Statistical Analysis Plan with Amendment 01

A Multicenter, Randomized, Double-Blind, Double-Dummy, Placebo-Controlled, Parallel-Group Study Comparing the Efficacy and Safety of 2 Dose Regimens (Intravenous/Subcutaneous and Subcutaneous) of TEV-48125 versus Placebo for the Prevention of Chronic Cluster Headache

Study Number TV48125-CNS-30057

NCT02964338

Statistical Analysis Plan with Amendment 01 Approval Date: 19 September 2018

Statistical Analysis Plan

Study Number TV48125-CNS-30057

A Multicenter, Randomized, Double-Blind, Double-Dummy, Placebo-Controlled, Parallel-Group Study Comparing the Efficacy and Safety of 2 Dose Regimens (Intravenous/Subcutaneous and Subcutaneous) of TEV-48125 versus Placebo for the Prevention of Chronic Cluster Headache

Phase 3

IND number: 129606; EudraCT number: 2016-003171-21

Statistical Analysis Plan Approval Date: 08 January 2018

Statistical Analysis Plan Amendment 01 Approval Date : 19 September 2018

Sponsor

Teva Branded Pharmaceutical
Products R&D, Inc.
41 Moores Road
Frazer, Pennsylvania 19355
United States

Prepared by: [redacted]

STATISTICAL ANALYSIS PLAN APPROVAL

Study No.: **TV48125-CNS-30057**

Study Title: **A Multicenter, Randomized, Double-Blind, Double-Dummy, Placebo-Controlled, Parallel-Group Study Comparing the Efficacy and Safety of 2 Dose Regimens (Intravenous/Subcutaneous and Subcutaneous) of TEV-48125 versus Placebo for the Prevention of Chronic Cluster Headache**

Statistical Analysis Plan for:

☐ **Interim Analysis** ☐ **Integrated Summary of Efficacy**

☒ **Final Analysis** ☐ **Integrated Summary of Safety**

Amendment: **1**

Author: [redacted]

[redacted] Sep 10, 2018

Approver: **Date**

[redacted] 19 SEP 2018.

Approver: **Date**

[redacted]

TABLE OF CONTENTS

TITLE PAGE 1
STATISTICAL ANALYSIS PLAN APPROVAL 2
AMENDMENT HISTORY 9
LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS 10
INTRODUCTION 12
1. STUDY ENDPOINTS 13
1.1. Primary Efficacy Endpoints 13
1.2. Secondary Efficacy Endpoints 13
1.3. Safety Endpoints 13
1.4. Immunogenicity Assessment Endpoints 14
1.5. Exploratory Endpoints 14
1.6. Wearable Sensor Substudy Endpoints 15
2. STUDY DESIGN 17
2.1. General Design 17
2.1.1. Wearable Sensor Substudy 18
2.2. Randomization and Blinding 19
2.3. Data Monitoring Committee 19
2.4. Sample Size and Power Considerations 19
2.5. Sequence of Planned Analyses 20
2.5.1. Planned Interim Analyses 20
2.5.2. Final Analyses and Reporting 20
3. ANALYSIS SETS 21
3.1. Intent-to-Treat Analysis Set 21
3.2. Safety Analysis Set 21
3.3. Full Analysis Set 21
3.4. Per-Protocol Analysis Set 21
4. GENERAL ISSUES FOR DATA ANALYSIS 22
4.1. General 22
4.2. Specification of Baseline Values 22
4.3. Region of Pooled Countries 22
4.4. Handling Withdrawals and Missing Data 22

4.5. Study Days and Visits....22
5. STUDY POPULATION....24
5.1. General....24
5.2. Patient Disposition....24
5.3. Demographics and Baseline Characteristics....24
5.4. Medical History....24
5.5. Prior Therapy and Medication....24
5.6. Childbearing Potential and Methods of Contraception....25
5.7. Physical Examinations....25
5.8. Study Protocol Deviations....25
6. EFFICACY ANALYSIS....26
6.1. General....26
6.1.1. Cluster Headache Attacks Data from E-diary....26
6.1.2. Data Derivation....26
6.1.3. Analysis Windows for E-diary Data....27
6.1.4. Handling Missing E-diary Data....28
6.1.4.1. Primary Efficacy Endpoint....29
6.1.4.2. Analyses Based on Monthly and 12-week Window....29
6.1.4.3. Analyses of Mean Change from Baseline Using MMRM Method Based on 1-Week Window....29
6.2. Primary Efficacy Endpoint and Analysis....30
6.2.1. Definition....30
6.2.2. Primary Efficacy Analysis....30
6.2.3. Sensitivity Analyses for the Primary Efficacy Endpoint....30
6.2.3.1. Analysis with Multiple Imputation Method....31
6.2.3.2. Mixed Model for Repeated Measures Analysis....31
6.2.4. Subgroup Analyses....32
6.3. Secondary Efficacy Endpoints and Analyses....33
6.3.1. Proportion of Patients with a ≥50% Reduction from Baseline in the Monthly Average Number of CH Attacks During the 12-Week Period After the First Dose of the IMP....33
6.3.2. Mean Change from Baseline in the Number of Cluster Headache Attacks....33
6.3.3. Mean Change from Baseline in the Weekly Average Number of Days with Use of Cluster-Specific Acute Headache Medications (Triptans and Ergot

Compounds) During the 12-Week Period After Administration of the First Dose of the IMP 34
6.3.4. Mean Change from Baseline in the Weekly Average Number of Days Oxygen is Used to Treat Chronic Cluster Headaches During 12-week Period After Administration of the First Dose of the IMP 34
6.3.5. Patient-Perceived Satisfactory Improvement at 1, 4, 8, and 12 Weeks After Administration of the First Dose of the IMP Relative to Baseline 34
6.4. Exploratory Efficacy Endpoints and Analysis 35
6.4.1. Proportion of Patients with ≥50%, 75%, or 100% Reduction from Baseline in Number of CH Attacks During the 12-Week Period (First 3 Exploratory Efficacy Endpoints) 35
6.4.2. Mean Change from Baseline in the Number of CH Attacks During Weeks 1, 2, 3, 4, 8, and 12 After Administration of the First Dose of the IMP 36
6.4.3. Mean Change from Baseline in the Number of CH Attacks During the 8-Week Period (Weeks 0 to 8) After Administration of the First Dose of the IMP 36
6.4.4. Mean Change from Baseline in the Number of CH Attacks During the 4-Week Period (Weeks 4 to 8) After Administration of the Second Dose of the IMP 36
6.4.5. Change from Baseline in the Severity of CH Attacks (Mild, Moderate, Severe and Very Severe) During Weeks 1, 2, 3, 4, 8, and 12 After Administration of the First Dose of the IMP 36
6.4.6. Mean Change from Baseline in the Weekly Average Number of Days with Use of any Acute Headache Medications During the 12-Week Period After Administration of the First Dose of the IMP 37
6.4.7. Cumulative Proportion of Responders Analysis 37
6.4.8. Mean Change from Baseline in the Monthly Average Number of CH Attacks of at Least Severe Severity During the 12-Week Period After Administration of the First Dose of the IMP 39
6.4.9. Changes from Baseline in Responses to Questionnaires That Measure Quality of Life, Satisfaction with Treatment, and Health Status 39
6.4.9.1. Change from Baseline in EuroQol-5 Dimension Questionnaire 39
6.4.9.2. Hospital Anxiety and Depression Scale 39
6.4.9.3. 12-Item Short-Form Health Survey 40
6.4.9.4. Impact on Partner and Family Questionnaire 41
6.4.9.5. Work Productivity and Activity Impairment Questionnaire: General Health V2.0 41
6.4.9.6. Patient Global Impression of Change Scale 42

7. MULTIPLE COMPARISONS AND MULTIPLICITY 43
8. SAFETY ANALYSIS 44
8.1. General....44
8.2. Duration of Exposure to Study Drug....44
8.3. Adverse Events....44
8.4. Injection Site Assessments....45
8.5. Hypersensitivity/Anaphylaxis....45
8.6. Electronic Columbia Suicide Severity Rating Scale....45
8.7. Deaths....46
8.8. Clinical Laboratory Tests....46
8.8.1. Laboratory Values Meeting Hy's Law Criteria....48
8.9. Physical Examinations....48
8.10. Vital Signs....48
8.11. Electrocardiography....49
8.12. Concomitant Medications or Therapies....49
9. TOLERABILITY VARIABLES AND ANALYSIS....50
10. PHARMACOKINETIC ANALYSIS....51
11. PHARMACOGENOMIC ANALYSIS....52
12. BIOMARKER ANALYSIS....53
13. IMMUNOGENICITY ANALYSIS....54
14. ANCILLARY STUDY ANALYSIS – WEARABLE SENSOR SUBSTUDY....55
15. PLANNED INTERIM ANALYSIS....56
15.1. Statistical Methods....56
15.2. Timing and Data....56
15.3. Result Dissemination....57
15.4. Decision Rules....57
16. STATISTICAL SOFTWARE....58
17. CHANGES TO ANALYSES SPECIFIED IN THE STUDY PROTOCOL....59
18. REFERENCES....60
APPENDIX A. E-DIARY QUESTIONNAIRE....61
APPENDIX B. LOGICS FOR CLUSTER HEADACHE DERIVATION....64
APPENDIX C. EUROQOL-5 DIMENSION QUESTIONNAIRE (EQ-5D)....65
APPENDIX D. HOSPITAL ANXIETY AND DEPRESSION SCALE (HADS)....67

APPENDIX E. 12-ITEM SHORT-FORM HEALTH SURVEY (SF-12) 69

APPENDIX F. IMPACT ON PARTNER AND FAMILY QUESTIONNAIRE - PATIENTS 71

APPENDIX G. IMPACT ON PARTNER AND FAMILY QUESTIONNAIRE - PARTNERS 73

APPENDIX H. WORK PRODUCTIVITY AND ACTIVITY IMPAIRMENT QUESTIONNAIRE: GENERAL HEALTH V2.0(WPAI:GH)....................................75

LIST OF TABLES

Table 1: Study Visits....................23
Table 2: Criteria for Potentially Clinically Significant Laboratory Values....................47
Table 3: Criteria for Potentially Clinically Significant Vital Signs....................48

AMENDMENT HISTORY

The Statistical Analysis Plan for study TV48125-CNS-30057 (study protocol with amendment 02 dated 01 May 2017) has been amended and reissued as follows:

Date	**Summary of changes**	**Reason for amendment**
01 May 2018	Section 6.4.5 New: $\frac{\sum most\ severe\ CH\ attack\ per\ day\ during\ a\ week}{Number\ of\ days\ with\ assessments\ recorded\ in\ the\ eDiary\ during\ a\ week}$ Old: $\frac{\sum most\ severe\ CH\ attack\ per\ day\ during\ a\ week}{Number\ of\ days\ with\ assessments\ recorded\ in\ the\ eDiary\ during\ a\ week} \times 7$	Correction
30 March 2018	Section 15.1, last paragraph Old: "….in the **weekly** average number of CH attacks during the **4-week** period after administration of the IMP." New: "….in the **monthly** average number of CH attacks during the **12-week** period after administration of the IMP."	Correction to typo
13 August 2018	Section 6.4.9.3, 3rd paragraph and last paragraph: Deleted wording for "**Computed raw scores for 8 domains" and "Computed raw scores".**	The software does not output computed raw scores

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	**Term**
ADA	antidrug antibodies
ANCOVA	analysis of covariance
ANOVA	analysis of variance
β-HCG	beta-human chorionic gonadotropin
CGRP	calcitonin gene-related peptide
CCH	chronic cluster headache
CH	cluster headache
CMH	Cochran-Mantel-Haenszel
CNS	central nervous system
CP	conditional power
CPRA	cumulative proportion of responder's analysis
CRF	case report form
CV	coefficient of variation
e-diary	electronic diary
ECG	electrocardiogram/electrocardiography
eC-SSRS	electronic Columbia Suicide Severity Rating Scale
EOT	end of treatment
EQ-5D	EuroQol-5 Dimension
FAS	full analysis set
HADS	Hospital Anxiety and Depression Scale
IMP	investigational medicinal product
IRT	interactive response technology
ITT	intent-to-treat
iv	intravenous(ly)
LS	least square
MCS	Mental Health Composite Scores
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MMRM	mixed model for repeated measures
NSAID	non-steroidal anti-inflammatory drug
PCS	Physical Composite Scores
PGIC	Global Impression of Change
PP	per-protocol

Abbreviation	**Term**
PPSI	Patient-Perceived Satisfactory Improvement
R&D	Research and Development
SAP	statistical analysis plan
sc	subcutaneous(ly)
SD	standard deviation
SE	standard error
SF-12	12-Item Short-Form Health Survey
SI	standard international
SOC	system organ class
SOP	standard operating procedure
ULN	upper limit of normal
UN	unstructured covariance
WHO Drug	World Health Organization Drug Dictionary
WPAI	Work Productivity and Activity Impairment

INTRODUCTION

This statistical analysis plan (SAP) describes the planned analysis and reporting for Teva Branded Pharmaceutical Products R&D, Inc. Study TV48125-CNS-30057, (a multicenter, randomized, double-blind, double-dummy, placebo-controlled, parallel-group study comparing the efficacy and safety of 2 dose regimens [intravenous/subcutaneous and subcutaneous] of TEV-48125 [fremanezumab] versus placebo for the prevention of chronic cluster headache), and was written in accordance with SOP GBP_RD_702.

The reader of this SAP is encouraged to read the study protocol for details on the conduct of this study, the operational aspects of clinical assessments, and the timing for completing the participation of a patient in this study.

The SAP is intended to be in agreement with the protocol, especially with regard to the primary and all secondary endpoints and their respective analyses. However, the SAP may contain more details regarding these particular points of interest, or other types of analyses (eg, other endpoints). When differences exist in descriptions or explanations provided in the study protocol and this SAP, the SAP prevails; the differences will be explained in the clinical study report.

1. STUDY ENDPOINTS

1.1. Primary Efficacy Endpoints

The primary efficacy endpoint of this study is the mean change from baseline (run-in period) in the monthly average number of cluster headache (CH) attacks during the 12-week period after administration of the first dose of the investigational medicinal product (IMP), ie, based on week 0 to 12 data.

1.2. Secondary Efficacy Endpoints

The secondary efficacy endpoints to further demonstrate efficacy are:

- the proportion of patients with a ≥50% reduction from baseline (run-in period) in the monthly average number of CH attacks over the 12-week period after the administration of the first dose of the IMP, ie, based on week 0 to 12 data
- the mean change from baseline (run-in period) in the number of CH attacks during the 4-week period after administration of the first dose of the IMP, ie, based on week 0 to 4 data
- the mean change from baseline (run-in period) in the number of CH attacks during the 4-week period after administration of the third dose of the IMP, ie, based on week 8 to 12 data
- the mean change from baseline (run-in period) in the weekly average number of days with use of cluster-specific acute headache medications (triptans and ergot compounds) during the 12-week period after administration of the first dose of the IMP, ie, based on week 0 to 12 data
- the mean change from baseline (run-in period) in the weekly average number of days oxygen is used to treat chronic cluster headache (CCH) during the 12-week period after administration of the first dose of the IMP, ie, based on week 0 to 12 data
- assessment of patient's perceived improvement, as measured by the Patient-Perceived Satisfactory Improvement (PPSI) at 1, 4, 8, and 12 weeks after administration of the first dose of the IMP relative to baseline (day 0)

1.3. Safety Endpoints

The safety endpoints are as follows:

- occurrence of adverse events throughout the study
- clinical laboratory (serum chemistry, hematology, coagulation, and urinalysis) test results at each visit
- vital signs (systolic and diastolic blood pressure, oral temperature, and pulse rate) measurements at each visit. Note: Oxygen saturation will be measured in cases of suspected anaphylaxis and severe hypersensitivity. Respiratory rate will also be measured in these cases but not as a standard vital sign.
- 12-lead electrocardiogram (ECG) findings at screening, baseline, and week 12

- use of concomitant medication during the study
- clinically significant changes in physical examinations, including body weight
- injection site reaction (ie, erythema, induration, and ecchymosis) and injection site pain assessments
- occurrence of hypersensitivity/anaphylaxis reactions
- suicidal ideation and behavior as measured by the electronic Columbia Suicide Severity Rating Scale (eC-SSRS)

1.4. Immunogenicity Assessment Endpoints

The immunogenicity endpoints are the following:

- antidrug antibody (ADA) incidence and characteristics (eg, titer, kinetics, and neutralizing activities)

1.5. Exploratory Endpoints

The exploratory endpoints are as follows:

- proportion of patients with ≥50% reduction from baseline (run-in period) in the weekly average number of CH attacks during the 12-week period after administration of the first dose of the IMP, ie, based on week 0 to 12 data
- proportion of patients with ≥75% reduction from baseline (run-in period) in the weekly average number of CH attacks during the 12-week period after administration of the first dose of the IMP, ie, based on week 0 to 12 data
- proportion of patients reaching 100% reduction in the number of CH attacks during the 12-week period after administration of the first dose of the IMP, ie, based on week 0 to 12 data
- mean change from baseline (run-in period) in the number of CH attacks during weeks 1, 2, 3, 4, 8, and 12, after administration of the first dose of the IMP
- mean change from baseline (run-in period) in the number of CH attacks during the 8-week period after administration of the first dose of the IMP, ie, based on week 0 to 8 data
- mean change from baseline (run-in period) in the number of CH attacks during the 4-week period after administration of the second dose of the IMP, ie, based on week 4 to 8 data
- change from baseline (run-in period) in the severity of CH attacks (mild, moderate, severe, and very severe) during weeks 1, 2, 3, 4, 8, and 12 after administration of the first dose of the IMP
- mean change from baseline (run-in period) in the weekly average number of days with use of any acute headache medications during the 12-week period after administration of the first dose of the IMP, ie, based on week 0 to 12 data

- total cumulative distribution of percent reduction (≥0%, ≥10%, ≥20%, ≥30%, ≥40%, ≥50%, ≥60%, ≥70%, or ≥80%) from baseline (run-in period) in the monthly average number of CH attacks during the 12-week period after the first dose of the IMP, ie, based on week 0 to 12 data
- cumulative distribution function plot of percent reduction (≥0%, ≥10%, ≥20%, ≥30%, ≥40%, ≥50%, ≥60%, ≥70%, or ≥80%) from baseline (run-in period) in the monthly average number of CH attacks during the 12-week period after the first dose of the IMP, ie, based on week 0 to 12 data
- mean change from baseline (run-in period) in the monthly average number of CH attacks of at least severe severity during the 12-week period after administration of the first dose of the IMP, ie, based on week 0 to 12 data
- changes from baseline (day 0) in responses to questionnaires that measure quality of life, satisfaction with treatment, and health status
- the relationship between biofluid biomarkers and fremanezumab concentrations
- the relationship between specific genetic polymorphisms within the calcitonin gene-related peptide (CGRP) receptor-ligand complex and headache-associated genes versus headache response, specific CH attack clinical features (eg. severity, IMP effects, IMP efficacy), and/or safety (eg, adverse events to IMP)
- the relationship between baseline levels and/or mean changes in levels from baseline (day 0) in biofluid biomarkers versus treatment, CH response status (onset/duration), and response status (ie, responders versus nonresponders)

1.6. Wearable Sensor Substudy Endpoints

Wearable sensor substudy exploratory endpoints are the following:

- change in total sleep time, sleep onset latency, wake after sleep onset, and sleep efficiency on days with and without CH attacks
- the change in the total time of sedentary, light, moderate, and vigorous activity before (6-hour period), during, and after (6-hour period) CH attack relative to days free of attacks
- the correlation between total time in sedentary, light, moderate, and vigorous activity; total sleep time; sleep onset latency; wake after sleep onset; and sleep efficiency and the frequency of CH attacks
- the correlation between total time in sedentary, light, moderate, and vigorous activity; total sleep time; sleep onset latency; wake after sleep onset; and sleep efficiency and headache intensity during CH attacks
- the correlation between total time in sedentary, light, moderate, and vigorous activity; total sleep time; sleep onset latency; wake after sleep onset; and sleep efficiency and headache duration during CH attacks

- the average total time in sedentary, light, moderate, and vigorous activity; total sleep time; sleep onset latency; wake after sleep onset; and sleep efficiency in patients with CCH
- the correlation between the total time in sedentary, light, moderate and vigorous activity; total sleep time; sleep onset latency; wake after sleep onset; and sleep efficiency and the genetic and biomarker profiles
- detection of time-dependent patterns composed of lower- and higher-level activity metrics (eg, activity score and sleep/wake patterns)
- cross-correlation between the time-dependent activity patterns and CH attacks
- prediction model for the coming CH attack, intensity, and duration. The input to the prediction model (ie, features) will be selected from the measurements collected by the digital wearable devices, information collected by the self-reporting devices, and aggregated information provided by the calculated endpoints (ie, various statistics)

2. STUDY DESIGN

2.1. General Design

This is a 16-week, multicenter, randomized, double-blind, double-dummy, placebo-controlled, parallel-group study to compare the efficacy and safety of 2 dose regimens of fremanezumab versus placebo in adult patients for the prevention of CCH. The study will consist of a screening visit, a run-in period lasting approximately 4 weeks (+3 days), and a 12-week double-blind treatment period. During the course of any CH attack, patients will be allowed to use acute medications to treat acute headaches, as needed.

Patients will complete a screening visit (visit 1) after providing written informed consent, and eligible patients will enter a run-in period lasting at least 4 weeks (+3 days) during which they will enter baseline CH attack information into an electronic diary device daily. Patients will return to the study center after completing the run-in period (visit 2 [week 0]). Patients who have at least 10 CH attacks during the run-in period and who continue to meet eligibility criteria (including entry of CH attack information in an electronic diary demonstrating compliance for 85% of days during the run-in period) will be randomly assigned at visit 2 (week 0) in a 1:1:1 ratio to 1 of 3 treatment groups (see details in Section 2.2).

Blinded treatment will be administered once monthly (ie, approximately every 4 weeks) for a total of 3 months. Final study assessments will be performed at the final visit for this study (visit 5), approximately 12 weeks after administration of the first dose of the IMP. Upon completion of the final study assessments, early withdrawal from the study, or discontinuation for any reason, patients will be offered the opportunity to enter a 32 week long-term safety study (as described in Study TV48125-CNS-30058) for safety and ADA evaluation without additional dosing.

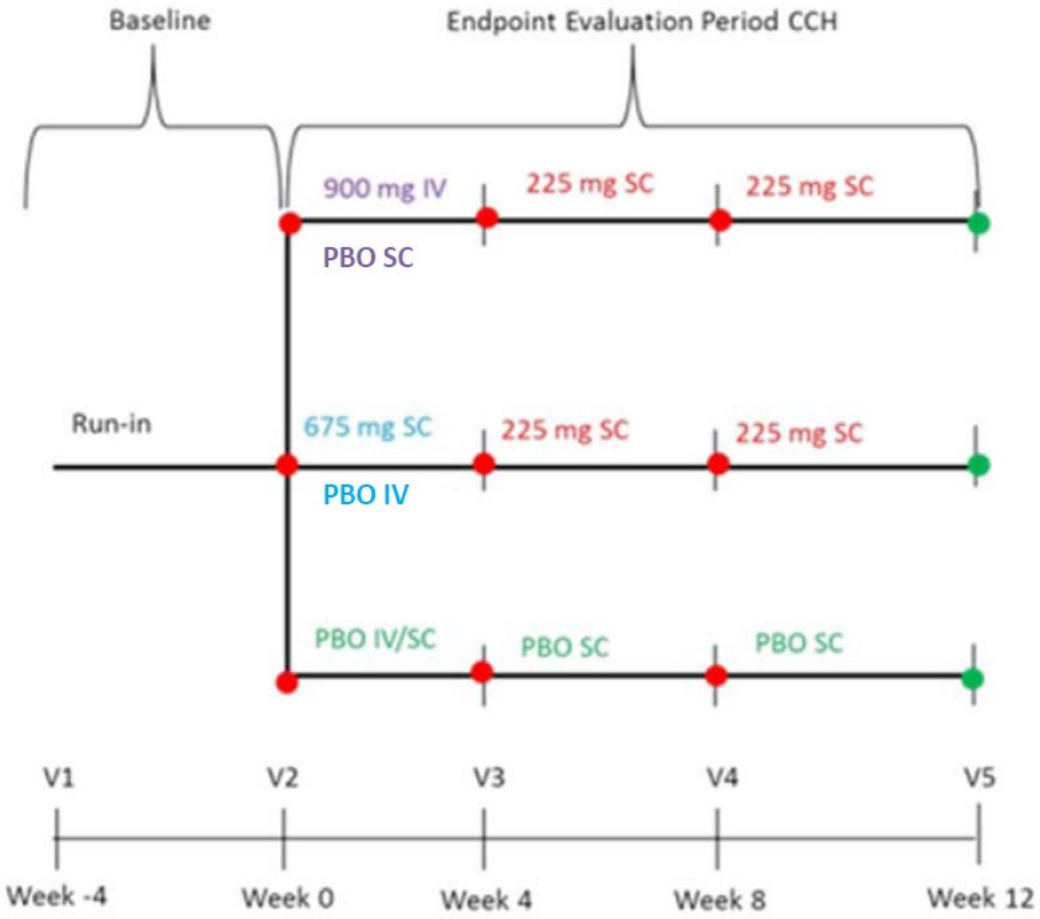
Patients who satisfactorily complete the study may be offered to enroll the long-term safety Study TV48125-CNS-30058 for 68 weeks (as described in this study protocol) to receive additional dosing and a final follow-up visit for safety and ADA evaluation. In any case, during the period of the long-term safety study, where patients are not receiving additional dosing (and are waiting for ADA evaluation), these patients should be treated with standard of care as appropriate. A separate protocol was issued for the long-term safety study.

CH attack information will be captured daily during the double-blind treatment period using an electronic diary device. Assessments of change in quality of life and health status (using the Hospital Anxiety and Depression Scale [HADS], EuroQol-5 Dimension [EQ-5D] questionnaire, 12-Item Short-Form Health Survey [SF-12], Impact on Partner and Family questionnaire, and Work Productivity and Activity Impairment [WPAI] questionnaire); satisfaction with treatment (using the PPSI and Patient Global Impression of Change [PGIC] scale); safety evaluations (including eC-SSRS); blood collection for pharmacokinetics, immunogenicity, biomarker, and pharmacogenomics (unless not allowed per local regulation) analyses; and urine sampling for biomarker analysis will be performed at prespecified time points.

The end of study is defined as the date the last patient attends the end of treatment (EOT)/early withdrawal visit. The study duration will be approximately 24 months from Q4/2016 to Q3/2018.

The study schematic diagram is presented in Figure 1. Study procedures and assessments with their timing are summarized in Table 2 of the study protocol.

Figure 1: Overall Study Schematic Diagram



CCH=chronic cluster headache; IV=intravenous; PBO=placebo; SC=subcutaneous; V=visit.
Note: Patients randomized to the 900-mg iv loading dose group will receive 900 mg of fremanezumab administered via an approximately 1-hour iv infusion followed by 3 placebo sc injections at visit 2 (week 0) and fremanezumab at 225 mg administered as single sc injections at visits 3 and 4 (weeks 4 and 8, respectively). Patients randomized to the 675-mg sc loading dose group will receive placebo administered via an approximately 1-hour iv infusion and fremanezumab at 675 mg administered as 3 sc injections (225 mg/1.5 mL) at visit 2 (week 0) and fremanezumab at 225 mg administered as single sc injections at visits 3 and 4 (weeks 4 and 8, respectively). Patients randomized to the placebo group will receive placebo as an approximately 1-hour iv infusion followed by placebo administered as 3 placebo sc injections at visit 2 (week 0) and placebo as single sc injections at visits 3 and 4 (weeks 4 and 8, respectively).

2.1.1. Wearable Sensor Substudy

A subset of patients in selected investigational sites will be offered the opportunity to participate in a substudy to understand the utility of physiological biomarker measures captured through wearable digital sensor devices as tools to monitor response to treatment and disease symptoms (eg, activity/sleep disruption).

Patients who are able to demonstrate appropriate use of the wearable device and are willing to comply with the requirements for use of the digital wearable device will be given the device and accessories at the baseline visit. The device will be worn continuously throughout the 12-week treatment period, and for patients who continue into the long-term safety study (Study TV48125-CNS-30058), the device will be worn continuously throughout the 40-week treatment period of that study. Refer to Section 8.6 of the study protocol for additional details.

2.2. Randomization and Blinding

Patients will be randomly assigned with stratification based on gender, country, and baseline concomitant preventive medication use (yes or no) in a 1:1:1 ratio to 1 of 3 treatment groups.

- **fremanezumab 900-mg intravenous (iv) loading dose group**: fremanezumab at 900 mg administered via an approximately 1-hour iv infusion and 3 placebo subcutaneous (sc) injections at visit 2 (week 0) followed by fremanezumab at 225 mg administered as single sc injections (225 mg/1.5 mL) at visits 3 and 4 (weeks 4 and 8, respectively)
- **fremanezumab 675-mg sc loading dose group**: placebo administered via an approximately 1-hour iv infusion and fremanezumab at 675 mg administered as 3 sc injections (225 mg/1.5 mL) at visit 2 (week 0) followed by fremanezumab at 225 mg administered as single sc injections (225 mg/1.5 mL) at visits 3 and 4 (weeks 4 and 8, respectively)
- **placebo group**: placebo administered via an approximately 1-hour iv infusion and 3 placebo sc injections at visit 2 (week 0) followed by placebo administered as single sc injections at visits 3 and 4 (weeks 4 and 8, respectively)

The sponsor, investigators, study staff (except for staff involved in bioanalytical analyses and interim analyses [by a third-party, unblinded statistician]), and patients will be blinded to treatment assignment throughout the study.

Randomization will be performed using electronic interactive response technology (IRT), a third-party vendor. The randomization code will be generated by IRT following specifications from the Biostatistics Department. A Teva statistician will be responsible for reviewing the dummy randomization codes, and the final randomization code will be maintained by the third-party vendor in a secure location.

2.3. Data Monitoring Committee

Not applicable.

2.4. Sample Size and Power Considerations

A sample size of 258 patients (86 evaluable patients completing the study per treatment group) will provide at least 90% power to detect a treatment difference of 4 CH attacks in the monthly average (assuming a common standard deviation [SD] of 8 CH attacks) at a 2-sided alpha level of 0.05. Assuming a 14% discontinuation rate, approximately 300 patients will be randomized in the trial.

2.5. Sequence of Planned Analyses

2.5.1. Planned Interim Analyses

An interim analysis for futility evaluation will be performed once 50% of patients (the first 150 patients) have completed 12-week assessments during the double-blind study period or have withdrawn from the study early. An independent statistician from a third party will perform evaluation. Details are provided in Section 15.

2.5.2. Final Analyses and Reporting

All analyses identified in this SAP will be performed after the final database lock for study completion or for early termination of study as result of futility evaluation at interim analysis. Details are provided in Section 15.

The study will not be unblinded until the final database lock and finalization of the SAP.

3. ANALYSIS SETS

3.1. Intent-to-Treat Analysis Set

The intent-to-treat (ITT) analysis set will include all randomized patients.

In the ITT analysis set, treatment will be assigned based on the treatment to which patients were randomized, regardless of which treatment they actually received.

3.2. Safety Analysis Set

The safety analysis set will include all randomized patients who receive at least 1 dose of the IMP.

In the safety analysis set, treatment will be assigned based on the treatment patients actually received, regardless of the treatment to which they were randomized, unless otherwise specified.

3.3. Full Analysis Set

The full analysis set (FAS) will include all patients in the ITT analysis set who receive at least 1 dose of the IMP and have at least 10 days of postbaseline efficacy assessment in the 12-week on the primary endpoint.

3.4. Per-Protocol Analysis Set

The per-protocol (PP) analysis set will consist of all patients who have completed the study without any violations of the inclusion/exclusion criteria or any violations or omissions of the drug administration.

The efficacy analysis for the primary and secondary endpoints will be repeated for the per-protocol analysis set.

4. GENERAL ISSUES FOR DATA ANALYSIS

4.1. General

Descriptive statistics for continuous variables include count (n), mean, SD, standard error (SE), median, minimum, and maximum. In addition, for fremanezumab concentration, percentage coefficient of variation (%CV) and geometric mean will also be calculated. Descriptive statistics for categorical variables include patient counts and percentages, and a missing category will be displayed as appropriate.

Summaries of potentially clinically significant abnormal values for clinical laboratory tests and vital signs values will include all postbaseline values (including scheduled, unscheduled, and early withdrawal visits).

4.2. Specification of Baseline Values

Baseline is the last observed data before the administration of the first dose of the IMP, unless otherwise noted. For data collected in the e-diary daily, baseline will be derived from the run-in period (including days from the informed consent up to the day prior to the first dose of the IMP administration) and prorated to 28 days if the number of days in the run-in period is not equal to 28. Details are provided in Section 6.1.

4.3. Region of Pooled Countries

The study is planned to be conducted in approximately 12 countries. The countries will be pooled to 2 regions (US/Canada and other).

4.4. Handling Withdrawals and Missing Data

For efficacy analyses using e-diary data, the missing data handling methods are provided in Section 6.1.4. For the efficacy analyses using non-e-diary data, missing data handling methods, if applicable, will be provided in the efficacy analysis section for that endpoint.

Dates that have incomplete information, such as only the month and year or just the year, will be estimated for the purpose of calculating variables that are dependent on time if necessary. Day will be estimated as the first day (01) of the month (if month and year of partial date are available) or middle (July 1) of the year (if only year is available), month will be estimated as July if year and day are available, unless otherwise noted. The imputations for partial dates are only for calculation purpose. Original date variables will not be imputed. Listings will list dates as collected.

Patients' date of birth will be collected by IRT. IRT will record the date of birth as the first day of the first month (January first) when only year is captured, and the first day of the selected month when month and year are captured.

4.5. Study Days and Visits

For by-visit summaries, if there are multiple assessments at a postbaseline visit then the last non-missing assessment at that visit will be used for the summary (this includes scheduled and unscheduled assessments), except for triplicate ECG assessments (see Section 8.11 for further details).

Study visits are detailed in Table 1.

Table 1: Study Visits

Visit #	**V1**	**V2**	**V3**	**V4**	**V5**
Week #	**Week -1**	**Week 0**	**Week 4**	**Week 8**	**Week 12**
Visit	**Screening**	**Baseline**	**Visit 3**	**Visit 4**	**EOT/early withdrawal (84 ±3 days)**

Notes: EOT=end of treatment; V2/Week 0 is day 1, which is the day eligible patients are randomized and receive first dose of IMP.

'Last Assessment' may be derived for analysis purpose and is defined as the last observed postbaseline data. For patients who withdraw from the study early, their data at the early withdrawal visit will be excluded from the by-visit sections but will be included in the Last Assessment section.

Study days are numbered relative to the first day of the IMP administration. The start of treatment (day 1) is defined as the date on which a patient takes the first dose of the IMP, as recorded on the case report form (CRF). Days will be numbered relative to treatment start (ie, ..., –2, –1, 1, 2, ...; with day 1 being the first day of the IMP administration and day –1 being the day before the first day of the IMP administration).

For data from the e-diary, weekly analysis windows (week 1, week 2, etc.) and 4-week/monthly analysis windows (weeks 1 to 4, weeks 5 to 8, and weeks 9 to 12) will be derived for the purpose of efficacy endpoint analyses. Details are provided in Section 6.1.3.

5. STUDY POPULATION

5.1. General

The ITT analysis set (see Section 3.1) will be used for all study population summaries unless otherwise specified. Summaries will be presented by treatment group and for all patients.

5.2. Patient Disposition

Data from patients screened; patients screened but not randomized and reason for not randomized; patients who are randomized; patients randomized but not treated; patients in the ITT, safety, and other analysis sets; patients who complete the study; patients who withdraw from the study; and patients continuing into the long-term safety study will be summarized using descriptive statistics. Data from patients who withdraw from the study will also be summarized by reason for withdrawal using descriptive statistics.

Patients who consent to the wearable sensor substudy and their completion status (completed and reason for not-completed the substudy) will be summarized using descriptive statistics.

This summary will include all patients.

5.3. Demographics and Baseline Characteristics

The demographic data will be collected at the screening visit after the patient signs the informed consent form. Patient's demographics data including age, age group (<40 years or ≥40 years), gender, race, race group (white or other), ethnicity, region (US/Canada or other), baseline weight (kg), baseline height (cm), and baseline body mass index (kg/m^2) will be summarized using descriptive statistics for all analysis sets. For continuous variables, treatment groups will be compared using an analysis of variance (ANOVA) with treatment group as a factor. For categorical variables, treatment groups will be compared using a Pearson's chi-square (or Fisher's exact test if cell sizes are too small).

Baseline characteristics including years since first CCH diagnosis, preventive medication use (yes or no) at screening or baseline, use of any triptans/ergots during the run-in period (yes or no), and number of CH attacks during the run-in period will be summarized for the ITT analysis set using descriptive statistics. No inferential analyses will be performed.

The years since first CCH diagnosis will be calculated as (date of informed consent - first date of CCH diagnosis + 1)/365.25. Rules for handling partial CCH diagnosis date are in Section 4.4.

5.4. Medical History

All medical history abnormalities will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The incidence of medical history abnormalities will be summarized using descriptive statistics by system organ class (SOC) and preferred term. Patients are counted only once in each SOC and only once in each preferred term.

5.5. Prior Therapy and Medication

All prior medications or therapy will be coded using the World Health Organization Drug Dictionary of medical codes (WHO Drug). The incidence of prior medications or therapy will be summarized by therapeutic class and preferred term using descriptive statistics. Patients are counted only once in each therapeutic class category, and only once in each preferred term category. Prior medications will include all medications taken prior to the administration of the first dose of the IMP.

The prior medications will be summarized by the following indications categories:

- preventive medication from Appendix H of the study protocol for CH
- preventive medication from Appendix H of the study protocol for other reason than CH
- butalbital for CH
- butalbital for other reason than CH
- triptans for CH
- triptans for other reason than CH
- ergots for CH
- ergots for other reason than CH
- non-steroidal anti-inflammatory drugs (NSAIDs) for CH
- NSAIDs for other reason than CH
- opioids for CH
- opioids for other reason than CH
- other

5.6. Childbearing Potential and Methods of Contraception

Information related to reproductive system findings will be collected at the screening visit. Data will be listed.

5.7. Physical Examinations

Patients with at least 1 abnormal finding (overall) and abnormal findings for each category will be summarized.

5.8. Study Protocol Deviations

Data from patients with any important protocol deviations during the study will be summarized overall and for each category using descriptive statistics.

6. EFFICACY ANALYSIS

6.1. General

The FAS will be used for all efficacy analyses unless otherwise noted. Analyses of the primary and secondary endpoints will be repeated for the PP analysis set.

For data not collected daily, baseline is the last observed data before the administration of the first dose of the IMP. For data collected daily in the e-diary, baseline will be derived from the run-in period (including days from the informed consent up to the day prior to the first dose of the IMP administration). Details are provided in Section 6.1.3. The efficacy baseline value derived from the run-in period are

- number of CH attacks
- number of CH attacks of at least severe severity
- number of days with use of cluster-specific acute headache medication (triptans and ergot compounds)
- number of days with use of any acute headache medication
- number of days oxygen is used
- severity of CH attacks

Summaries will be presented by treatment group as randomized unless otherwise noted. In addition to inferential statistics, descriptive statistics will be presented in by-visit or by-analysis-window summaries.

6.1.1. Cluster Headache Attacks Data from E-diary

CH attacks will be recorded in the e-diary daily. Corresponding CH attack questions are in Appendix A (questions A1, A2, and B1 through B15). Patients will record headache(s) occurred during last 48 hours. The recording window on e-diary device is “Yesterday morning 6:00 am to today at 5:59 am” or “The day before yesterday morning 6:00 am to yesterday at 5:59 am”; a maximum of 48 hours of recall time will be accepted to record CH attacks for the day before yesterday.

The derivation logics for CH attacks are in Appendix B. A CH attack will be endorsed if required criteria in Appendix B are met.

Baseline and postbaseline average number of CH attacks for a patient within an analysis window will be calculated based on data available and prorated to 28 days using formula [1A] as the monthly average for analyses based on monthly windows or prorated to 7 days using formula [1B] as the weekly average or weekly value for analyses based on weekly average or weekly windows.

6.1.2. Data Derivation

The **monthly *average*** of an efficacy variable will be calculated and prorated to 28 days as follows:

$$\frac{\sum efficacy\ endpoint\ data\ during\ the\ period}{Number\ of\ Days\ with\ assessments\ recorded\ in\ the\ eDiary\ during\ the\ period} \times 28 \quad [1A]$$

The **weekly *average or weekly value*** of an efficacy variable will be calculated and prorated to 7 days as follows:

$$\frac{\sum efficacy\ endpoint\ data\ during\ the\ period}{Number\ of\ Days\ with\ assessments\ recorded\ in\ the\ eDiary\ during\ the\ period} \times 7 \quad [1B]$$

The “efficacy endpoint data” will be “number of CH attacks”, “days with use of cluster-specific acute headache medications (tripants and ergot compounds)”, “days oxygen is used to treat CCH”, or “days with use of any headache medication”, etc.

The period will be the run-in period (including days from informed consent up to the day prior to the first dose of the IMP administration) for baseline calculation or an analysis window (eg. weeks 1 to 4, weeks 5 to 8, weeks 9 to 12, etc.) for post-baseline calculations.

The ***percentage of reduction*** from baseline in the monthly average number of CH attacks will be calculated as follows:

$$\frac{baseline\ value - postbaseline\ value}{baseline\ value} \times 100\% \quad [2]$$

where baseline value and postbaseline value are the monthly average calculated using [1A].

Note, the calculation of a proportion based on the weekly average or monthly average would yield the same results. Therefore to be consistent, all proportions will be calculated based on the monthly average number of CH attacks.

6.1.3. Analysis Windows for E-diary Data

Baseline window is the run-in period and includes days from the informed consent up to the day prior to the first dose of the IMP administration. The baseline average value for an efficacy variable will be calculated based on data recorded in the run-in period and prorated to 28 days using formula equation [1A] as the monthly average for analyses based on monthly windows or prorated to 7 days using formula [1B] as the weekly average or weekly value for analyses based on weekly average or weekly windows.

Note, per eligibility checks at visit 2, patients with more than 28+3 days in the run-in period will not be eligible and will not be randomized for the study. However, in case there are patients with more than 31 days of the e-diary data collected during the run-in period, all data will be used for the baseline calculation by prorating the data to 28 days for the monthly average or 7 days for the weekly average.

Postbaseline 4-week (monthly) analysis windows will be determined based on the actual dosing date as follows:

- weeks 1 to 4: from the first dose of the IMP administration to the day before IMP administration at visit 3 (week 4); for endpoints based on week 0 to 4 data
- weeks 5 to 8: from the IMP administration at visit 3 to the day before the IMP administration at visit 4 (week 8); for endpoints based on week 4 to 8 data

- weeks 9 to 12: from the IMP administration at visit 4 to the day before the EOT visit (visit 5/week 12); for endpoints based on week 8 to 12 data

Notes:

There will be no IMP administration at visit 5, and patients will return the e-diary at the visit; for patients who withdraw from the study early, the end day for the last window will be the day before EOT/early withdrawal visit; a 4-week analysis window may contain <28, 28, or >28 days.

The monthly average of an efficacy variable for each analysis window will be calculated based on data recorded in the e-diary and prorated to 28 days using formula [1A]. Missing data handling methods are provided in Section 6.1.4.

Postbaseline 1-week (weekly) analysis windows will be determined based on days in the 4-week analysis windows (derived based on the actual dosing date as described above) as follows:

- week 1 = day 1 to day 7
- week 2 = day 8 to day 14
- week 3 = day 15 to day 21
- week 4 = day 22 to day 28*

of analysis window weeks 1 to 4
(Month 1; Dose 1 to Dose 2)

- week 5 = day 1 to day 7
- week 6 = day 8 to day 14
- week 7 = day 15 to day 21
- week 8 = day 22 to day 28*

of analysis window weeks 5 to 8
(Month 2; Dose 2 to Dose 3)

- week 9 = day 1 to day 7
- week 10 = day 8 to day 14
- week 11 = day 15 to day 21
- week 12 = day 22 to day 28*

of analysis window weeks 9 to 12
(Month 3; Dose 3 to EOT)

* This analysis window may contain <7, 7, or >7 days.

For patients who withdraw from the study early, the last analysis window may be <7 days.

The weekly value of an efficacy variable for each analysis window will be calculated based on data recorded in the e-diary and prorated to 7 days using formula [1B] if applicable. Missing data handling methods are provided in Section 6.1.4.

6.1.4. Handling Missing E-diary Data

This section includes missing data handling for e-diary data. For other type of data, the missing data handling methods are provided in the analysis section for the endpoint if applicable.

6.1.4.1. Primary Efficacy Endpoint

The monthly average number of CH attacks during 12-week period will be calculated based on data available and prorated to 28 days using formula [1A]. Note, patients who have < 10 days of e-diary data will be excluded from FAS and PP analysis set (Section 3).

As sensitivity analyses, the primary endpoint will be analyzed using the following methods for handling missing data :

- Multiple imputation (MI) method to impute missing weekly data
- An mixed model for repeated measures (MMRM) method to analyze the weekly data

Detail is provided in Section 6.2.3.

6.1.4.2. Analyses Based on Monthly and 12-week Window

In analyses using analysis of covariance (ANCOVA) based on 12-week window, the monthly average of an efficacy variable will be calculated based on data available and prorated to 28 days using formula [1A]; the weekly average of an efficacy variable will be calculated based on data available and prorated to 7 days using formula [1B].

Note, patients who have < 10 days of e-diary data will be excluded from the FAS and PP analysis set (Section 3).

In analyses using MMRM based on monthly analysis windows

- for patients who have ≥10 days of e-diary data in an analysis window, the monthly average number of CH attacks will be calculated based on data available in that analysis window and prorated to 28 days using formula [1A];
- for patients who have <10 days of e-diary data in an analysis window, the monthly average number of CH attacks for the analysis window will be considered as missing.

In analyses based on proportion based on weeks 0 to 12, the average number of CH attacks will be calculated based on data available and prorated to 28 days using formula [1A] as the monthly average. Response status will be determined based on the prorated average using formula [2].

6.1.4.3. Analyses of Mean Change from Baseline Using MMRM Method Based on 1-Week Window

For analyses of mean change from baseline using an MMRM method based on 1-week analysis window, missing data will be handled as follows:

- For patients who have ≥3 days of e-diary data in a week, the weekly value of an efficacy variable will be calculated based on data available and prorated to 7 days using formula [1B].
- Otherwise the weekly value of an efficacy variable will set to missing.

6.2. Primary Efficacy Endpoint and Analysis

6.2.1. Definition

The primary efficacy endpoint of this study is the mean change from baseline (run-in period) in the monthly average number of CH attacks during the 12-week period after administration of the first dose of the IMP.

CH attacks will be derived from the e-diary data as described in Section 6.1.1. The 12-week period is the period from administration of the first dose of the IMP to the EOT/early withdrawal visit. Missing data handling methods are provided in Section 6.1.4.1.

6.2.2. Primary Efficacy Analysis

The hypothesis testing for the primary analysis is:

$$H_o : \delta_1 = \delta_2 \quad vs \quad H_a : \delta_1 \neq \delta_2$$

where δ_1 and δ_2 are the estimates of mean change from baseline in the monthly average number of CH attacks for the fremanezumab treatment group and the placebo group, respectively. To control overall type I error rate, the comparison of each active treatment group (fremanezumab 900-mg iv loading dose group and fremanezumab 675-mg sc loading dose group) versus the placebo will be tested using Hochberg's step-up method as described in Section 7.

The primary endpoint will be analyzed using an ANCOVA method. The model will include baseline preventive medication use (yes or no), gender, region (US/Canada or other), and treatment as fixed effects; the baseline number of CH attacks as a covariate. The least square (LS) mean and SE for each treatment group, LS means and corresponding 95% confidence intervals for the treatment differences (fremanezumab - placebo), and associated p-values will be presented.

The analysis will be performed on the FAS and PP analysis sets.

Example SAS code for ANCOVA

```
PROC MIXED DATA=<name>;
   CLASS TRTP SEX REGION BPMU;
   MODEL CHG= BASE BPMU SEX REGION TRTP /S ;
   LSMEANS TRTP /DIFF=CONTROL("Placebo") CL ALPHA=0.05;
   ODS OUTPUT LSMEANS= <name>  DIFFS= < name> ;
RUN;
```

Where TRTP denotes the planned treatment group; BPMU denotes the baseline preventive medication use (yes or no); CHG denotes the change from baseline; and BASE denotes the baseline number of CH attacks. In bold are the SAS key words.

6.2.3. Sensitivity Analyses for the Primary Efficacy Endpoint

Sensitivity analyses will be conducted to explore the impact of missing data in the primary efficacy analysis.

6.2.3.1. Analysis with Multiple Imputation Method

An MI method will be utilized to impute the monthly missing data. Monthly analysis windows will be derived using the algorithm described in Section 6.1.3. The number of CH attacks within each monthly analysis window will be calculated and prorated to 28 days using formula [1A].

Data with the monthly average number of CH attacks will be processed using the following steps:

- For patients with partial e-diary data (ie, have <10 days of data) in a monthly window, the number of CH attacks in that monthly window will be reset to missing before the MI procedure.
- For patients in the active treatment groups who withdraw from the study early with reasons of adverse event or lack of efficacy, missing values will be imputed using data from the placebo-treated patients by assigning these patients to the placebo group for data imputation purpose.
- Run SAS PROC MI procedure to create 10 sets of data.
- Within each imputed data set, for patients with partial e-diary data in a monthly window (ie, <10 days of data), the number of CH attacks in the monthly window will be replaced by

  $$\sum(\text{observed } CH\ attacks\ during\ a\ monthly\ window\ ) + (28 - X) * \frac{imputed\ value}{28}$$

  Where X denotes the number of days with e-diary within a monthly window.
- The monthly average number of CH attacks ***during the 12-week period*** after the first dose of the IMP will be the average of the values from 3 monthly analysis windows.

Each set of imputed data will be analyzed using an ANCOVA method in a manner analogous to the primary endpoint as described in Section 6.2.2. The LS means and SEs from each analysis will be output to a SAS data set.

The SAS MIANALYZE procedure will be used to generate the final LS means and SE for the treatment groups and the treatment differences (fremanezumab - placebo) as well as p-values associated with treatment differences. The 95% confidence intervals for the treatment differences will also be constructed.

The analysis will be performed on the ITT analysis set.

6.2.3.2. Mixed Model for Repeated Measures Analysis

MMRM analysis will be utilized to estimate the mean change from baseline in the monthly average number of CH attacks during the 12-week period after administration of the first dose of the IMP.

Postbaseline data will include data from three monthly analysis windows derived using the algorithm described in Section 6.1.3. Missing data will be handled as described in Section 6.1.4.2.

The MMRM will include baseline preventive medication use (yes or no), gender, region (US/Canada or other), treatment, month (months 1 to 3), and month-by-treatment interaction as fixed effects and baseline number of CH attacks as a covariate. The unstructured covariance structure (UN) will be used to model intra-subject correlation. LS mean and SE for each treatment group, LS means and corresponding 95% confidence intervals for the treatment differences (fremanezumab - placebo), and the associated p-values from the overall results for treatment comparisons on the average of months 1 to 3 will be presented to support the primary analysis.

Example SAS codes for MMRM analysis:

```
PROC MIXED DATA=<name> METHOD=REML;
    CLASS USUBJID SEX BPMU REGION ATPTN TRTP;
    MODEL CHG=BASE BPMU SEX REGION TRTP ATPTN TRTP*ATPTN/S;
    REPEATED ATPTN/ SUB=USUBJID TYPE=UN;
    LSMEANS TRTP TRTP*ATPTN/ DIFF CL ALPHA=0.05;
    ODS OUTPUT LSMEANS= <name> DIFFS= < name>;
RUN;
```

Where TRTP denotes the planned treatment group; BPMU denotes the baseline preventive medication use (yes or no); CHG denotes the change from baseline; BASE denotes the baseline number of CH attacks; and ATPTN denotes the month or the analysis window.

The analysis will be performed on the FAS.

6.2.4. Subgroup Analyses

The following subgroup analyses will be performed for primary endpoint and selected secondary endpoints (mean change from baseline in the number of CH attacks [the 2nd and 3rd secondary endpoints]; mean change from baseline in the number of days with use of cluster-specific acute headache medication [the 4th secondary endpoint]; mean change from baseline in the number of days oxygen is used [the 5th secondary endpoint]):

- Age group (< 40 and ≥40 years old)
- Race group (white and other)
- Sex (male and female)
- Baseline preventive medication use (yes and no)
- Region (US/Canada and other)

Data will be analyzed in a manner analogous to the method described in the section for each of these endpoints.

Modification to SAS codes: adding a ‘by’ statement for the subgroup variable; If the subgroup variable is in MODEL statement, remove the subgroup variable from MODEL statement accordingly.

The subgroup analyses will be performed on the FAS.

6.3. Secondary Efficacy Endpoints and Analyses

The secondary efficacy endpoints are listed in Section 1.2. Analyses will be based on the FAS and PP analysis set.

As exploratory analyses, endpoints analyzed using an ANCOVA method will be also analyzed using MMRM method in a manner analogous to the analysis as described in Section 6.2.3.2.

6.3.1. Proportion of Patients with a ≥50% Reduction from Baseline in the Monthly Average Number of CH Attacks During the 12-Week Period After the First Dose of the IMP

This secondary efficacy endpoint will be analyzed using Cochran-Mantel-Haenszel (CMH) test stratified by baseline preventive medication use (yes or no). Descriptive statistics (count and percent) and p-value for Row Mean Scores difference will be presented.

The 12-week period is the period from administration of the first dose of the IMP to the EOT/early withdrawal visit. CH attacks within the 12-week period will be derived from the e-diary data as described in Section 6.1.1 and prorated to 28. Missing data handling methods are provided in Section 6.1.4.2.

The proportion of reduction for a patient will be calculated using formula [2]. Responders will be those with ≥50% reduction.

<u>Example SAS code for CMH:</u>

```
PROC FREQ DATA=<name>;
   TABLES BPMU*TRTP*AVALC / CMH;
   OUTPUT OUT=<name> CMH;
RUN;
```

Where TRTP denotes the planned treatment group; BPMU denotes the baseline preventive medication use (yes or no); and AVALC denotes the responses with yes for responders (with ≥50% CH attacks reduction) and no for non-responders (with <50% CH attacks reduction).

6.3.2. Mean Change from Baseline in the Number of Cluster Headache Attacks

The following 2nd and 3rd secondary efficacy endpoints will be analyzed based on the monthly average number of CH attacks using an MMRM method in a manner analogous to the analysis as described in Section 6.2.3.2.

- the mean change from baseline (run-in period) in the number of CH attacks during the 4-week period after administration of the first dose of the IMP, ie, based on week 0 to 4 data
- the mean change from baseline (run-in period) in the number of CH attacks during the 4-week period after administration of the third dose of the IMP, ie, based on week 8 to 12 data

Postbaseline data will include data from three 4-week analysis windows (weeks 1 to 4, weeks 5 to 8, and weeks 9 to 12) derived using the algorithm described in Section 6.1.3. Data will be prorated to 28 days as the monthly average. CH attacks within each analysis window will be derived from the e-diary data as described in Section 6.1.1. Missing data handling methods are provided in Section 6.1.4.2

The analysis will be based on data from all three 4-week (monthly) analysis windows. Results for each endpoint will be subsetted from PROC MIXED outputs. For the 2nd secondary endpoint, the results from the first analysis window (weeks 1 to 4) will be presented; for the 3rd secondary endpoint, the results from the last analysis window (weeks 9 to 12) will be presented.

6.3.3. Mean Change from Baseline in the Weekly Average Number of Days with Use of Cluster-Specific Acute Headache Medications (Triptans and Ergot Compounds) During the 12-Week Period After Administration of the First Dose of the IMP

Use of triptans and ergot compounds for CH will be recorded in the e-diary. The 12-week period is the period from administration of the first dose of the IMP to the EOT/early withdrawal visit. The weekly average number of days with use of triptans or/and ergots for CH for a patient during 12-week period will be calculated and prorated to 7 days using formula [1B]. Missing data handling method is provided in Section 6.1.4.2.

Data will be analyzed using an ANCOVA method in a manner analogous to the primary endpoint as described in Section 6.2.2.

6.3.4. Mean Change from Baseline in the Weekly Average Number of Days Oxygen is Used to Treat Chronic Cluster Headaches During 12-week Period After Administration of the First Dose of the IMP

Use of oxygen will be recorded in the e-diary. The 12-week period is the period from administration of the first dose of the IMP to the EOT/early withdrawal visit. The weekly average number of days using oxygen for a patient during the 12-week period will be calculated and prorated to 7 days using formula [1B]. Missing data handling method is provided in Section 6.1.4.2.

Data will be analyzed using an ANCOVA method in a manner analogous to the primary endpoint as described in Section 6.2.2.

6.3.5. Patient-Perceived Satisfactory Improvement at 1, 4, 8, and 12 Weeks After Administration of the First Dose of the IMP Relative to Baseline

The PPSI scale will be completed in the e-diary at home at week 1(day 7) and in the investigational site tablet at the visits 2, 3, 4, and 5. Patients will mark the level of CH-associated pain and indicate the level of pain using the following scale compared with 4 weeks ago:

- 1 = Much worse
- 2 = Moderately worse
- 3 = Slightly worse
- 4 = Unchanged
- 5 = Slightly improved
- 6 = Moderately improved

- 7 = Much improved

PPSI will be defined as change in pain that corresponds with a minimal rating of "slightly improved" (Protocol Section 6.7). For analysis purpose, a dichotomous scale of "Responder" (scales 5 to 7) or "Non-responder" (scales 1 to 4) will be derived.

The percentage of patients in dichotomous scale of "Responder" or "Non-responder" at week 1, visit 3, 4, and 5 will be analyzed using a CMH method in a manner analogous to the first secondary endpoint as described in Section 6.3.1.

Modification to SAS code in Section 6.3.1: adding a 'by' statement for visit.

Raw scales will be summarized using descriptive statistics. A missing category will be presented if applicable.

6.4. Exploratory Efficacy Endpoints and Analysis

Exploratory efficacy endpoints are listed in Section 1.5. Analyses will be based on the FAS.

As exploratory analyses, endpoints analyzed using an ANCOVA method will be also analyzed using MMRM method in a manner analogous to the analysis as described in Section 6.2.3.2.

6.4.1. Proportion of Patients with ≥50%, 75%, or 100% Reduction from Baseline in Number of CH Attacks During the 12-Week Period (First 3 Exploratory Efficacy Endpoints)

The following exploratory efficacy endpoint will not be analyzed since the result would be the same as these from the first secondary endpoint (≥50 reduction in monthly average). Note, the calculation of a proportion based on the weekly average or monthly average would yield the same result.

- proportion of patients with ≥50% reduction from baseline (run-in period) in the weekly average number of CH attacks during the 12-week period after administration of the first dose of the IMP, ie, based on week 0 to 12 data

The following exploratory efficacy endpoints will be analyzed using a CMH method in a manner analogous to the first secondary endpoint as described in Section 6.3.1. The proportion of reductions will be calculated based on **monthly average number** of CH attacks to be consistent with the calculation of the first secondary endpoint.

- proportion of patients with ≥75% reduction from baseline (run-in period) in the weekly average number of CH attacks during the 12-week period after administration of the first dose of the IMP, ie, based on week 0 to 12 data
- proportion of patients reaching 100% reduction in the number of CH attacks during the 12-week period after administration of the first dose of the IMP, ie, based on week 0 to 12 data

The 12-week period is the period from administration of the first dose of the IMP to the EOT/early withdrawal visit. CH attacks within the 12-week period will be derived from the e-diary data as described in Section 6.1.1 and prorated to 28 days using formula [1A]. Missing data handling methods are provided in Section 6.1.4.2.

The proportion of reduction for a patient will be calculated using formula [2]. Responders will be those with ≥75% or 100% reduction, respectively.

The monthly analyses will be presented.

6.4.2. Mean Change from Baseline in the Number of CH Attacks During Weeks 1, 2, 3, 4, 8, and 12 After Administration of the First Dose of the IMP

This exploratory efficacy endpoint will be analyzed using an MMRM in a manner analogous to the analysis as described in Section 6.2.3.2.

Postbaseline data will include data from twelve 1-week analysis windows derived using the algorithm described in Section 6.1.3. Missing data handling methods are provided in Section 6.1.4.3. If days with e-diary data in a 1-week analysis window are more than 7, CH attacks for the analysis window will be prorated to 7 days using formula [1B].

Analyses will be based on data from all 12 analysis windows. Results for weeks 1, 2, 3, 4, 8, or 12 will be subsetted from PROC MIXED outputs afterwards.

LSMEANS statement in PROC MIXED will be changed to:

LSMEANS TRTP*ATPTN/ **DIFF CL ALPHA=0.05;**

6.4.3. Mean Change from Baseline in the Number of CH Attacks During the 8-Week Period (Weeks 0 to 8) After Administration of the First Dose of the IMP

This exploratory efficacy endpoint will be analyzed using an MMRM in a manner analogous to the analysis as described in Section 6.2.3.2.

Postbaseline data will include data from analysis windows of weeks 1 to 4 and weeks 5 to 8 derived using the algorithm described in Section 6.1.3. Missing data handling methods are provided in Section 6.1.4.2. The overall results from PROC MIXED will be presented.

LSMEANS statement in PROC MIXED will be changed to:

LSMEANS TRTP / **DIFF CL ALPHA=0.05**;

6.4.4. Mean Change from Baseline in the Number of CH Attacks During the 4-Week Period (Weeks 4 to 8) After Administration of the Second Dose of the IMP

The exploratory efficacy endpoints will be analyzed based on the monthly average number of CH attacks using an MMRM method in a manner analogous to the analysis as described in Section 6.3.2.

The analysis will be done in the same PROC MIXED procedure as described in Section 6.3.2. The results from the second analysis window (weeks 5 to 8) will be subsetted from the output of the PROC MIXED and will be presented.

6.4.5. Change from Baseline in the Severity of CH Attacks (Mild, Moderate, Severe and Very Severe) During Weeks 1, 2, 3, 4, 8, and 12 After Administration of the First Dose of the IMP

Pain severity for each CH attack will be recorded in the e-diary as follows:

1 = Mild

2 = Moderate

3 = Severe

4 = Very Severe

The analysis will be based on daily average derived from the numerical rating (1, 2, 3, or 4). If there is more than 1 CH attack for a day, the worst severity of CH attacks for that day will be used for the analysis. For patients who do not have CH attacks in a day, the severity rating for that day will be set to 0.

The weekly average severity of CH attacks will be calculated as follows:

$$\frac{\sum most\ severe\ CH\ attack\ per\ day\ during\ a\ week}{Number\ of\ days\ with\ assessments\ recorded\ in\ the\ eDiary\ during\ a\ week}$$

Change from baseline in the severity of CH attacks will be analyzed based on the calculated weekly average severity rating using an MMRM in a manner analogous to the analysis as described in Section 6.2.3.2. Postbaseline data will include data from all twelve 1-week analysis windows.

Analyses will be based on data from all 12 weekly analysis windows. Results for weeks 1, 2, 3, 4, 8, or 12 will be subsetted from PROC MIXED outputs afterwards.

LSMEANS statement in PROC MIXED will be changed to:

LSMEANS TRTP*ATPTN/ **DIFF CL ALPHA=0.05;**

6.4.6. Mean Change from Baseline in the Weekly Average Number of Days with Use of any Acute Headache Medications During the 12-Week Period After Administration of the First Dose of the IMP

Use of any acute headache medication will be recorded in the e-diary. The 12-week period is the period from administration of the first dose of the IMP to the EOT/early withdrawal visit. The weekly average number of days with use of any acute headache medication (triptan, ergot, opioids, or butalbital) for CH within the 12-week period for a patient will be calculated and prorated to 7 days using formula [1B]. The missing data handling method for analysis of mean change from baseline is provided in Section 6.1.4.2.

Data will be analyzed using an ANCOVA method in a manner analogous to the primary endpoint as described in Section 6.2.2.

6.4.7. Cumulative Proportion of Responders Analysis

- total cumulative distribution of percent reduction (≥0%, ≥10%, ≥20%, ≥30%, ≥40%, ≥50%, ≥60%, ≥70%, or ≥80%) from baseline (run-in period) in the monthly average number of CH attacks during the 12-week period after the first dose of the IMP, ie, based on week 0 to 12 data

- cumulative distribution function plot of percent reduction (≥0%, ≥10%, ≥20%, ≥30%, ≥40%, ≥50%, ≥60%, ≥70%, or ≥80%) from baseline (run-in period) in the monthly average number of CH attacks during the 12-week period after the first dose of the IMP, ie, based on week 0 to 12 data

The 12-week period is the period from administration of the first dose of the IMP to the EOT/early withdrawal visit. CH attacks within the 12-week period will be derived from the e-diary data as described in Section 6.1.1 and prorated to 28. Missing data handling methods are provided in Section 6.1.4.2. The percentage of reduction from baseline will be calculated using formula [2].

A cumulative proportion of responder's analysis (CPRA) will be provided for the percent reduction from baseline in the monthly average number of CH attacks during the 12-week period. This technique is a special case of the empirical cumulative distribution function, which requires that the data be specified as the cumulative proportion of patients who have values above all specified cut-off values of interest (0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%).

Let Y_{ki} denote the percent reduction from baseline in the monthly average number of CH attacks during the 12-week period for patient i in treatment group k, where $k = 1,2,3$ and $i = 1, \ldots, n_k$, n_k is the total number of patients in treatment group k.

Let C_j denote the j^{th} cut-off value of interest, where $j = 1, \ldots, \mathrm{J}$. For the cut-off value ranges from 0% to 80% by 10, $C_1 = 0$, $C_2 = 10, \ldots, C_9 = 80$.

The cumulative proportion of responders will be calculated as follows:

$$P_{kj} = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathrm{I}\,(Y_{ki} \geq C_j),$$

where

$$\mathrm{I}\,(Y_{ki} > C_j) = \begin{cases} 1, \text{ if } Y_{ki} \geq C_j \\ 0, \text{ if } Y_{ki} < C_j \text{ or if } Y_{ki} \text{ is missing} \end{cases}$$

Summary statistics for total cumulative reduction at ≥ 0%, ≥10%, ≥20%, ≥30%, ≥40%, ≥50%, ≥60%, ≥70%, or ≥80% will be presented.

The CPRA graph will place responder levels of interest (0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%) on the x-axis and graph the associated proportion of patients calculated for that response level on the y-axis for each treatment group. A 2-dimensional graph will display data by plotting P_{kj} (y-axis) verse C_j (x-axis), with k overlaid curves representing k treatment groups.

6.4.8. Mean Change from Baseline in the Monthly Average Number of CH Attacks of at Least Severe Severity During the 12-Week Period After Administration of the First Dose of the IMP

The mean change from baseline in the monthly average number of CH attacks of at least severity will be analyzed using an ANCOVA in a manner analogous to the endpoints in Section 6.2.2.

The derivation of input data for this analysis is the same as the primary analysis except only counting CH attacks with severity of severe or very severe.

6.4.9. Changes from Baseline in Responses to Questionnaires That Measure Quality of Life, Satisfaction with Treatment, and Health Status

These will include analyses based on the quality of life, satisfaction with treatment, and health status measured by the EQ-5D questionnaire, HADS, SF-12, Impact on Partner and Family questionnaire, WPAI questionnaire, and PGIC scale.

6.4.9.1. Change from Baseline in EuroQol-5 Dimension Questionnaire

The 5-level EQ-5D questionnaire (Appendix C) is a standardized questionnaire that assesses overall state of health. The EQ-5D questionnaire consists of 2 parts. In part 1, patients will rate their health in 5 domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression using a scale of 1 to 5 as follows:

1=no problems

2=slight problems

3=moderate problems

4=severe problems

5= extreme or unable to

In part 2, patients will rate their health state on a 0 to 100 mm visual analog scale; a rating of 0 means the worst imaginable health state, and a rating of 100 means the best imaginable health state. Data will be collected in the investigational site tablet at visit 2, 3, and 5.

Part 1 data will be summarized using descriptive statistics by domain and visit. A missing category will be presented if applicable.

For part 2 data, change from baseline at visit 3 and 5 in the health status (visual analog scale) will be analyzed using an MMRM in a manner analogous to the analysis as described in Section 6.2.3.2. Postbaseline data will include data from visits 3 and 5. Results from overall and each visit will be presented.

6.4.9.2. Hospital Anxiety and Depression Scale

The HADS is a 14-item scale developed to measure anxiety (7 items) and depression (7 items). Each item is scored on a 4-point scale from 0 to 3. Total scores for depression and anxiety range from 0 to 21; a total score of 0 to 7 is normal, 8 to 10 is borderline abnormal, and 11 to 21 is abnormal. Scoring details are provided in Appendix D.

Patients will complete the HADS in the investigational site tablet at visits 2, 3, and 5. Total score for each sub-scale (depression or anxiety) will be calculated by summation of raw scores from the 7 items for each sub-scale. Categorical score level (normal, borderline abnormal, and abnormal) will be derived accordingly based on the total scores for each sub-scale.

For total score calculations of each sub-scale, missing score(s) will be handled as follows:

- If having scores for at least 4 items, the missing score will be inferred using the mean score of the remaining items for the patient. The total score for the sub-scale will be calculated as follows:

  $$\frac{\sum scores\ from\ N\ items}{N} \times 7$$

  Where N (≤ 7 and ≥ 4) is the number of non-missing items for a sub-scale.
- If missing score for more than 4 items, the total score for that sub-scale will be invalid (missing).

Change from baseline at visits 3 and 5 in total score for each sub-scale (depression and anxiety) will be analyzed using an MMRM in a manner analogous to the analysis as described in Section 6.2.3.2. Postbaseline data will include data from visits 3 and 5. Results from overall and each visit will be presented.

Raw data and total scores for each sub-scale (depression and anxiety; numerical and categorical level) will also be summarized using descriptive statistics by visit. A missing category will be presented if applicable.

For summaries with the total scores, a footnote will be added indicating 0-7 = Normal; 8-10 = Borderline abnormal; 11-21 = Abnormal.

6.4.9.3. 12-Item Short-Form Health Survey

The 12-Item Short-Form Health Survey version 2 (SF-12v2) is a generic health survey containing 12 items (questions) to measure functional health and well-being rated in 8 health domains:

- general health (1 item)
- physical functioning (2 items)
- role functioning - physical (2 items)
- role functioning - emotional (2 items)
- bodily pain (1 item)
- mental health (2 items)
- vitality (1 item)
- social functioning (1 item)

Detail descriptions of each item are provided in Appendix E. Patients will complete the SF-12 in the investigational site tablet at visits 2, 3 and 5.

Physical Composite Scores (PCS-12) and Mental Health Composite Scores (MCS-12) will be computed using scoring software (Scoring Software v5.0).

Change from baseline at visits 3 and 5 in PCS-12 and MCS-12 will be analyzed using an MMRM in a manner analogous to the analysis as described in Section 6.2.3.2. Postbaseline data will include data from visits 3 and 5. Results from overall and each visit will be presented.

Raw data will be summarized using descriptive statistics by visit.

6.4.9.4. Impact on Partner and Family Questionnaire

The patients and partners/family members of patients participating in the study, if applicable, will complete an Impact on Partner and Family questionnaire at visits 2, 3, and 5. Responses will be recorded in the investigational site tablet. The questionnaire for patients is provided in Appendix F; for partners in Appendix G.

Patients and partners/family members of patients will respond the questions based on impact of CH “during the last 12 weeks” at baseline visit and impact of CH “since last assessment” at postbaseline visits.

Raw data will be summarized using descriptive statistics.

6.4.9.5. Work Productivity and Activity Impairment Questionnaire: General Health V2.0

The generic version of the Work Productivity and Activity Impairment: General Health (WPAI: GH; Appendix H) questionnaire measures the overall effect of health on productivity at work and daily activities. The specific health problems version of the WPAI questionnaire allows investigators to attribute productivity and activity impairment issues to specific health conditions. Patients will complete the WPAI questionnaire in the investigational site tablet at visits 2, 3, and 5.

The scores for the following 4 domains will be derived based on the WPAI: GH questionnaire. Derived scores will be multiplied by 100 to express in percentages. Higher percentages indicate greater impairment.

1. Percent work item missed due to health: $\frac{Q2}{Q2+Q4}$
2. Percent impairment while working due to health: $\frac{Q5}{10}$
3. Percent overall work impairment due to health: $\frac{Q2}{Q2+Q4} + \left[\left(1 - \frac{Q2}{Q2+Q4}\right) \times \frac{Q5}{10}\right]$
4. Percent activity impairment due to health: $\frac{Q6}{10}$

Change from baseline at visits 3 and 5 in above calculated domain scores will be analyzed using an MMRM in a manner analogous to the analysis as described in Section 6.2.3.2. Postbaseline data will include data from visits 3 and 5. The results for overall and visits 3 and 5 will be presented.

Raw data will be summarized using descriptive statistics by visit.

6.4.9.6. Patient Global Impression of Change Scale

The PGIC is a validated generic tool for the assessment of overall change in the severity of illness following treatment. Patients will be asked to rate the change in their overall health and well-being since the study start (the time after patient received the first IMP dose) on a 7-point scale as the follows:

- 1 = Much worse
- 2 = Moderately worse
- 3 = Slightly worse
- 4 = Stayed the same
- 5 = A little better
- 6 = Moderately better
- 7 = Much better

Patients will record responses to the PGIC scale in the electronic diary device at home at week 1 (day 7) and in the investigational site tablet at visits 3, 4, and 5.

For analysis purpose, a dichotomous scale of Yes or No will be derived. A favorable change is score of 5-7, which means there is improvement (Yes) with the treatment. If the response is 1-4, it is considered no improvement (No).

The percentage of patients' dichotomous scale of Yes or No rated by PGIC assessments at each visit will be analyzed using a CMH method in a manner analogous to the first secondary endpoint as described in Section 6.3.1.

Raw data will be summarized using descriptive statistics for each visit. Missing category will be presented if applicable.

7. MULTIPLE COMPARISONS AND MULTIPLICITY

Hochberg's step-up method will be implemented to test primary and secondary endpoints while controlling the overall type 1 error rate at 0.05.

The treatment comparisons will be for the fremanezumab 900-mg iv loading dose group versus placebo [H_{01}] and the fremanezumab 675-mg sc loading dose group versus placebo [H_{02}]. According to Hochberg step-up method, the rules of multiple comparisons for 2 null hypotheses for the primary endpoint are as follows:

1. If the resulting 2-sided p-values from both comparisons are ≤0.05 (reject both null hypotheses), then the next comparison of interest will be interpreted inferentially at the alpha level of 0.05 and statistical significance will be claimed for both active dose groups. Testing of secondary efficacy measures will be continued.
2. If the resulting 2-sided p-value from 1 of comparisons is >0.05 (fail to reject the null hypothesis) and other resulting 2-sided p-value is ≤0.025 ($\frac{\alpha}{2}$ = 0.025; reject the null hypothesis), then the result from the comparison with p-value ≤0.025 will be interpreted inferentially at the alpha level of 0.05, and statistical significance will be claimed for this active dose group**.** No further comparisons will be interpreted inferentially.
3. If the resulting 2-sided p-value from both comparisons are >0.05 (fail to reject both null hypotheses), then no statistical significant will be claimed and no comparisons will be interpreted inferentially.

If the resulting 2-sided p-values from both comparisons for the primary endpoint are ≤0.05 (reject both hypotheses; case 1 as described above), testing of secondary efficacy measures will proceed at an error rate of 0.05 in the sequential manner as specified in Section 1.2. The same testing rules as described above will be followed. This process will continue either until all comparisons of interest for secondary endpoints are interpreted inferentially or until the point at which the resulting 2-sided p-value for a comparison of interest is >0.05. At the point where a 2-sided p-value >0.05, the testing process stops. No further comparisons will be interpreted inferentially.

For the last secondary endpoint (PPSI assessments), only the assessment at week 12 will be included in the multiple comparisons.

8. SAFETY ANALYSIS

8.1. General

The safety analysis set will be used for all safety analyses. Summaries will be presented using descriptive statistics by treatment group and all fremanezumab as actually received unless otherwise stated.

8.2. Duration of Exposure to Study Drug

At the visit 2, each patient will receive a 1-hour iv infusion and 3 sc injections of fremanezumab or placebo. At visits 3 and 4, patients will receive 1 sc injection of fremanezumab or placebo.

Duration of the treatment period (days) for a patient is defined as the number of days a patient is in the double-blind treatment period and calculated as date of EOT visit - first date of the IMP + 1. For patients who are lost to follow-up, the date of EOT visit is defined as the date of last IMP administration + 27. Duration of iv infusion (minutes) will be calculated for the infusion as stop date/time of infusion - start date/time of infusion. Dosing visit interval (days) is defined as the number of days between dosing visit (dose 1 and dose 2; dose 2 and dose 3) and calculated as date of next dosing - date of current dosing.

Duration of the treatment period (days), duration of iv infusion (minutes), dosing visit interval (days), the number of doses received, total number of injections received (placebo or fremanezumab), and reasons for IMP not administrated will be summarized using descriptive statistics. Summary will be presented by treatment group.

IMP administration and accountability data will be listed.

8.3. Adverse Events

Adverse events will be recorded from time informed consent is obtained through the end of study participation.

The following are considered protocol-defined adverse events of special interest to be sent to the sponsor's Global Patient Safety and Pharmacovigilance Department for evaluation: ophthalmic-related adverse events of at least moderate severity, events of possible drug-induced liver injury (aspartate aminotransferase or alanine aminotransferase ≥3 × the upper limit of normal [ULN], total bilirubin ≥2 × the ULN, or international normalized ratio >1.5), Hy's Law events, or events of anaphylaxis and severe hypersensitivity reactions.

All adverse events will be coded using MedDRA (version 18.1). Each patient will be counted only once in each preferred term or SOC category for the analyses of safety. Summaries will be presented for all adverse events (overall and by severity), adverse events determined by the investigator to be related to test IMP (defined as related or with missing relationship) (overall and by severity), serious adverse events, and adverse events causing withdrawal from the study. Adverse events with the missing flag indicating serious will be excluded from the summary of serious adverse events but included in the summary of non-serious adverse events.

Summaries for injection site and infusion-related adverse events and protocol-defined adverse events of special interest will be presented (overall and by severity).

Listings for deaths, serious adverse events, adverse events leading to discontinuation, injection site and infusion-related adverse events, and protocol-defined adverse events of special interest will be presented. In addition, listings for MedDRA dictionary terms for adverse event descriptions and adverse event preferred terms by patient number and treatment group will be presented.

Spontaneous abortion or an elective abortion due to developmental anomalies will be reported as a serious adverse event (protocol Section 7.2). These serious adverse events will be listed separately if applicable.

Summaries will include treatment-emergent adverse events which are defined as adverse events occurring at or after the first dose of the IMP. The listing will include all adverse events recorded.

Adverse events for patients who did not meet screening criteria will be listed.

8.4. Injection Site Assessments

Injection site assessments will be performed immediately (+10 minute) and 1 hour (±15 minutes) after receiving each dose of the IMP at visits 2, 3 and 4. The injection site(s) will be assessed for erythema, induration, and ecchymosis. More details are in Section 7.11 of the study protocol.

Injection-site reactions should be recorded as adverse events. Injection-site related adverse events will be summary as indicated in Section 8.3.

8.5. Hypersensitivity/Anaphylaxis

Patients will be assessed for suspected anaphylaxis reaction during and after administration of the IMP (through 1 hour postdose) at visits 2, 3, and 4. Data will be summarized using descriptive statistics.

The number of patients with suspected anaphylaxis reactions and number of suspected anaphylaxis reactions per patient will be summarized using descriptive statistics.

The relative time of suspected event will be calculated as date/time of suspected event - date/time of most current IMP administration and summarized using descriptive statistics. If a patient has more than one suspected anaphylaxis reactions, the earliest time will be used for the calculation.

Data will be listed.

8.6. Electronic Columbia Suicide Severity Rating Scale

The eC-SSRS Baseline/Screening version will be completed at visit 1, and the eC-SSRS Since Last Visit version will be completed at all other visits (visits 2, 3, 4, and 5), including unscheduled visits. Any positive findings on the eC-SSRS Since Last Visit version require evaluation by a physician or doctoral-level psychologist.

Data for patients with positive findings (having suicidal ideation or suicidal behavior) will be listed.

8.7. Deaths

If any patient dies during the study, all relevant information will be discussed in the patient narrative included in the clinical study report.

8.8. Clinical Laboratory Tests

Clinical laboratory tests (serum chemistry, hematology, coagulation, and urinalysis; see protocol Appendix M) will be performed at all visits (visit 1 [screening] through visit 5 [EOT]) using the central laboratory. All clinical laboratory test results outside of the reference range will be judged by the investigator as belonging to one of the following categories:

- abnormal and not clinically significant
- abnormal and clinically significant

A laboratory test result that is judged by the investigator as clinically significant will be recorded on the source documentation and the CRF as an adverse event, and monitored as described in Section 7.1.2 of the protocol.

Laboratory test results will be presented in standard international (SI) units in summaries. Laboratory values and changes from baseline to each visit and Last Assessment will be summarized using descriptive statistics. Shifts (below [low], within [normal], and above [high] the normal range) from baseline to each postbaseline visit and the Last Assessment will be summarized using patient counts. Baseline is defined as the last observed data before the administration of the first dose of the IMP (also see Section 4.2).

The potentially clinically significant abnormal values will be derived using criteria specified in Table 2 based on all postbaseline values (including scheduled, unscheduled, and withdrawal visits). The overall incidence of potentially clinically significant abnormal values will be summarized for laboratory variables using descriptive statistics by treatment group. Listings for patients who have potentially clinically significant abnormal laboratory data will be presented.

Table 2: Criteria for Potentially Clinically Significant Laboratory Values

Test		**Criterion value**
Serum chemistry		
ALT		≥3x ULN
AST		≥3x ULN
ALP		≥3x ULN
GGT		≥3x ULN
LDH		≥3x ULN
BUN		≥10.71 mmol/L
Creatinine		≥177 μmol/L
Bilirubin (total)		≥34.2 μmol/L
Hematology		
Hematocrit	Men	<0.37 L/L
	Women	<0.32 L/L
Hemoglobin	Men	≤115 g/L
	Women	≤95 g/L
WBC counts		≤3 x 10^9/L ≥20 x 10^9/L
Eosinophils		≥10%
ANC		≤1 x 10^9/L
Platelet counts		≤75 x 10^9/L ≥700 x 10^9/L
Urinalysis		
HGB		≥2 unit increase from baseline
Glucose		≥2 unit increase from baseline
Ketones		≥2 unit increase from baseline
Total protein		≥2 unit increase from baseline

ALP=alkaline phosphatase; ALT=alanine aminotransferase; ANC=absolute neutrophil count AST=aspartate aminotransferase; BUN=blood urea nitrogen; GGT=gamma- glutamyl transpeptidase; HGB=hemoglobin; LDH=lactate dehydrogenase; RBC=red blood cell; ULN=upper limit of normal range; WBC=white blood cell

Serum beta-human chorionic gonadotropin (β-HCG) tests will be performed for all women of childbearing potential at visits 1 and 5. Urine β-HCG tests will be performed for women of childbearing potential at visits 2 through 4. Pregnancy test results will be listed.

Current menstruating status (yes or no) will be collected at all visits and listed.

8.8.1. Laboratory Values Meeting Hy's Law Criteria

All occurrences of possible drug-induced liver injury that meet Hy's law criteria as defined in the Section 7.1.5.1 of the study protocol will be included in serious adverse events reporting.

8.9. Physical Examinations

Physical examinations will be performed at visits 1, 2, and 5. Any physical examination finding that is judged by the investigator as a clinically significant change (worsening) compared with a baseline value will be considered an adverse event, recorded on the CRF, and monitored as described in Section 7.1.2 of the study protocol.

Abnormal physical examination findings will be listed.

Weight and height will be summarized and listed with vital signs data.

8.10. Vital Signs

Vital signs (pulse, systolic and diastolic blood pressure, and body temperature) will be measured at each (visit 1 through 5). Weight will be measured at visits 1, 2, and 5. Any vital sign value that is judged by the investigator as a clinically significant change (worsening) from a baseline value will be considered an adverse event, recorded on the source documentation and transcribed onto the CRF, and monitored as described in Section 7.1.2 of the protocol.

Vital signs (including weight) values and changes from baseline to each visit and the Last Assessment will be summarized using descriptive statistics. The incidence of potentially clinically significant abnormal values will be summarized for selected vital signs using descriptive statistics. Baseline is defined as the last observed data before the administration of the first dose of the IMP (also see Section 4.2).

Table 3 specifies the criteria for identifying vital signs as potentially clinically significant abnormal values. Note that in order to qualify as potentially clinically significant abnormal, a value needs to meet both criteria below: ie, have a value beyond the criterion value and a change of at least the magnitude specified in the change relative to baseline column. The potentially clinically significant abnormal vital signs values will include all postbaseline values (including scheduled, unscheduled, and early withdrawal visits) for the summaries.

Table 3: Criteria for Potentially Clinically Significant Vital Signs

Vital Sign	Criterion value	Change relative to baseline
Pulse	≥120 bpm	Increase of ≥15 bpm
	≤50 bpm	Decrease of ≥15 bpm
Systolic blood pressure	≥180 mm Hg	Increase of ≥20 mm Hg
	≤90 mm Hg	Decrease of ≥20 mm Hg
Diastolic blood pressure	≥105 mm Hg	Increase of ≥15 mm Hg
	≤50 mm Hg	Decrease of ≥15 mm Hg
Temperature	≥38.3°C	Change of ≥1.1°C

bpm=beats per minute

Height will be measured at screening visit, and data will be listed in the vital sign listing.

8.11. Electrocardiography

Triplicate 12-lead ECGs will be collected at visits 1, 2, and 5. Any ECG finding that is judged by the investigator as a potentially clinically significant change (worsening) compared with a baseline value will be considered an adverse event, recorded on the source documentation and in the CRF, and monitored as described in Section 7.1.2 of the protocol.

For ECG variables, the mean of recorded results from the 3 measurements at a visit will be calculated. The mean results and mean changes from baseline to EOT and Last Assessment will be summarized using descriptive statistics. Baseline is determined based on the last set of observed data before the administration of the first dose of the IMP (also see Section 4.2).

For ECG findings, the worst value of recorded findings at a visit will be used for analysis. Baseline ECG findings and shifts (normal, abnormal not clinically significant, and abnormal clinically significant) from baseline to overall (worst value for a patient) and the Last Assessment (worst value of recorded findings from the last visit) will be summarized using patient counts.

8.12. Concomitant Medications or Therapies

Concomitant medications, treatments, or procedures will be collected up to the end of study.

All concomitant medications will be coded using the WHO Drug. The incidence of concomitant medications will be summarized using descriptive statistics by therapeutic class and preferred term. Patients are counted only once in each therapeutic class category, and only once in each preferred term category. The concomitant medications will include all medications taken after administration of the first IMP.

The subset of medications or therapies will be summarized by the indication categories as indicated in Section 5.5.

9. TOLERABILITY VARIABLES AND ANALYSIS

Tolerability was not specifically defined for this study.

10. PHARMACOKINETIC ANALYSIS

There are no prespecified pharmacokinetic endpoints.

Fremanezumab plasma concentration will be summarized using descriptive statistics at each planned sampling time point for each of the active treatment groups (samples from patients who received placebo will not be analyzed). The summary will be based on the safety analysis set. The plasma concentration will be listed by active treatments, scheduled visits and time points.

11. PHARMACOGENOMIC ANALYSIS

Pharmacogenomic analysis results will be summarized for each gene tested. An attempt will be made to correlate clinical observations (pharmacokinetics, safety, efficacy, or other effects) with the genotypes observed. Additional pharmacogenomic analysis may be conducted at a later time and will be reported in a separate addendum report.

This analysis is not included in this SAP.

12. BIOMARKER ANALYSIS

Biomarker analysis will include logistic regression, receiver operating characteristic curves, and summary statistics. Results will be reported separately. Measurements will be made using validated assays.

This analysis is not included in this SAP.

13. IMMUNOGENICITY ANALYSIS

A summary of immunogenicity results will be provided, and the incidence of immunogenicity will be calculated. The impact of immunogenicity on the pharmacokinetics profile, IMP efficacy, and clinical safety will be evaluated. This ADA impact analysis will be reported separately.

This analysis is not included in this SAP.

14. ANCILLARY STUDY ANALYSIS – WEARABLE SENSOR SUBSTUDY

A subset of patients (n=45, approximately 15 patients from each treatment group) from each pivotal study (TV48125-30056 and TV48125-30057), approximately 90 patients in total, will be asked to wear a sensor monitoring system (digital wearable device) on the wrist to track sleep patterns and activity patterns. These patients will be allowed to continue in the substudy during the long-term safety study (TV48124-30058).

Analyses for endpoints from the wearable sensor substudy will include summary statistics and multimodal the algorithms to monitor physiological activity, sleep/wake activity, and treatment responses. Results will be reported separately.

These analyses will include data from both pivotal studies and the long-term safety study and are not included in this SAP. A separate analysis plan will be provided.

15. PLANNED INTERIM ANALYSIS

An interim analysis for futility will be performed once 50% of patients (the first 150 patients) have completed 12-week assessments during the double-blind study period or have withdrawn from the study early. The futility will be assessed using conditional power (CP). The study may be terminated if the CP is less than 25%. An independent statistician from a third party will perform the analysis.

15.1. Statistical Methods

The independent statistician may consider using the normal distribution based z-test as an approximation to calculate the CP for the interim analysis. Using the z-test approximation, the CP calculation can be simplified as:

$$\mathrm{CP(n, N, \alpha)} = 1 - \Phi\left(c(z_{1-\alpha/2} - \frac{n\Delta 1 + (N-n)\Delta}{\sqrt{2n\sigma_1^2 + 2(N-n)\sigma^2}}) \right)$$

Where $c = \sqrt{\frac{n * \sigma_1^2 + (N-n) * \sigma^2}{(N-n)\sigma^2}}$

$Z_{1-\alpha/2}$ is the critical value from the standard normal distribution; N is the total number of patients planned for each treatment group and n is the number of patients in each treatment group at interim analysis. Δ_1 and σ_1 are the observed treatment difference and common SD in the interim analysis, respectively. Δ and σ are the unknown treatment difference and common SD of the future data, respectively.

Under the condition of $n=N-n$, $\Delta=\Delta_1$ and $\sigma=\sigma 1$ the CP can be determined by the effect size Δ/σ as

$$\mathrm{CP(n, N, \alpha)} = 1 - \Phi\left(\sqrt{2}(z_{1-\alpha/2} - \frac{n\Delta 1 + (N-n)\Delta}{\sqrt{2n\sigma_1^2 + 2(N-n)\sigma^2}}) \right)$$

$$\mathrm{CP(n, N, \alpha)} = 1 - \Phi\left(\sqrt{2}(z_{1-\alpha/2} - \sqrt{n}\frac{\Delta 1}{\sigma 1}) \right)$$

The critical variable for the interim analysis is the primary efficacy variable in this study: mean change from baseline in the monthly average number of CH attacks during the 12-week period after administration of the IMP. Observed treatment difference and SD and will be calculated using ANOVA.

15.2. Timing and Data

The interim analysis will be performed once 50% of patients (the first 150 patients, 50 patients per treatment group) have completed 12-week assessments or withdrawn from study early.

Data for the interim analysis futility will be 100% clean. Analysis dataset(s) for CH attacks will be provided to the independent statistician.

The Teva statistician will request to have a copy of the randomization code sent to the independent statistician prior to the interim analysis while all persons who are involved in the study will remain blinded. The production and execution of the analysis and the randomization code will be conducted and kept in a secure location by the independent statistician.

15.3. Result Dissemination

The report of the interim analysis will only include binary results (yes or no) to indicate whether or not the CP for an active treatment group is less than 25%.

15.4. Decision Rules

Decision of whether or not to continue the study will be based on CPs from the 2 comparisons (fremanezumab 900-mg iv loading dose group versus placebo or fremanezumab 675-mg sc loading dose group versus placebo).

If both comparisons satisfy the stopping criterion (CP <25%), this study may be considered to be terminated.

If the CP for 1 of the 2 comparisons is ≥25%, the study will continue without changes.

16. STATISTICAL SOFTWARE

All data listings, summaries, and statistical analyses will be generated using SAS®.

17. CHANGES TO ANALYSES SPECIFIED IN THE STUDY PROTOCOL

None.

18. REFERENCES

Scoring Software (v5.0) Copyright© 2004, 2007, 2009, 2010, 2016 QualityMetric Incorporated
24 Albion Road, Bldg 400. Lincoln, R.I. 02865, U.S.A.

APPENDIX A. E-DIARY QUESTIONNAIRE

The recording window on e-diary device is "Yesterday morning 6:00 AM to today at 5:59 AM" or "The day before yesterday morning 6:00 AM to yesterday at 5:59 AM". A maximum of 48 hours of recall time will be accepted.		
A1	Did you have any Headache?	
A2	How many episodes of headache did you have?	NUMBER
Questions to be asked for each attack		
B1	At what approximate time did the attack begin?	TIME
B2	At what approximate time did the attack end?	TIME
B3	How bad was your headache (head pain) when it was at its worst	MILD MODERATE SEVERE VERY SEVERE
B4	Was the headache restricted to one side of your head?	Yes, my headache was restricted to the left side of my head Yes, my headache was restricted to the right side of my head No, my headache happened in both sides
During any time of your headache, did you experience the following in the same side of you headache?		
B5	Red eye and/or tearing (lacrimation)?	
B6	Stuffy nose (nasal congestion) and/or runny nose?	
B7	Swelling around your eyes (eyelid or palpabrea edema)?	
B8	Sweating on your forehead or face?	
B9	Did your forehead or face become red (flushing)?	
B10	Sensation of full ear (fullness of your ear)?	
B11	Dropping of the upper eyelid (ptosis)	
B12	On the side of pain, your pupil became smaller than on the other side (miosis)?	
B13	Did you become agitated or restless (walking around the room, bending your head, etc.)?	
B14	Did you use oxygen at any time during the attack?	
B15	Did you take any medications during the attack?	GO MED CHEST
Comparing this attack with typical attacks that you experienced BEFORE THIS STUDY, would you say that:		

C1	The severity of this attack was	SAME as before study MORE SEVERE than before study LESS SEVERE than before study
C2	The duration of this attack was	SAME as before study LONGER than before study SHORTER than before study
D1	How many hours did you sleep?	NUMBER
D2	How rested/recovered did you feel when you wake up?	Totally rested Somewhat rested; not rested at all
D3	How much time did you have trouble concentrating on activities that you needed to do?	NONE OF THE TIME SOME OF THE TIME MOST OF THE TIME ALL TIME
D4	How interested (motivated, energized) were you in doing daily activities?	VERY INTERESTED SOMEWHAT INTERESTED NOT INTERESTED
	In general how was your day?	It was a good day Neutral (not good or bad) It was a bad day
D5	Which of the following situations best describe your work/school performance, when you did not have a headache? (select only one)	I worked/studied normally My working/studying ability was impaired by less than 50% My work/studying ability was impaired by 50% or more I missed work/school
D6	During your work or school day, approximately how much of the time did you feel that you were working more slowly or taking longer to complete tasks than usual or expected? (select only one)	None of the time Some of the time About half of the time or more
D7	On average, how much of the time were you very tired, asleep, or feeling drained? (select only one)	None of the time Some of the time Most of the time All of the time

D8	Which of the following situations best describe your ability to perform household chores, when you did not have a headache? (select only one)	I performed household activities normally My ability to perform household chores was impaired by less than 50% My ability to perform household chores was impaired by 50% or more I could not perform household chores Not applicable
D9	How engaged were you with your partner's or children's activities, when you didn't have a headache? (select only one)	Very engaged Somewhat engaged Not engaged Not applicable

APPENDIX B. LOGICS FOR CLUSTER HEADACHE DERIVATION

CLUSTER HEADACHE ATTACK	
OPTION 1	
A1	**YES**
B3	**SEVERE, VERY SEVERE**
B4	**RESTRICTED TO LEFT OR RIGHT**
AND	
B5-B12	**YES TO AT LEAST ONE**
OR	
B13	**YES**
OPTION 2	
A1	**YES**
B3	**MILD, MODERATE, SEVERE, VERY SEVERE**
B4	**RESTRICTED TO LEFT OR RIGHT**
AND	
B14	**YES**
OR	
B15	**YES TO USE OF ACUTE MEDICATION (TRIPTAN OR ERGOT OR OPIOIDS OR BUTALBITAL)**
AND	
B5-B12	**YES TO AT LEAST ONE**
OR	
B13	**YES**

APPENDIX C. EUROQOL-5 DIMENSION QUESTIONNAIRE (EQ-5D)

[redacted]

[redacted]

APPENDIX D. HOSPITAL ANXIETY AND DEPRESSION SCALE (HADS)

[redacted]

[redacted]

APPENDIX E. 12-ITEM SHORT-FORM HEALTH SURVEY (SF-12)

[redacted]

[redacted]

APPENDIX F. IMPACT ON PARTNER AND FAMILY QUESTIONNAIRE - PATIENTS

[redacted]

[redacted]

APPENDIX G. IMPACT ON PARTNER AND FAMILY QUESTIONNAIRE - PARTNERS

[redacted]

[redacted]

APPENDIX H. WORK PRODUCTIVITY AND ACTIVITY IMPAIRMENT QUESTIONNAIRE: GENERAL HEALTH V2.0(WPAI:GH)

[redacted]

[redacted]