Effect of *COMT* Genetic Polymorphisms on Response to Propranolol Therapy in Temporomandibular Disorder

NCT number: 02437383
Document Date: 21 February 2018

Statistical Analysis Plan

Protocol Title: Effect of *COMT* Genetic Polymorphisms on Response to Propranolol Therapy in Temporomandibular Disorder

Protocol Number: 14-067-E

Version Number and Date:	Version 3.0; 21 February 2018
Funding Sponsor:	National Institute of Dental and Craniofacial Research
Study Principal Investigator:	Inna Tchivileva, MD
SAP Prepared by:	Robert James / Gary Slade BDSc, Dip DPH, PhD
Analyses to be Conducted by:	Robert James / Gary Slade BDSc, Dip DPH, PhD

DOCUMENT VERSION CONTROL

Version Number	Date	Comments/Changes
1.0	22 May 2013	Created by the study team prior to external review of the protocol.
2.0	27 June 2017	Added NIDCR protocol number and protocol short title; updated based on protocols V4.0-11.0

Version Number	Date	Comments/Changes
3.0	21 February 2018	Minor changes involving grammar, wordsmithing, punctuation, and other editorial changes have been made throughout the document. Added and removed text to reflect updates based on protocol v12.0.

APPROVALS

Approved: _____ Date: _____

Inna Tchivileva, MD
Principal Investigator
UNC School of Dentistry

Date: _____

_____

Gary Slade, BDSc, Dip DPH, PhD
Co-Principal Investigator
UNC School of Dentistry

Date: _____

_____

Robert James, MS
Statistical Scientist
Data Coordinating Center
Rho, Inc

Date: _____

_____

Samuel Arbes, DDS, MPH, PhD
Principal Investigator
Data Coordinating Center
Rho, Inc.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AE	Adverse event
COMT	Catechol-O-methyltransferase
DCC	Data Coordinating Center
DSMB	Data and Safety Monitoring Board
ECG	Electrocardiogram
GLMM	Generalized Linear Mixed Model
HIT	Headache Impact Test
ITT	Intention-to-Treat Population
LPS	Low pain sensitive
NIDCR	National Institute of Dental and Craniofacial Research
NIH	National Institutes of Health
PP	Per-Protocol Population
PPT	Pressure pain threshold
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SF-12	Short Form 12 Health Survey
TEAE	Treatment emergent adverse event
TMD	Temporomandibular disorder
TMJ	Temporomandibular joint
Val Met	Valine allele of the *COMT* gene
V1	Study visit 1, the study of the randomization and start of study drug administration
V2	Study visit 2 at 1 week post-randomization
V3	Study visit 3 at 5 week post-randomization
V4	Study visit 4 at 9 week post-randomization

1. PURPOSE OF THE ANALYSES

The purpose of the analyses is to quantify and test associations described in the primary, secondary, and exploratory objectives of the protocol.

2. PROTOCOL SUMMARY

A total of 200 participants aged 18 to 65 years with chronic temporomandibular disorder (TMD), at three clinic sites, will be randomized in a 1:1 ratio to propranolol or placebo in a parallel-group, masked clinical trial. Participants will attend 6 clinic visits over a period of 12 to 15 weeks.

Potential participants will be prescreened by telephone or at a clinic visit. During the Screening and Baseline Visit, participants will be consented, evaluated for eligibility, and assessed for baseline characteristics. Baseline procedures will include clinical tests and examinations and the administration of questionnaires.

At the Randomization and Titration Visit, participants will be randomized to propranolol or placebo and will begin a 10-week drug treatment phase that is divided into 1 week of drug titration, 8 weeks of drug maintenance, and 1 week of drug tapering. During treatment, many of the baseline assessments will be repeated. The final study visit will occur 1 week after drug tapering ends.

The primary endpoint will be a weekly pain index derived from a daily symptom diary. Secondary endpoints will be additional ratings of clinical pain; examiner assessments of experimental pain sensitivity; participant ratings of physical function, emotional function, and global improvement; occurrence of symptoms and adverse events; and use of rescue medications. DNA analyses will be conducted to identify polymorphisms in the gene that encodes catechol-O-methyltransferase (COMT), an enzyme that metabolizes catecholamines such as epinephrine, norepinephrine, and dopamine and is associated with responses to pain.

2.1 Study Objectives

2.1.1 Primary Objective

The primary objective is to investigate the efficacy of extended-release propranolol compared to placebo in the reduction of the pain index in participants with TMD at week 9 of treatment (at Visit 4).

2.1.2 Secondary Objectives

The secondary objectives are:

- To investigate gene-by-treatment group interaction to determine whether the efficacy of extended-release propranolol in the reduction of the pain index varies according to participants' polymorphisms in the coding region of the *COMT* gene.

- To investigate the efficacy of extended-release propranolol compared with placebo using secondary endpoints: additional pain ratings; examiner assessments of experimental pain sensitivity; participant ratings of physical function, emotional function, global impression of change; occurrence of symptoms and AEs; and use of rescue medications.

2.1.3 Exploratory Objectives

The exploratory objectives are:

- To investigate whether the efficacy of extended-release propranolol in the reduction of the pain index varies according to participants' polymorphisms including, but not limited to three other genetic regions: a) the *COMT* promoter area; b) adrenergic receptor β2 *(ADRB2)* gene; and c) the adrenergic receptor β3 *(ADRB3)* gene.

- To investigate whether the efficacy of extended-release propranolol in the reduction of the pain index varies according to various phenotypic characteristics.
- To collect and archive biospecimens for future biological and statistical analyses.

2.2 Overall Study Design and Plan

This investigation is a two-arm, parallel-group, placebo-controlled, masked, Phase II clinical trial. Participants will attend 6 clinic visits over a period of 12 to 15 weeks, and the trial will take approximately 3 years to complete. The study schematic and schedule of events are provided in Sections 12.1 and 12.2.

Throughout this statistical analysis plan (SAP), V1, V2, V3, V4, and V5 refer respectively to study visit 1 (randomization and start of drug treatment), visit 2 (1 week post randomization), visit 3 (5 weeks post randomization), visit 4 (9 weeks post randomization), and visit 5 (11 weeks post randomization and 1 week after the planned termination of study drug administration).

2.3 Study Population

A total of 200 participants aged 18 to 65 years of either sex and any race/ethnicity with chronic temporomandibular disorder (TMD), at three clinic sites, will be enrolled.

2.4 Treatment Regimens

Participants will be randomized to propranolol or placebo and will begin a 10-week drug treatment phase that is divided into 1 week of drug titration, 8 weeks of drug maintenance, and 1 week of drug tapering. The titration dose and the tapering dose are 60 mg once per day by mouth whereas the maintenance dose is 60 mg twice per day by mouth.

2.5 Treatment Group Assignments or Randomization

At the Randomization and Titration Visit, participants will be randomized 1:1 to propranolol or placebo. Randomization will be stratified by site and conducted within blocks to insure equal

numbers of participants per arm. Randomization will be accomplished by a web-based randomization system developed and maintained by the Data Coordinating Center (DCC).

2.6 Sample Size Determination

This study plans to randomize 200 participants, which will provide 94% power (alpha = 0.05) to detect a difference between a 27% reduction in pain intensity in the propranolol arm and a 9% reduction in the placebo arm (primary hypothesis). Assuming haplotype distributions and within-haplotype treatment effects that are similar to those observed in the pilot study, the same sample size would provide 61% power (alpha = 0.05) to detect an interaction between haplotype and treatment group (secondary hypothesis).

Power calculations were made using the GLMPOWER procedure in SAS, which assumes a data structure of one observation per person. This is a simplification of the situation that will occur with the linear mixed model, described in Section 10.4.1, where efficacy will be evaluated using a data structure where most participants will have three observations per person. The simplified power calculations are shown here because conventional software for power calculations does not deal with the more complex situation of linear mixed models.

3. GENERAL ANALYSIS AND REPORTING CONVENTIONS

The following is a list of general analysis and reporting conventions to be applied for this study.

- Categorical variables will be summarized using counts (n) and percentages (%) and will be presented in the form n (%).
- Moment statistics including mean and standard deviation will be reported at 1 more significant digit than the precision of the data.
- Order statistics including median, min, and max will be reported to the same level of precision as the original observations. If any values are calculated to have more significant digits then the value should be rounded so that it is the same level of precision as the original data.
- Following SAS default rules, the median will be reported as the average of the two middle numbers if the dataset contains even numbers.
- Test statistics including t and z test statistics will be reported to two decimal places.
- P-values will be reported to 3 decimal places if greater than 0.001. If less than 0.001, then p-values will be reported as <0.001. P-values will be reported with a leading 0, e.g., 0.05 rather than .05.
- No preliminary rounding should be performed; rounding should only occur after analysis. To round, consider digit to right of last significant digit: if < 5 then round down, if >=5 then round up.

- All listings will be sorted in order of treatment group (e.g., active, placebo), participant, and time of assessment (e.g., visit, time, and/or event).

If departures from these general conventions are present in specific sections of this SAP, then those conventions will take precedence over these general conventions.

The items listed below will apply to all data analyses.

- For categorical variables, summary statistics will consist of the number and percentage of patients in each category. All percentages will be rounded to one decimal point. The number and percentage of patients will always be presented in the form XX (XX.X) where the percentage is in parentheses. To ensure completeness, all summaries for categorical and discrete variables will include all categories, even if none of the patients had a response in a particular category. Unless otherwise noted, for all percentages, the number of patients in the analysis population for the treatment group who have an observation will be the denominator.

- For continuous variables, summary statistics will consist of the number of patients, mean, median, standard error of the mean (SE), minimum, and maximum values. The summary statistic n will be the number of patients with non-missing values. All means and medians will be reported to one more significant digit than the values being analyzed. Standard errors will be reported to two more significant digits than the values being analyzed. The minimum and maximum will be reported to the same number of significant digits as the values being analyzed.

- For tests of hypothesis of treatment group differences, the associated P-value will be reported.

- Version 9.4 or later of the SAS® software package will be used to produce all summaries, listings, statistical analyses, and graphical presentations.

4. ANALYSIS POPULATIONS

4.1 Intention-to-Treat Population

Section 11.3.1 of the protocol defines the intention-to-treat (ITT) sample as follows:

> Consistent with the primary statistical method that will use a linear mixed model, the intention-to-treat (ITT) sample is defined as all study participants who have at least one valid measurement of the primary endpoint (at Visit 2 or later. A "valid" assessment for the primary endpoint is defined as at least 4 daily reports of pain intensity and duration collected in the 7-day period prior to the visit (Section 2.5.1 of the protocol). Treatment group allocation will be defined as the study arm into which the subject was randomized regardless of the actual study treatment received. By necessity, the ITT sample excludes participants who provide no data for the primary endpoint after randomization. Note that follow-up

assessments will be included in the ITT analysis when sufficient diary data are collected at the follow-up assessment, even if there were protocol violations. Likewise, the ITT sample will include all available follow-up assessments from any discontinued participants.

4.2 Per-Protocol Population

As stated in Section 11.3.2 of the protocol, the per-protocol (PP) sample will be a subset of the ITT sample, excluding any assessments made at or after a visit at which the participant is found to have any major protocol deviation as listed in Section 5.2 below and excluding all assessments for participants whose compliance with study medication is less than 60% over all the study visits.

4.3 Safety Population

Section 11.3.3 of the protocol defines the safety population as follows:

The safety sample will include all randomized participants who received at least one dose of study medication. Participants in the safety sample will be analyzed with the treatment group according to the medication they actually received, regardless of their randomized assignment.

All adverse events (AEs) will be recorded on a case report form regardless of study population. For safety analyses, all AEs will be provided in listings, but only treatment emergent adverse events (TEAEs) within the safety population will be reported in summary tables.

5. STUDY PATIENTS

5.1 Disposition of Patients

The screened population will consist of all consented participants and the randomized population will consist of all screened participants who are randomized. Any participant who does not meet the criteria for randomization with be discontinued from the trial. Participants who withdraw from the study voluntarily or are withdrawn by an investigator will be discontinued from the study. Study staff will complete a study disposition form in the electronic data capture system, indicating the reason for discontinuation. The disposition of all participants and the study populations to which they belong will be presented in the participant listings and summarized by treatment allocation in a disposition table.

5.2 Protocol Deviations

The protocol deviations listed below are related to the use of investigational product and concomitant therapies and will be considered major protocol deviations. These deviations will exclude a participant’s data from the PP analysis (Section 4.2).

- Participant is unable to reach the maintenance dosage of the study drug or placebo or is unable to remain on the maintenance dosage.
- Initiation of opioid medications.
- Initiation of TMJ surgical treatment (including TMJ arthrocentesis).
- Initiation of therapeutic injections for the management of pain.
- Initiation of occlusal appliance therapy.
- Initiation of any of the following non-pharmacological therapies: acupuncture, biofeedback, and/or TENS.
- Initiation after Visit 1 and usage between Visits 1 and 4 of short-acting, non-prescription analgesics (e.g., NSAIDs, aspirin, and/or acetaminophen) for pain that exceeds the definition of episodic use as described in Section 5.7.1 of the protocol.

These and other types of protocol deviations will be reported during the study. The types of protocol deviations will be summarized descriptively for both the Safety and ITT populations by treatment group. A listing of the protocol deviations of individual participants will be included in the participant listings.

6. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Demographic and other baseline characteristics will be summarized descriptively for both the ITT and the PP populations. While no formal statistical testing will be done, all demographic and baseline characteristics will be assessed for clinically significant differences between the treatment groups. P-values from these comparisons will be considered descriptive.

To further assess the overall health of the study participants prior to study treatment, any significant medical history will be included in the participant listings.

All prior and concomitant medications will be coded to preferred drug names and therapeutic drug class using the *WHO Drug Dictionary*. Use of concomitant medications will be summarized for each therapeutic drug class and each preferred drug name. Concomitant medications will include medications reported on the Concomitant Medications page of the CRF.

7. MEASUREMENTS OF TREATMENT COMPLIANCE

Participants will be asked to bring their medication containers to each visit during the treatment period. Study staff will count the number of capsules returned and will enter the number on a case report form in the electronic data capture system. Participants will also record the study medication use in their Daily Symptom Diary that will be reviewed by the study staff at Visits 2, 3, 4, and 5. Participant compliance with study medication will be defined as taking 60% or more of the study medication over all study visits. In reports, pill counts will supersede the Daily Symptom Diary in determining compliance. Treatment compliance will be summarized by treatment group and ITT population.

8. EFFICACY EVALUATION

8.1 Handling of Dropouts or Missing Data

There will be no imputation for missing data. Missing data will be accommodated in the linear mixed model.

8.2 Efficacy Variables

The efficacy variables are listed and described in Table 8-1.

Table 8-1 Efficacy Variables

Efficacy Variables	Description and Derivation
Primary	
Net Change in the Mean Pain Index	The weekly mean pain index will be computed for the week prior to randomization and for each week during the treatment phase. The weekly mean pain index will be based on at least 4 daily reports of facial pain intensity and duration collected in the 7-day period prior to the visit. When there are fewer than 4 daily reports, the weekly mean pain index will be considered missing. The net change in the mean pain index will be calculated as the weekly mean pain index at V2, V3, or V4 minus the weekly mean pain index at V1.
30% Responder Status	The binomial 30% responder status will be positive if the participant experiences a decrease of 30% from V1 in their TMD weekly mean pain index at V4.
50% Responder Status	The binomial 50% responder status will be positive if the participant experiences a decrease of 50% from V1 in their TMD weekly mean pain index at V4.
Secondary	
Net Change in the Mean Pain Intensity	The weekly mean pain intensity will be based on at least 4 daily reports of facial pain intensity collected in the 7-day period prior to the visit. When there are fewer than 4 daily reports, the weekly mean pain intensity will be considered missing. The net change in the mean pain intensity will be calculated as the weekly mean pain intensity at V2, V3, or V4 minus the weekly mean pain intensity at V1.
Net Change in the Mean Pain	The weekly mean pain duration will be based on at least 4

Efficacy Variables	**Description and Derivation**
Duration	daily reports of facial pain duration collected in the 7-day period prior to the visit. When there are fewer than 4 daily reports, the weekly mean pain duration will be considered missing. The net change in the mean pain duration will be calculated as the weekly mean pain duration at V2, V3, or V4 minus the weekly mean pain duration at V1.
Quantitative sensory tests of sensitivity to experimental pain (QST form)	For each visit when it was measured, Thermal Pain Tolerance, recorded in degrees Celsius, will be computed as the mean of 4 pain tolerance trials. For each visit when it was measured, Thermal Pain Threshold recorded in degrees Celsius, will be computed as the mean of 4 pain threshold trials.
Pressure Pain Threshold (DC/TMD exam)	For each visit when they were measured, five variables, recorded in kg, representing mean Pressure Pain Threshold at each of five anatomical locations: a) Temporalis muscle, b) Masseter muscle; c) TMJ; d) Trapezius muscle; and e) lateral epicondyle eminence (item 8 of DC/TMD Exam form). At each site, the mean response to two trials, each performed bilaterally, will be computed. For each trial, threshold is defined as the amount of pressure at which the participant first perceives the stimulus to be painful..
Facial pain ratings (SF-McGill questionnaire)	For each visit when it was administered, two subscales, each a continuous variable, will be computed from the Short Form McGill Pain Questionnaire: a) Affective component, and b) Sensory component. In addition, 4 variables assessing facial pain will be analyzed: a) Present Pain Intensity (Item II of the SF-McGill); b) Weekly Pain Intensity (Item III of the SF-McGill); c) Weekly Pain Duration (Item IV of the SF-McGill); and d) Weekly Fatigue (Item V of the SF-McGill).
Clinical measures of limitation in jaw function (DC/TMD Exam)	For each visit when they were measured, the following clinical measures will be derived. Three measurements of range of jaw motion will be analyzed, each recorded in millimeters: a) Pain-free jaw opening; b) Maximum unassisted jaw opening; and c) Maximum assisted jaw opening (Item 4 of the DC/TMD form) The count of clinical pain masticatory structure locations will be computed (range=0 to 8) representing locations

Efficacy Variables	**Description and Derivation**
	where pain is reported during the 30-day period prior to examination: temporalis muscles, masseter muscles, other masticatory muscles, and TMJs, each reported bilaterally (Item 1a of DC/TMD form). The count of examination hyperalgesia masticatory structure locations will be computed (range=0 to 8) representing places in which pain is elicited during jaw provocation during the examination: temporalis muscles, masseter muscles, other masticatory muscles, and TMJs, each reported bilaterally (Items 4, 5 and 6 of DC/TMD form). Jaw provocation includes jaw mobility maneuvers and standardized pressure palpation. The count of examination familiar pain masticatory structure locations will be computed (range=0 to 8) representing places in which pain is elicited during jaw provocation during the examination: temporalis muscles, masseter muscles, other masticatory muscles, and TMJs, each reported bilaterally, and confirmed by the subject to be familiar to pain experienced in the prior 30 days (Items 4, 5 and 6 of DC/TMD form). Jaw provocation includes jaw mobility maneuvers and standardized pressure palpation.
Vital signs	For each visit when it was measured, systolic and diastolic blood pressure (recorded in mmHg) and heart rate (recorded in beats per minute) will be computed, each as the mean of three repeated measurements taken 2 min apart and recorded as Vital Signs in RAVE.
Subjective health status (SF-12v2 questionnaire)	For each visit when it was administered, published scoring algorithms will be used to compute two continuous variables from the SF-12v2 questionnaire: a) Physical Component Subscale and b) Mental Component Subscale (SF-12v2 questionnaire).
Headache Impact (HIT-6 questionnaire)	For each visit when it was administered, published scoring algorithms will be used to compute the Total Score for headache impact, a continuous variable.
Graded chronic pain ratings (GCPS questionnaire)	For each visit when it was administered, published scoring algorithms will be used to compute two continuous measurements (Characteristic Pain Intensity and Pain Interference) and one derived ordinal score (Grade of Chronic Pain) from the GCPS questionnaire.

Efficacy Variables	Description and Derivation
Anxiety and Depression (HADS questionnaire)	For each visit when it was administered, published scoring algorithms will be used to compute two continuous measurements: Anxiety, and Depression scores.
Patient global impression of change (PGIC questionnaire)	The ordinal rating of change, measured at the subject's visits 3 and 4, will be reported as two secondary outcome variables.
Subjective measure of jaw functional limitation (JFLS questionnaire)	For each visit when it was administered, published scoring algorithms will be used to compute four continuous scores of subjective jaw function from the JFLS questionnaire: a) Chewing Limitation b) Opening Limitation; c) Verbal & Emotional Expression Limitation; and d) Combined Global Measure.
Psychological Stress (PSS questionnaire)	For each visit when it was administered, published scoring algorithms will be used to compute a continuous summary score of psychological stress from the PSS questionnaire.
Somatic symptoms (SCL-90R questionnaire)	For each visit when it was administered, published scoring algorithms will be used to compute the Somatization Scale from the SCL-90R questionnaire.
Sleep quality (PSQI questionnaire)	For each visit when it was administered, published scoring algorithms will be used to compute the global measure of sleep quality from the PSQI questionnaire.

8.3 Analysis Methods

The ITT population will be used in all analyses unless otherwise stated. The alpha level for determining significance will be set to 0.05 for all inference tests. The mixed models analyses described below automatically adjust for missing values; there will be no further imputation or adjustments for missing values. During the analysis of the primary hypothesis, treatment allocation will be masked (i.e., groups will be labeled A and B) during the statistical programming and unmasked only after the statistical programming code has been finalized.

8.3.1 Primary Efficacy Analyses

The primary statistical analysis to evaluate the primary hypothesis will use all data from the ITT analytical sample in a linear mixed model. Mixed models will be used to analyze the treatment effect at V4 (Proc Mixed, SAS, version 9.4 or greater, SAS Institute, Inc., Cary, North Carolina). The mixed model will include the net change in the mean pain index as the dependent variable. The model's independent variables will include fixed effects for site, sex, race-ethnicity, the

mean pain index at V1 (as a baseline covariate), visit (V2, V3 and V4), treatment allocation (active versus placebo), and a visit by treatment interaction. Participants will be included as a random effect in the mixed model. The mixed model will be run twice assuming either a compound symmetry or an unstructured variance structure. The mixed model achieving the lowest Akaike information criterion will be selected as the best model and used for reporting of results. The model's denominator degrees of freedom will be adjusted for any possible effects of sample imbalances due to any missing values using Kenward Roger's method in SAS. For descriptive purposes (and regardless of the P-value for efficacy at Visit 4), efficacy estimates at each visit (2, 3 and 4) will be described in terms of adjusted means and 95% confidence intervals.

If there is a treatment effect, then "responder analyses" will be performed to compare the responder status of the treatment allocation groups. Participants will have their binomial responder status classified at both the 30% and 50% response threshold levels. The odds ratios of treatment group allocation on the responder status of participants at both the 30% and 50% threshold levels will be estimated using logistic regression with covariate adjustments for site and V1 weekly mean pain index. The odds ratio for the effect of treatment allocation on responder status for both the 30% and 50% threshold levels will be reported with their 95% confidence limits and their associated P-values.

As secondary efficacy analyses, the above primary hypothesis analyses will be repeated on the PP population. The results of the ITT and PP population based analyses will be reported separately. If it appears reasonable to assume that the results of the primary hypothesis are independent of choice of the analysis population (ITT or PP), then all subsequent analyses will be performed on the ITT population only. Stratified analysis will estimate ITT treatment effects separately for the two sexes and for the main racial-ethnic groups. The analysis will be descriptive, reporting means and 95% confidence intervals. There will be no hypothesis tests of efficacy within demographic subgroups because of the expected small numbers of males and minorities.

8.3.2 Secondary Efficacy Analyses

For secondary efficacy measures that are continuous variables or counts (Table 8-1), the effect of treatment arm on the net change in the secondary efficacy measures will be analyzed using the same statistical model as in the primary efficacy analyses but with derived scores for each measure used as the dependent variables. In summary, data from all applicable visits will be analyzed with linear mixed models as described for the primary hypothesis. For most secondary efficacy measures, the data were collected at V1, V3 and V4, although the schedule differed for a few data forms as shown in Appendix A. Outcomes will be reported at each study visit using number of subjects, the mean ± standard error, and 95% confidence limits of the mean, median, min, and max. Tests for significance of treatment group differences in the change from V1 to V4 will be evaluated with linear mixed models as described for the primary hypothesis. P-value testing for the main effects of treatment allocation will be reported for overall treatment effect and for the treatment effects at each study visit.

For secondary efficacy variables measured on an ordinal scale, the non-parametric signed-rank Wilcoxon test will instead be used to test for change between visits within each treatment group,

and there will be no adjustment for covariates. Patients' Global Impression of Change is the one secondary-efficacy measure that was not measured at V0 or V1, so the test for treatment group differences in that outcome will be evaluated using the Wilcoxon rank sum.

To test the secondary hypothesis of gene-by-treatment interaction, the linear mixed model described for the primary analysis will be extended by adding three sets of predictor variables: (1) the number of *COMT* LPS haplotypes (0, 1, or 2) will be modeled as a fixed effect (both as a continuous variable and as a categorical variable); (2) two-way interactions between haplotype effect and treatment allocation and between haplotype effect and visit sequence; and (3) a three-way interaction term between the haplotype effect, treatment allocation, and visit sequence. The gene-by-treatment interaction at V4 is the interaction of interest. A contrast statement within the mixed model will test for a gene-by-treatment interaction at V4 with a P-value ≤ 0.05 considered statistically significant. Both an unstructured and compound symmetry variance structure, as well as, fitting the haplotype variable as a continuous (linear) and a categorical variable will be tested. The statistical model with the lowest Akaike information criterion will consider the best model and used for inference testing. Residual analysis will be used to confirm that there are no major violations of model assumptions. Model adjusted means will be reported with their 95% confidence intervals. For descriptive purposes, a second set of secondary analyses will be performed on the valine allele of the *COMT* gene using valine allele (Val/Met allele) counts in an identical manner as for the LPS haplotype counts described. A third set of secondary analyses will adapt the approach for analysis of the LPS haplotype, restricting the model to non-Hispanic whites. This is consistent with the approach of some genetic researchers who advocate racial-group restriction as a preferable analytic method to control for population stratification.

The participants' use of rescue medications for TMD pain as well as over-the-counter analgesics (namely, NSAIDs, acetaminophen, and aspirin) for the treatment of any type of pain during a period from V1 to V4 will be summarized by treatment group, both overall and by study visit. The number of days on over-the counter analgesics and the number of days on over-the counter analgesics as a ratio of days on study drug will be summarized (n, mean and standard error) by treatment group, both overall and by study visit. Log-transformed linear regression models will test for treatment group effects on the total number of days a participant is on rescue medications for the treatment of TMD-related pain and (in a separate analysis) for over-the-counter analgesics for the treatment of any pain. The days on rescue medications (or over-the-counter analgesics for any pain) will be log-transformed and used as the dependent variable. If there are some participants with 0 days on rescue medications (or over-the-counter analgesics), then a log(x+1) transformation of days will be used to prevent log(0) values. Study site, mean pain index at V1, and days on study (log-scale) will be used as covariates in the statistical models. Adjusted mean estimates and their 95% confidence intervals of the model adjusted days on rescue medications (or analgesics for any pain) will be reported with their treatment group effect p-values. If residual analysis suggests that model assumptions are unreasonable, then the following alternative model will be used to test for treatment group effects on the use of rescue medications (or over-the-counter analgesics for any pain). This alternative model will be a generalized linear mixed model using a negative binomial link with days on rescue medications (or over-the-counter analgesics for any pain) as the dependent variable and covariates for study site, mean pain index at V1, and for days on study.

There are approximately 20 participants that were enrolled early at UB who did not have the "type of pain" documented when taking over-the-counter analgesics. It cannot be known whether the over-the-counter analgesics were taken due to TMD pain (i.e., rescue medications) and thus these 20 participants will be excluded from the above treatment group effects on the rescue medication use analysis. These subjects will be included in the use of over-the-counter analgesics for any pain.

All of the above secondary efficacy analyses are secondary to and supportive of the primary hypothesis analysis, and thus no statistical corrections will be made for multiplicity for any of the secondary analyses. P-values ≤ 0.05 will be considered statistically significant.

Additional secondary analyses of the incidence of adverse events are described in Section 9.

8.3.3 Exploratory Efficacy Analyses

The mixed model described for the secondary hypothesis of the LPS and Val/Met haplotypes will be modified by specifying genetic variants for the *COMT* promoter, *ADRB2* gene, and *ADRB3* gene. Other genetic variants may also be included in the mixed model. The only difference in these mixed model analyses is that the *ADRB2 and ADRB3* gene types will be included as categorical fixed effects rather than the linear continuous fixed effects in the LPS and Val/Met haplotype mixed models.

Other stratum-specific estimates of treatment effect will be explored in an analogous manner, namely through the use of mixed models in which the stratification variable is a categorical fixed effect. Each stratum will be a mutually exclusive category of study participants, classified according to the following characteristics assessed either at V0 or V1:

- Body Mass Index, computed from measurements recorded at V1, will be used to create two strata, with the dichotomy based on one of two established thresholds ≥ 25 or ≥ 30. The decision as to which threshold will be made after inspecting the distribution of BMI, and select the threshold that is closest to the median.

- Smoking status, assessed using the Smoking questionnaire completed at V0, will be used to create two strata: current tobacco smokers or not (the latter therefore including former smokers and never smokers). Smokeless tobacco and other sources of nicotine will not be used for this classification.

- Migraine headache will be classified using responses to the headache questionnaire (HAQ) administered at V1. Two strata will be created, classified according to presence or absence of one or more of four types of migraine sub-types (Migraine without aura, Migraine with aura, Chronic migraine, or Probable migraine without aura).

- Degree of examiner-assessed tenderness to body palpation, recorded at Item 4 of the Fibromyalgia Questionnaire at V0, will be used to create two strata, based as closely as possible on a median split of the total number of tender points.

- Published scoring algorithms will be used to compute the catastrophizing subscale of the revised Coping Strategies Questionnaire (CSQ-R) measured at V1. The two strata will be based as closely as possible on a median split of the subscale score.

- Self-rated general health at V1, reported at item 1 of the SF-12v2 questionnaire, will be used to create two strata based as closely as possible on a median split of the ordinal rating (high, low).

- Anxiety and depression as measured using the two subscales of the Hospital Anxiety and Depression Scale (HADS) administered at V1. Two strata will be created based as closely as possible on a median split of each of the subscale scores (high, low).

- Sleep quality, as measured using the Global Score of the Pittsburgh Sleep Quality Index (PSQI) administered at V1. Two strata will created based as closely as possible on a median split of Global Score (high, low).

- Subjects' rating of confidence in treatment outcome reported at V1 will be used to create two strata based as closely as possible on a median split of the ordinal rating (high, low).

Exploratory analyses of the incidence of adverse events are described in Section 9.

9. SAFETY EVALUATION

9.1 Overview of Safety Analysis Methods

All safety analyses will be carried out using the Safety population. The following assessments will be used to evaluate the safety of propranolol for participants with TMD:

- Adverse events (AEs)
- Vital signs
- Unanticipated Problems

9.2 Adverse Events

All AEs will be coded to a system organ class (SOC) and preferred term (PT) using the *Medical Dictionary for Regulatory Activities* (MedDRA) version 18.0. All tabular AE summaries will be for treatment-emergent AEs (TEAEs). TEAEs will be defined as those AEs with an onset on or after the date of first dose of study drug (study visit 1: V1). If an AE was recorded prior to the first dose of study drug and there was an increase in its severity, the AE will be considered a TEAE. All other AEs will be classified as non-treatment-emergent. Non-treatment-emergent AEs will be flagged in all AE data listings.

An overall summary table will be developed to report the number of events and the incidence of participants having at least one event in the following categories:

- TEAEs

- TEAEs indicated as serious (SAEs)
- TEAEs that lead to study drug discontinuation
- TEAEs with an outcome of death
- TEAEs that were reported as having a definite, probable, or possible relation to study drug
- TEAEs reported as having a severity rating of severe

In addition, a summary table of TEAEs classified by system organ class (SOC) and preferred term (PT) will be provided for each of the following:

- TEAEs
- TEAEs by maximum severity
- TEAEs by relationship to study drug
- TEAEs by week of treatment

The summary of all TEAEs will present both the number of TEAEs and the incidence of TEAEs by severity, and relationship to study drug. When reporting the number of TEAEs, if the same TEAE occurs for a patient on multiple occasions, the event will be counted once for each occurrence. When reporting the incidence, a patient will only be counted once if they ever experience an event within the SOC and ever experience the individual PT.

If for any TEAE (system organ class or preferred term) the incidence rate is 5% or more in either treatment arm then a logistic regression will be performed and descriptive p-values of treatment arm differences will be reported.

For the summary of TEAEs by severity, if the severity of the TEAE is not reported, then the severity of the AE will be counted as severe. If the same TEAE occurs for a patient on multiple occasions, the TEAE will be categorized according to the highest severity rating for that TEAE in that patient. For the summary of TEAEs by relationship to study drug, if the relationship is missing, it will be counted as definite/certain. If the same TEAE occurs for a patient on multiple occasions, the TEAE will be categorized according to the closest relationship to study drug reported for that TEAE in that patient.

9.3 Deaths, Serious Adverse Events, and Other Significant Adverse Events

In addition to these summary tables, the incidence of adverse events leading to death, the incidence of SAEs, and incidence of TEAEs that led to study drug being discontinued will be presented in summary tables and listings. The summary listings will provide all of the information reported for that AE and will include the number of days from date of first dose to date of the occurrence of the AE.

9.4 Vital Signs and Other Observations Related to Safety

Listings will include all participant vital signs and any other comments as provided by the examining clinician. Vital signs statistics will be summarized in table form by treatment group and study visit. These tables will include summary statistics appropriate for variable type as listed in Section 3.

9.5 Unanticipated Problems

Unanticipated problems will be included in the listings.

10. INTERIM ANALYSES AND DATA MONITORING

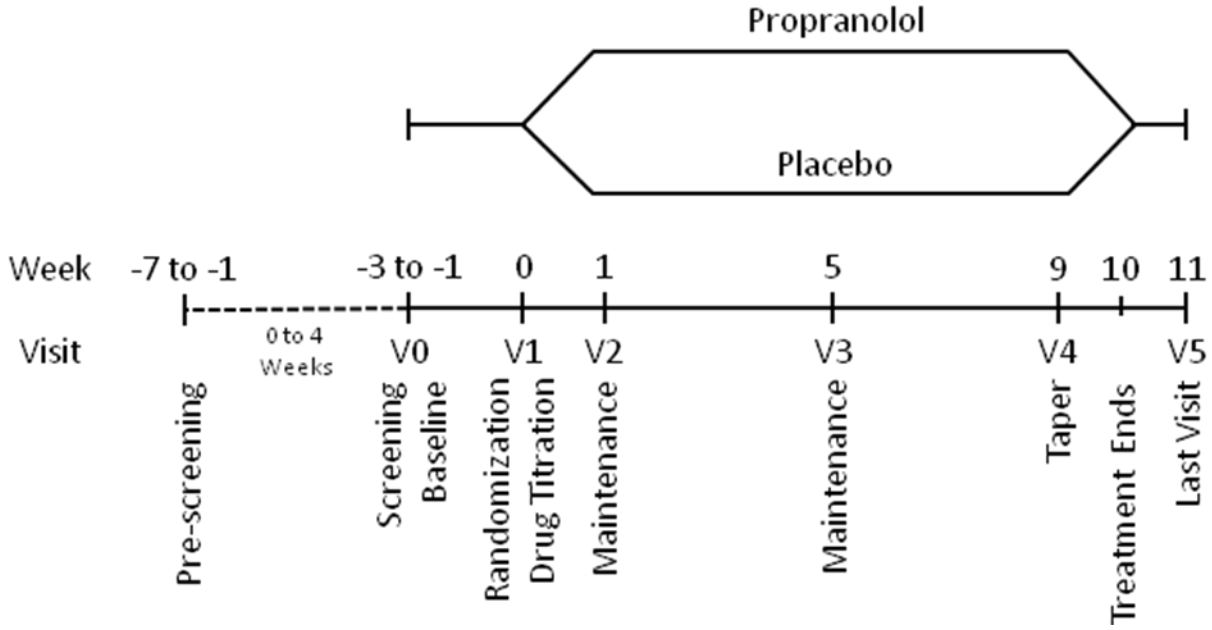
No interim analyses have been planned.

11. CHANGES TO THE ANALYSES PLANNED IN THE PROTOCOL

No changes to the analyses planned in the protocol have been made.

12. APPENDICES

12.1 Study Schematic



12.2 Schedule of Events

Study Phase:	Pre-screening[a]	Screening and Baseline	Randomization and Treatment (1 week titration, 8 weeks maintenance, and 1 week taper)				Follow-Up
Clinic Visit:		V0	V1[b]	V2	V3	V4	V5
Study Day:	≤ 28 days prior to V0	7 to 21 days prior to V1	0	7 (+ 3)	35 (± 7)	63 (+ 6)	77 (+ 7)
Procedure							
Contact Information	x						
Prescreening Interview Script	x						
Informed Consent[c]		x					
Eligibility Review	x	x	x				
AUDIT Questionnaire		x					
Smoking Questionnaire		x					x
Alcohol Consumption per week			x	x	x	x	x
Demographic Information		x					
Medical History and Review	x	x	x	x	x	x	x
Concomitant Medications		x	x	x	x	x	x
Concomitant Therapies		x	x	x	x	x	x
Dispense New Daily Symptom Diaries		x	x	x	x	x	
Adverse Event Review			x	x	x	x	x
Collect Daily Symptom Diaries			x	x	x	x	x
Assess Compliance with Daily Symptom Diaries			x	x	x	x	x
Collect Confidence in Treatment Outcome			x				
Randomization			x				
Dispense Study Drug			x[d]	x	x	x	
Collect Study Drug Container(s)				x	x	x	x
Assess Compliance with Study Drug				x	x	x	x
Clinical Examinations and Tests							
Physical Measurements		x					

Study Phase:	**Pre-screening[a]**	**Screening and Baseline**	**Randomization and Treatment (1 week titration, 8 weeks maintenance, and 1 week taper)**				**Follow-Up**
Clinic Visit:		**V0**	**V1[b]**	**V2**	**V3**	**V4**	**V5**
Study Day:	**≤ 28 days prior to V0**	**7 to 21 days prior to V1**	**0**	**7 (+ 3)**	**35 (± 7)**	**63 (+ 6)**	**77 (+ 7)**
Vital Signs		x	x	x	x	x	x
Urine Pregnancy Test (females)		x	x	x	x	x	x
Manual Tender Point Exam		x					
TMD Examination		x			x	x	
Heat and Pressure Pain Tests		x			x	x	
12-Lead ECG		x					
Blood Draw			x			x	
Schedule Next Visit	x	x	x	x	x	x	
Outcome Measure Questionnaires							
Symptom Inventory		x	x	x	x	x	x
Fibromyalgia Questionnaire		x					
Headache Questionnaire[e]			x				
SF-McGill Pain Questionnaire		x		x	x	x	x
Graded Chronic Pain Scale[e]			x		x	x	
SF-12 Health Survey v2[e]			x		x	x	
Jaw Functional Limitation Scale[e]			x		x	x	
Perceived Stress Scale[e]			x		x	x	
Hospital Anxiety and Depression Scale[e]			x		x	x	
Pittsburgh Sleep Quality Index[e]			x		x	x	
Headache Impact Test (HIT-6) [e]			x		x	x	
SCL-90R Somatization Scale[e]			x		x	x	
Coping Strategies Questionnaire-Revised[e]			x				
Patient Global Impression of Change[e]					x	x	

AUDIT = Alcohol Use Disorders Identification Test; ECG = electrocardiogram; SCL-90 = Symptom Checklist -90R; SF = Short Form; V = Visit

[a] May occur by phone or at a clinic visit, and may be combined with the Screening and Baseline Visit (Visit 0).

[b] If Visit 1 cannot occur within 3 weeks of Visit 0 and the participant is to remain in the study, then Visit 0 will be repeated as an unscheduled visit. The timing of Visits 3-6 and their windows are established from the date of Visit 1.

[c] Includes consent for study participation, consent to store biological specimens for future studies, and Health Insurance Portability and Accountability Act statement, if applicable.

[d] The first dose of study drug should be taken in the evening of the Day 1.

[e] Questionnaires will be distributed at Visit 0 and the participant will return completed questionnaires at Visit 1.