Dronabinol and Vaporized Cannabis in Chronic Low Back Pain

Statistical analysis plan

Principal investigators:
Thomas D. Marcotte, Ph.D.
Barth Wilsey, M.D. (retired)

Protocol Statisticians:
Bin Tang, Ph.D.
Florin Vaida, Ph.D.

April 24, 2020

1. Summary of Study Design

This study is a prospective, randomized, placebo controlled, double blind, phase 2/3 clinical trial of dronabinol and vaporized cannabis for the treatment of chronic low back pain. A total of 120 (n = 120) subjects with chronic low back pain, aged between 19-70 years will be randomized in a ratio of 1:1:1 to the three treatment arms (40 participants per arm): 1) vaporized 3.7% Δ9-THC /5.6% cannabidiol plus placebo oral pills, 2) oral Δ9-THC (dronabinol) plus vaporized placebo, and 3) double placebo. The three arms are shortly called 1) vaporized cannabis; 2) dronabinol; 3) placebo. After randomization, each patient will receive study medication at home for 8 weeks after which they will be tapered off of investigational treatments over two weeks. The three study periods include: i) Titration (weeks 0-4); ii) Treatment (weeks 5-8); iii) Tapering (Weeks 9-10). During the titration period, patients will undergo titration of vaporized medication from 4 puffs to 18 puffs, or oral medication from 5 mg qd to 10 mg tid per day. Subjects will be dropped from the study by the end of Week 4 if they are unable to tolerate the minimum amounts of medication described above, and replaced by another volunteer. During the treatment period, subjects will be expected to consume the amount of study medication that they titrated themselves up to by the end of week 4. During the tapering period, patients will slowly withdraw from medication over two weeks. During treatment at home (weeks 1-8), patients will maintain a paper daily diary to record their intake of study medications, and breakthrough pain medications, as well as pain relief measurements. Outcome measure (such as self-report forms and neuropsychological testing) will be collected during the six study visits. If participating in Driving Simulation, subjects with driver licenses will be scheduled for two experimental visits.

The variables being assessed include vital signs, pain evaluation (cold pressor test), driving simulation, iPad performance testing, blood draws, neuropsychological tests, self-reports (BDI-II, Neuropathic Pain Scale, side effects). The primary outcome will be measured using self-reported average numerical pain intensity during the past 24 hours. A secondary outcome measure will be the level of use of breakthrough pain medication, mood, attention, verbal learning, fine motor coordination, psychomimetic side-effects, and driving performances.

The three main objectives of this study are to 1) assess whether treatment with vaporized whole plant cannabis or oral Δ9-THC reduces spontaneous and evoked pain more than placebo, and whether there are differences between the two active treatments. 2) examine the effects of vaporized whole plant cannabis and oral Δ9-THC (dronabinol) on mood, neuropsychological function, and psychomimetic side-effects (high, stoned, etc.) compared to placebo and to each other. 3) examine the acute effects (after receiving stable treatment for 4 weeks) of vaporized whole

plant cannabis and oral Δ9-THC compared to placebo and each other on driving skills.

2. Baseline characteristics, treatment, and study enrollment summaries

2.1 Demographic and medical and laboratory characteristics

These measures will be tabulated and summarized by treatment arm using Mean (SD), N (%), and Median (IQR), as appropriate. Fisher's exact test for categorical variables and analysis of variance (ANOVA) for numeric variables will be used to test the overall differences between the three treatment arms.

2.2 Study enrollment, Randomization, and dropout

The 120 eligible participants will be randomized to Arm A (vaporized Δ9-THC (3.5%)/placebo oral pill, n=40), to Arm B (dronabinol/vaporized placebo THC, n=40), or Arm C (double placebo, n=40). With a 16% dropout rate, we expect 33 subjects completing the 8-week visit for each arm. The completers who have a valid driver's license will implement driving simulation. The number (%) of subjects randomized, number (%) of subjects completing the 8-week-treatment, and number (%) of patients lost to follow-up will be summarized overall and by treatment arm. Fisher's exact test will be used to compare the arms in the proportions of lost to follow-up. Reasons (if known) for lost to follow-up will also be recorded. The comparisons are following an order of importance: arm A vs. arm B; arm A vs. arm C; arm B vs. arm C. Since the three comparisons have different levels of importance, no correction will be done for multiple comparisons.

The dropouts and completers will be compared, overall and stratified by treatment arm, with respect to their baseline demographics, medical characteristics, and the prevalence and timing of adverse events. The statistical analysis will use Fisher's exact test for binary and nominal variables, Student's t test or Wilcoxon rank-sum test for continuous variables, and log-rank test for time-to-event analysis, censored at the time of dropout.

2.3 Treatment Adherence and Discontinuation

The number (%) of subjects who discontinue assigned study treatment permanently, dates and reasons for treatment discontinuation will be summarized overall and by treatment arm, and will be compared between arms using Fisher's exact test with the strategy of the comparison described in 2.2.

3. Safety and treatment toxicity

3.1 Adverse events

The number (%) of subjects with adverse events will be tabulated and summarized overall and by treatment arm, with the grade of the adverse event (1-4) included. The study week and reasons for adverse events will be included. Using the Kruskal-Wallis rank sum test, the number of adverse events per individual and maximal grade for each individual were compared among the three treatment arms. The proportion of participants with severe adverse events (leading to permanent treatment determination) will be compared between treatment arms with Fisher's exact test using the strategy of the comparison described in 2.2.

4. Endpoints

4.1 Primary endpoint

The primary outcome will be self-reported daily average numerical pain intensity during the past 24 hours. The comparisons are following as following: 1) primary comparison: arm A vs. arm B; 2) secondary comparison: arm A vs. arm C; exploratory comparison: arm B vs. arm C. Since the three comparisons have different levels of importance, no correction will be done for multiple comparisons.

4.1.1 Primary analysis

The primary analysis will be the comparison of self-reported pain intensity between arm A and arm B. An intent-to-treat analysis will be performed on data from participants who complete baseline. Data transformation (e.g., log10, square root) may be performed to improve distribution normality as appropriate. Response profiles analysis will be performed if the trend of pain intensity over time is not linear. We will compare different covariance pattern models (e.g., unstructured, compound symmetry) based on AIC and identify a covariance model with the best fit to our data. This is a randomized controlled trial so that we can remove main group effect with assumptions about equal mean responses (i.e., pain scores) at baseline for the three treatment arms. Weekly or biweekly average of pain intensity from the daily diary will be calculated as a primary outcome. Time will be treated as a factor with 6 levels (i.e., baseline, weeks 1, 3, 5, 7, 8), and the mean response profiles will be compared between the arms. The treatment effect over time will be assessed by the significance of the treatment arm x time interaction using Wald test. Also, average change at the treatment period (weeks 5-8) from baseline will be compared between the groups. (Aim 1)

4.1.2 Secondary analysis of primary endpoint

The secondary and exploratory comparisons of pain intensity will be performed for arm A vs. arm C, and arm B vs. arm C respectively. The same statistical analyses will be performed as described above in the primary analysis. (Aim 1)

4.2 Secondary endpoints

The secondary outcomes include the level of use of breakthrough pain medication, pain tolerance and sensitivity, chronic low back pain (CLBP) impact score, neuropathic pain scale (NPS, a sum of 10 pain descriptors), profile of mood states (total mood disturbance scores), Beck Depression Inventory II, ability to learn and recall measured using Hopkins Verbal Learning Test (HVLT), measures of concentration, psychomotor speed, and graphomotor abilities by Wechsler Adult Intelligence Scale -III (WAIS-III) digit symbol test, marijuana subscale (M-scale), motor coordination and speed evaluated by grooved pegboard test. In addition, subjective and psychoactive side effects (e.g., good drug effect, drug high) and measures (e.g., lane tracking, car following) in driving simulation will be examined as secondary outcomes in a substudy, where completers with a valid driver's license will receive acute administration of treatment at the driving simulation visit.

All comparisons of secondary outcomes between arms will be with an order of importance as described above for the primary endpoint: 1) primary comparison: arm A vs. arm B; 2) secondary comparison: arm A vs. arm C; exploratory comparison: arm B vs. arm C. No correction will be done for multiple comparisons.

4.2.1 Comparison of outcomes measured at 6 visits between treatment arms

The similar analysis strategies as described above for primary endpoint will be used for the secondary outcomes measured at 6 visits. Briefly, the interaction effect of treatment arm and time (6 visits) on outcome will be tested using a covariance pattern model (e.g., unstructured model). Average change at the treatment period (weeks 5-8) from baseline will be compared between the groups. Non-significant interaction terms will be removed from the models (Aim 2).

4.2.2 Comparison of cannabis withdrawal scale between treatment arms

Cannabis withdrawal intensity and negative impact of withdrawal will be assessed via telephone calls at week 8 and week 10. Linear regression will be used to compare changes in intensity and negative

impact of withdrawal symptoms from week 8 to week 10 between the arms.

4.2.3 Comparison of subjective and psychoactive side effects between treatment arms

Psychoactive effects will be assessed at 1, 2, 3, 4, 5, and 6 hours after the first acute administration of treatment at the driving simulation visit in a substudy. Area under the time concentration curve (AUC) will be used as a measure for the aggregate effect of response over a period of time (during 1-6 hours after the 1st dose). AUC will be compared between the arms using linear regression (Aim 2).

4.2.4 Comparison of driving performance between treatment arms

The driving simulation assessments will be conducted in the substudy. (1) Driving measurements (standard deviation of lateral deviation (SDLP) for lane tracking and coherence for car following) following 8 weeks of study medication administration will be compared between treatment arms using ANOVA, to determine whether driving skills will be different after stable cannabis treatment. Acute cannabis effects will be then tested using a covariance pattern model (e.g., unstructured model) including fixed effects of treatment arm, time (minutes), and their interaction. (2) We will also include terms for pain (separately using intensity and relief), its interaction with treatment, and time too, the latter to test for recovery of the cannabis effect back to the Hour 1 baseline level. (3) The correlation (Pearson or Spearman, as appropriate) between blood Δ9-THC and each of SDLP and coherence will be estimated at each hour using a 95% confidence interval applied to the Fisher Z-transformation, then back-transformed. Secondary analyses will use independent t-tests or Wilcoxon rank sum tests to compare differences between the means of each of coherence and SDLP for high and low blood Δ9-THC levels split at 7 ng/ml. (Aim 3)

4.3 Exploratory analysis:

4.3.1 We will assess the association of daily pain intensity and the dose of treatment (mean daily number of puffs or pills) using a mixed-effects model including fixed effects of dose, treatment regimen, time (i.e., day or week), and dose x treatment regimen interaction, and subject-specific random effect. A significant interaction indicates the association of the dose and pain intensity differs in the three treatment arms. Use of breakthrough pain medication will be then included in the model to remove its confounding effect if it's present. A Non-significant interaction will be removed from the models.

4.3.2 Cumulative distributions of maximal pain reduction from baseline among treatment groups will be compared using Kolmogorov-Smirnov tests. This test (nonparametric) will indicate if a given treatment is likely to produce more overall pain reduction than another treatment regimen and takes into account an expected delay in effects of dronabinol compared to whole plant cannabis.

4.3.3 To test if a treatment effect on pain may be linked to mood and neuropsychological side effects of treatment, a mechanistic analysis using the mediation model methods of Preacher and Hayes {Preacher KJ and Hayes AF: SPSS and SAS procedures for estimating indirect effects in simple mediation models, 2004, *Behavior Research Methods,* 36:717-731} will be used, allowing for multiple mediators to be tested simultaneously.

5. Sample Size and Accrual

5.1 Power Estimation

Power estimates for each outcome assumed a 5% significance level. Totally, N=120 subjects will be randomized in 1:1:1 ratio to the three treatment arms, that is, 40 subjects to each arm. Using the same 16% drop-out rate experienced in our previous study and that 87% of the completers possess a valid driver's license, a national average published by the Department of Transportation, power was calculated for 33 subjects per treatment arm (total N=99) for Aims 1 and 2, and 29 subjects per treatment arm (total N=87) for Aim 3. We have 80% power to detect effect sizes equivalent to Cohen's d=0.54 and d=0.58 for sample sizes 33 and 29 respectively. This effect size is based on differences between groups of changes from baseline at the end of the study (week 8). The power is calculated using specific formulas for a time-by-treatment interaction in a longitudinal linear mixed-effects model {Diggle PJ, Heagarty P, Liang KY, Zeger SL: Analysis of Longitudinal Data, 2nd edition, 2002, Oxford University Press, Oxford, p.29}. Power is expected to be higher for larger effect sizes and fewer dropouts.

6. Data and Safety Monitoring Committee (DSMC)

Progress reports will be presented to the DSMC every 6 months. Two reports (open and closed reports) will be generated and presented in two sessions.

Key study personnel and all DSMC members will attend the open session; *the open report* will present data in aggregate form (not by arm). Open session attendees will review presented summarized data for data completeness and accuracy (e.g., review ranges for numeric variables to catch any outliers/unusual observations).

The closed session will include only the study statistician and members of the DSMC; *the closed report* will present data by treatment arm, but in blinded fashion,

i.e., arms will be presented as arm X, arm Y, arm Z. Treatment labels can be revealed to the DSMC members upon their request if major concerns arise. Statistical comparison of relevant characteristics between the arms will be done. No interim analysis comparing primary outcomes between the arms will be performed in order to preserve the power of the primary analysis and the overall Type I error at 0.05 level till the end of the study, unless requested by the DSMC. Members of the DSMC will be asked to maintain confidentiality related to the interim data presented in the closed report until the end of the trial.

Number of available observations (N) will be given for each variable to give a sense of data collection completeness. The reasons of data entry lagging, leading to a large portion of missing values, will be provided.

6.1 The information included in open report

The data will be presented in aggregate form, but not by arm. The measures will be tabulated and summarized using Mean (SD), N (%), and Median (IQR), as appropriate.

A. *Outline of the study design*: 1) aims, 2) design summary, 3) database freezing date, 4) confidentiality statement

B. *Enrollment and study status*: the number of 1) screened subjects; 2) exclusion and the reasons; 3) enrolled subjects and projected enrollment (graph of enrollment by month); 4) subjects completing the entire study; 5) withdrawal from the study and the reasons; 6) missing visit(s) and the reasons.

C. *Demographic characteristics of the cohort*: 1) age, 2) sex, 3) education, 4) Race

D. *Baseline study relevant characteristics of the cohort*: 1) baseline of pain intensity; 2) baseline levels of the driving performance

E. *Baseline lab characteristics*: vital signs, e.g. heart rate and blood pressure

F. *Safety data and study-related adverse events*: 1) type and severity (grade) of adverse events; 2) number of adverse events; 3) number of subjects with adverse events; 4) number of deaths related to adverse events; 5) unanticipated problems.

G. Protocol deviations

6.2 The information included in closed report

The data will be presented by blinded arms (arm X, arm Y, and arm Z), except for a notice of aggregate. The measures will be tabulated and summarized in aggregate form or by treatment arm using Mean (SD), N (%), and Median (IQR), as appropriate. The items with an asterisk (*) will be compared between arms. Kruskal-Wallis test and Fisher's Exact test will be used to compare numeric and categorical variables, respectively.

A. *Outline of the study design*: 1) aims, 2) design summary, 3) database freezing date, 4) confidentiality statement

B. *Enrollment and study status*: the number of 1) screened subjects (aggregate); 2) exclusion and the reasons (aggregate); 3) enrolled subjects and projected enrollment (graph of enrollment by month) (aggregate); 4) subjects completing the entire study; 5) *withdrawal from the study and the reasons; 6) missing visit(s) and the reasons.

C. *Demographic characteristics of the cohort*: 1) age, 2) sex, 3) education, 4) Race

D. *Baseline study relevant characteristics of the cohort**: 1) baseline of pain intensity; 2) baseline levels of the driving performance

E. *Baseline lab characteristics*: vital signs, e.g. heart rate and blood pressure

F. *Safety data and study-related adverse events*: 1) type and severity (grade) of adverse events*; 2) number of adverse events; 3) number of subjects with adverse events*; 4) number of deaths related to adverse events*; 5) unanticipated problems.

G. Protocol deviations

H. Summary of the overall study status (closed report only)