period (Week 1, Week 2, Week 3), and treatment-by-time-period interaction, as well as the continuous fixed covariates of baseline value, and log (number of compliant days within each weekly time period divided by 7) as the offset in the model. In case of nonconvergence, pooled investigative site may be excluded from the model. Directional consistency of treatment effects from this model and the primary analysis MMRM model as specified in Section 5.5.8.1 will be examined.

In addition, as another form of sensitivity analysis, residuals from the primary analysis MMRM model will be examined and outliers identified. Consistency of results before and after removing patients with outlier residuals will be examined.



5.6. Interim Analyses

Up to 2 interim analyses were planned for Study CGAL. Interim analysis 1 during SP III may be conducted that may result in increasing the sample size or stopping the trial for futility. Details were documented in the Statistical Analysis Center SAP, ERB supplement, and DMC Charter. However, this interim analysis for sample size re-estimation will not happen due to enrollment infeasibility. The DMC will still independently monitor patient safety during this trial.

The other interim analysis that was planned will be conducted after all patients have had the opportunity to complete 8 weeks of treatment (SP III) and, thus, will be the final analysis of the primary efficacy endpoint. The interim analysis will be conducted using internal unblinded study team members who do not have direct interaction with sites.

5.7. Unblinding Plan

Interim analysis will be conducted by unblinded study team members who do not have direct interaction with sites. All study personnel with direct interaction with sites are kept blinded to individual patient treatment information.

5.8. Reports to Be Generated at Each Interim and Final Database Lock

5.8.1.1. Reports to Be Generated at Interim Database Lock

For the interim analysis, the database will be locked after all randomized patients have had the chance to complete 8 weeks of treatment in SP III. However, some patients will still be ongoing in SP IV at the time of the database lock. Data up to the data cutoff date in the locked database from all study phases will be used, and analyses specified in this SAP will be performed. However, only analyses conducted for SP III at interim analysis will be considered as the final

analyses. The analyses including data from SP IV will be rerun and updated when the completed data are available at the final database lock.

5.8.1.2. Reports to Be Generated at Final Database Lock

For final database lock, all analyses including tables, figures and listings that use data from SP IV will be generated.

5.9. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements. These analyses will be the responsibility of the Sponsor.

Analyses provided for the CTR requirements include the following:

A summary of AEs will be provided as a dataset that will be converted to an XML file. Both Serious Adverse Events and "Other" Adverse Events are summarized by treatment group and by MedDRA PT.

- An AE is considered "Serious" whether or not it is a TEAE.
- An AE is considered in the "Other" category if it is both a TEAE and is not serious. For each "Serious" AE and "Other" AE, for each term and treatment group, the following are provided:
  - the number of participants at risk of an event
  - the number of participants who experienced each event term
  - the number of events experienced
- Consistent with www.ClinicalTrials.gov requirements, "Other" AEs that occur in fewer than 5% of patients in every treatment group may not be included if a 5% threshold is chosen.
- AE reporting is consistent with other document disclosures; for example, the CSR, manuscripts, and so forth.

6. References

Diggle P, Liang K, Zeger S. Analysis of longitudinal data. Oxford: Clarendon Press; 1994.

Guy W. ECDEU assessment manual for psychopharmacology, revised 1976. Rockville, MD: National Institute of Mental Health, Psychopharmacology Research Branch. p 217-222. Available at: https://archive.org/details/ecdeuassessmentm1933guyw. Accessed September 29, 2014.

Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. 1997;53(3):983-997.

Koch GG, Tangen CM, Jung JW, Amara IA. Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. *Stat Med*. 1998;17(15-16):1863-1892.

Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso J. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Information Journal.* 2008;42(4):303-319.

Permutt T. Sensitivity analysis for missing data in regulatory submissions. *Stat Med*. 2015;35(17):2876-2879.

Phipson B, Smyth GK. Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Stat Appl Genet Mol Biol.* 2010; 9:Article39.

Pocock S, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*. 1975;31(1):103-115.

The Sumatriptan Cluster Headache Study Group. Treatment of acute cluster headache with sumatriptan. N Engl J Med. 1991;325(5):322-326.

Zink RC, Koch GG. NParCOV3: a SAS/IML macro for nonparametric randomization based analysis of covariance. *J Stat Softw*. 2012;50(3):1-17.

7. Appendices

Appendix 1. Description of Important Protocol Deviations

Category	Subcategory	Study Specific Term	Source
Informed Consent Form (ICF)	Informed consent not obtained		Programmable
	Improper consent	ICF not signed prior to initiation of protocol procedures	Nonprogrammable
Eligibility	Inclusion/ Exclusion	At Visit 1 patients must have a history of episodic cluster headache	Nonprogrammable
		Age < 18 or > 65 years old at study entry	Nonprogrammable
		Female patients who have a positive serum pregnancy test prior to Visit 3	Programmable
		Randomized patients had prior or current exposure to CGRP antibody	Nonprogrammable
		Corrected QT (QTcB) interval > 470 msec for women and >450 for men prior to Visit 3	Nonprogrammable
		PR > 220, or conduction delay of QRS>120 prior to Visit 3	Nonprogrammable
		SBP >160 mm Hg or DBP >100 mm Hg on 2 or more blood pressure assessments prior to Visit 3	Programmable
		Evidence of ischemia /qualitative findings of ST or J-point elevation, excluding early repolarization	Nonprogrammable
		History of MI, UA, PCI, CABG, or DVT/PE within 6 months of screening	Nonprogrammable
		Have planned cardiovascular surgery or percutaneous coronary angioplasty	Nonprogrammable

Category	Subcategory	Study Specific Term	Source
Eligibility	Inclusion/ Exclusion	Any lifetime history of vasospastic angina or stroke	Nonprogrammable
		Clinical evidence of peripheral vascular disease or a diagnosis of Raynaud's Phenomenon	Nonprogrammable
		Have any history of intracranial or carotid aneurysm, intracranial hemorrhage, stroke	Nonprogrammable
		Have a history of intracranial tumors or significant head trauma that preclude study participation	Nonprogrammable
		Have a clinically significant elevation of ≥2 X ULN for ALT, or ≥1.5 X ULN for TBIL or ALP prior to V3	Nonprogrammable
		Have a positive urine drug screen for substances of abuse not allowed prior to randomization	Programmable
		Completion of less than 5 of 7 days of the daily ePRO diary during the baseline assessment	Programmable
		Baseline weekly cluster headache attack: (a) ≥2 consecutive days without attack, or (b) <4 total attacks, or (c) >8 attacks per day	Programmable
		Body mass index (BMI) ≥40 kg/m2 at baseline.	Programmable
		Use within 14 days prior to SP II or in SP II/III of any of the medications described in I/E 9a	Programmable
		Use within 30 days prior to SP II or in SP II/III of any of the medications described in I/E 9b	Programmable

Category	Subcategory	Study Specific Term	Source
Study Procedures	Other	Use of Botox within 4 month prior to SP II and during study	nonprogrammable
		Use of other excluded meds during study	Nonprogrammable
		Use of verapamil at doses higher than allowed at baseline and during study	Programmable
		Missing any scheduled or unscheduled C-SSRS	Programmable
		Missing all triplet measurements of blood pressure or pulse at any scheduled visit	Programmable
		Missing entire chemistry or hematology panel	Programmable
		No ECG measurements during a study phase	Programmable
Investi-gational Product	Patient took medication not fit for use		Nonprogrammable
	Unblinding		NonProgrammable
	Other	IP lost or stolen	Nonprogrammable
		Dose planned but not given–date of injection missing	Programmable
		Dosing interval outside specified limits of 21-37 days for double-blind treatment phase	Programmable
	Dosing Error		Nonprogrammable
Safety	SAEs		Nonprogrammable
	Other	Positive pregnancy test	Nonprogrammable
Data Quality	Treatment Assignment/ Randomization Error	IWRS data entry errors that impact patient stratification	Programmable
	Treatment Assignment/Randomization Error	Randomized after screening failure, no study drug dispensed	Programmable

Category	Subcategory	Study Specific Term	Source
Data Quality	Other	Primary efficacy compliance rate ≤50% in any weekly interval during DB treatment phase	Programmable
	Data Entry Issues	Patients did not report oxygen use in number of times in eDiary	nonprogrammable
Admini-strative Oversight	Patient Privacy Violation		Nonprogrammable
	Suspected Misconduct		Nonprogrammable
	Other	Post training; switching roles blinding to unblinded vice versa without prior medical team approval	Nonprogrammable
		Unqualified or untrained site personnel administer (C-SSRS)	Nonprogrammable
		Quality issue at site or vendor	Nonprogrammable

Abbreviations: ALT = alanine aminotransferase; ALP = alkaline phosphatase; BMI = body mass index; CABG = coronary artery bypass grafting; C-SSRS = Columbia-Suicide Severity Rating Scale; DBP = diastolic blood pressure; DVT = deep vein thrombosis; ECG = electrocardiogram; ePRO = electronic patient-reported outcome; ICF = informed consent form; I/E = inclusion/exclusion criteria; IP = investigational product; IWRS = interactive web-response system; MI = myocardial infarction; PCI = percutaneous coronary intervention; PE = pulmonary embolism; PR = pulse rate; SAE = serious adverse event; SBP = systolic blood pressure; SP = study phase; TBIL = total bilirubin; UA = unstable angina; ULN = upper limit normal.

Leo Document ID = 6d5b9e94-d91f-45f5-8bbd-335b57713af7

Approver: PPD
Approval Date & Time: 05-Apr-2018 19:35:35 GMT
Signature meaning: Approved