Study: # 7343 entitled EFFECT OF CLONAZEPAM ON CANNABIS WITHDRAWAL AND RELAPSE IN TREATMENT-SEEKING PATIENTS: COMBINED INPATIENT/OUTPATIENT STUDY

PI: John Mariani, M.D.

NCT# NCT02913924

Statistical Analysis Plan (Grant): 9/19/2013

Amended 4/15/2016 (DSMP)

3. Data Analysis a. Outcome Measures and Covariates

i. Primary outcome measures: 1) Cannabis use: The time to sustained relapse defined as 5 days of cannabis use over a 7 day period as recorded by the Timeline Followback method and confirmed by creatinine-normalized quantitative urine THC levels, or time to dropout of study, whichever comes first (survival outcome).

ii. Secondary outcome measures: These measures are designed to capture changes in cannabis consumption patterns and other symptoms not measured by the primary outcome measures.

1) Abstinence during the final two weeks (study weeks 11 and 12) of the study (dichotomous)

2) Retention in treatment (survival outcome)

3) Days per week of cannabis use (continuous longitudinal)

4) Cannabis withdrawal: measured by weekly Marijuana Withdrawal Checklist (continuous longitudinal)
5) Cannabis craving: measured by weekly Marijuana Craving Questionnaire (continuous longitudinal)

6) Sleep disturbance: measured by the Medical Outcomes Study—Sleep Scale (continuous longitudinal)

iii. Covariates: 1) Demographic characteristics (e.g., gender, age) 2) Baseline severity of cannabis use (measured by the quantity of cannabis per using day at baseline)

iv. Other measures: Adverse effects, as measured by the Systematic Assessment for Treatment and Emergent Events (SAFTEE), will be assessed, including potential effects on compliance and outcome.

b. Sample Size and Randomization: The Research Core will supervise randomization procedures. A total of 72 patients will be recruited over a 2.5-year period and randomized to the double-blind treatment trial, with 36 patients randomized to each of the two treatment groups, nabilone/ XR-NTX arm and placebo arm. The randomization sequence will be balanced in blocks of random size (2, 4, 6) to prevent clinicians from guessing what the next patient's treatment might be.

c. Intent to Treat / Dropouts and Missing Data: The primary analyses in this study will be on the intent-to treat (ITT) sample, i.e. on all randomized patients. If patients drop out of the study, they will be regarded as relapsed in the primary outcome analysis. This seems a safe assumption as it seems most likely dropouts will relapse. Cannabis consumption and cannabis abstinence, are assessed repeatedly over time and we will try to collect them at all study assessment points. We will account for unobserved data in secondary outcome variables using longitudinal mixed effects models (MEM) (Brown and Prescott 1999, Diggle et al. 2002) using PROC GLIMMIX in SAS®. MEMs do not require complete measurements data to estimate the outcome variable. The inferences from analyses with missing data are valid provided that they are "missing at random" (Little & Rubin, 2002). 'Missing at random' (i.e. the missing mechanism does not depend on the value of the unobserved outcome) is un-testable in most

medical research and in our study as well. One can assume either parametric or semi-parametric models for the missingness that does depend on the unobserved outcome value and do the analysis (Diggle and Kenward 1994, Kenward 1998, Rotnitzky et al. 1998, Liu et al. 1999, Scharfstein et al. 1999). Comparison of the inferences from assuming various models for the missingness provides a measure of the validity of the efficacy estimate from the model that assumes missing 'at random'. One can also compute a local sensitivity index which measures the change in the estimated treatment effect in a neighborhood of the 'missing at random' model for missingness (Rotnitzky et al. 2001). We plan to perform a sensitivity analysis based on these two approaches to assess the effect of the assumption of missing 'at random' on the inference.

d. Significance Testing and Preliminary Analyses: All tests for main effects will be performed at two-tailed significance α=5%. Before performing specific analyses (described below), we will examine all variables for outliers. The distributions of all continuous variables will be checked for normality, and transformations will be employed, if necessary, before applying specific parametric techniques. The distribution of demographic variables (ethnicity, gender, age) and other covariate measures of at baseline in the treatment arms will be examined and described in terms of means, standard deviations, proportions and 95% confidence intervals. The covariates (specified in Section 3.a.iii) may be associated with treatment outcome. For this reason, we will adjust for these covariates in all models used to test the study hypotheses. These covariates will be included in all models as main effects regardless of their statistical significance or whether they differ between treatment groups. In the secondary analyses, we will adjust also for baseline value of the outcome variable where appropriate. This adjustment will be based on the inclusion of main effects for the baseline.

e. Hypotheses Testing i. Primary hypotheses:

Hypothesis 1: Nabilone and XR-NTX will significantly increase time to relapse as compared to placebo. The following Cox Proportional Hazards model will be used: (1) $\log\{hi(t)\} = \log\{h0(t)\} + \beta 1Ij + \beta 2Uij$ where hij(t) is the hazard for individual i in the treatment group j at time t; h0(t) is the baseline hazard function; Uij is the vector of covariates; and Ij is the indicator variable for treatment with nabilone and XR-NTX. Significant Iij indicates that the hazard rate of each treatment group is different. ii. Secondary hypotheses:

Hypothesis 2: Nabilone and XR-NTX will significantly promote abstinence from cannabis use as compared to placebo during the final 2 weeks of study (weeks 11 and 12). The following logistic regression model will be used: (2) $\text{logit}(Yij = 1) = \beta 0 + \beta 1Ij + \beta 2Uij + \varepsilon ij$ where Yij is a dichotomous measure indicating whether the i th subject in the treatment group j has used cannabis during the final 2 weeks of study, with Yij = 1 for non-abstinent and Yij = 0 for abstinent during the final 2 weeks of study; Uij is the vector of all appropriate covariates; Ij is the indicator variable for treatment with nabilone and XR-NTX; and εij is a random error term. Significant Ij indicates that the odds of abstinence from cannabis during the final 2 weeks are different in each group.

Hypothesis 3: Nabilone and XR-NTX will significantly increase time to dropout as compared to placebo. To test the secondary hypothesis 3, we will use the same model as in (1) to test whether retention is different between the two groups.

Hypothesis 4: Over time, the days of weekly cannabis use will be significantly less in the Nabilone and XR NTX group as compared to the placebo group. This hypothesis will be tested using the following

longitudinal mixed effects model: (3) $Y_{ijt} = \beta_0 + \beta_1 I_j + \beta_2 t + \beta_3 t*I_j + \beta_4 U_{ij} + s_{ij} + \varepsilon_{ijt}$ where Y_{ijt} is the days of cannabis used over a one-week period by the i th subject in the treatment group j at week t (t =1, 2, …, 12); U_{ij} is the vector of covariates; I_j is the indicator variable for treatment with nabilone and XR-NTX; s_{ij} is a random intercept for subject i and ε_{ijt} is a random error term. Significant interaction $t*I_j$ indicates that the effect of each treatment group is different over time (that corresponds to rejecting null hypothesis that $\beta_3 = 0$). If so, the effect of time will be estimated for each group separately and the groups will be compared (using contrast) in the last time point t=12. If the interaction term is not significant (i.e., the difference between the treatment arms does not change over time), we will refit the model without the interaction term and test the significance of the main effect of treatment (i.e. rejecting the null hypothesis that $\beta_1 = 0$).

Hypothesis 5: Over time, subjects in the nabilone and XR-NTX group will experience significant reduction in the pattern of cannabis withdrawal symptoms (measured as mean Marijuana Withdrawal Checklist score per week) compared to the subjects in the placebo group.

Hypothesis 6: Over time, subjects in the nabilone and XR-NTX group will exhibit significantly reduced cannabis cravings (as measured by Marijuana Craving Questionnaire) compared to subjects in the placebo group.

Hypothesis 7: Over time, subjects in the nabilone and XR-NTX group will experience significantly less sleep disturbance (as measured by the Medical Outcomes Study Sleep scale) compared to subjects in the placebo group. To test the secondary hypotheses 5, 6, and 7 we will use the same model as in (3) where Y_{ijt} is the appropriate outcome variable (as specified for each hypothesis) for the i th subject in the treatment group j at week t (t =1, 2, …, 12). We will first test the interaction effect of group and time as described for Hypothesis 4.

iii. Exploratory Analyses We will explore whether there are differences between the two groups with respect to the number of adverse effects, side effects, and tolerability in relation to the primary outcome. We will also explore whether co-occurring psychiatric disorders, gender, concurrent treatment exposure, trauma history, or impulsivity are predictors of treatment outcome.

f. Power Analysis: The sample size of 72 was chosen to ensure sufficient power (at least 80%) of a two-sided test with level of significance $\alpha=0.05$ for detecting difference between the two experimental treatments with respect to time to relapse, based on the Cox proportional hazards model. We expect that the nabilone and XR-NTX group will have greater time to relapse than the placebo group. The rate of retention for the placebo group at 12 weeks can be expected to be 15% based on our prior studies. With 36 per arm, we will be able to detect a difference in retention (without relapse) rates between placebo versus nabilone and XR-NTX of approximately 25% (i.e. 15% on placebo vs 39.6% on nabilone and XR-NTX). This corresponds to a hazard ratio of 0.49 (Machin et al. 1997). The proposed sample size will also enable us to estimate the retention rate of the treatment group with a 95% confidence interval, ±15%, in the worst case. The computation above was performed assuming that the covariates (i.e., baseline cannabis use and age) do not have a predictive power. The 25% rate difference is consistent with a medium to large sized effect, which is what we are seeking in this early Phase II trial. This represents an effect that would be clinically meaningful and likely to make a substantial impact on the success of XR-NTX treatment in clinical practice

Amended for Data and Safety Monitoring Plan 4/15/2016

A2. Primary and Secondary Outcome Measures Primary Outcome Measure: (1) Cannabis withdrawal (mood, sleep quality and duration, food intake) (2) Cannabis relapse defined as (1) time to any MJ use as recorded by the timeline follow back method and confirmed by urine metabolite levels, or (2) time to dropout, whichever comes first, and (3) medication compliance (% of pills taken) or time to discontinuation. Participants who relapse will continue in the trial to obtain secondary outcome measures. Secondary Outcome Measures: (1) Abuse liability (medication liking, desire to take again) (2) Cognitive task performance (3) Side effects and tolerability.

A4. Power calculation and sample size Cannabis Withdrawal. We hypothesize that clonazepam will decrease symptoms of cannabis withdrawal (e.g. ,irritability, sleep efficiency). With a sample size of 80, the test of a single contrast at the 0.050 level will have >90% power to detect a 25% decrease in these ratings under conditions of medication administration compared to placebo, assuming the same between-level correlation and SD than our earlier studies. Cannabis Relapse: The time to relapse as recorded by the Timeline Follow Back method and confirmed by creatinine-normalized quantitative urine THC levels, or time to dropout of study, whichever comes first (survival outcome). The sample size of 80 was chosen to ensure sufficient power (at least 80%) of a two-sided test with level of significance 5% for detecting difference between the two experimental treatments with respect to time to relapse, based on the Cox proportional hazards model. We expect that the clonazepam group will have greater time to sustained relapse than the placebo group. The rate of survival for the placebo group can be expected to be 15% based on our prior studies. With 40 per arm, we will be able to detect a difference in proportion of surviving in placebo versus clonazepam of approximately 27% (i.e. 15% on placebo vs 41.81% on clonazepam). This corresponds to a hazard ratio of 0.46 (Machin et al. 1997). The power computation was performed assuming that the covariates (i.e., baseline cannabis

use and age) do not have any predictive power. If they do, the power of the primary hypothesis will be higher than 80%. The 26.81% rate difference is consistent with a medium to large sized effect, which is what we are seeking in this early Phase II trial. This represents an effect that would be clinically meaningful and likely to make a substantial impact on the success of clonazepam treatment in clinical practice. The proposed sample size n=40 for clonazepam group will also enable us to estimate the relapse and retention rates of the treatment group with a 95% confidence interval, ±15.5%, in the worst case.

C. DATA MANAGEMENT AND ANALYSIS C1. Data Acquisition and Transmission Research assistants trained in behavioral observation techniques will assist in data collection during laboratory sessions. Subjective responses, performance effects, food intake and sleep will be obtained electronically using our custom-designed software programs. Some data (e.g., sleep questionnaires) will be recorded manually, and then entered into an Excel spreadsheet. C2. Data Entry Methods As noted above, a portion of the data is automated and a portion is recorded on data sheets, and transferred manually to an Excel spreadsheet. C3. Data Analysis Plan Subjective effects, food intake, sleep, and task performance will be analyzed using repeated measures analysis of variance (ANOVA) models with planned comparisons calculated with SUPERANOVA, SPSS or SYSTAT packages. We will use a cluster analysis of peak subjective effects ratings to reduce the overall number of comparisons (Haney et al., 2013a,b). Tests of differences will be based on F-statistics with degrees of freedom corrected, depending on the observed within-subject correlation of the measures, using the method of Huynh and Feldt. Planned contrasts will be single degree of freedom comparisons using the appropriate interaction error term. Results will be considered statistically significant at $p < 0.05$, using 2- tailed tests. The primary analyses in this study will be on the intent-to-treat (ITT) sample, i.e. on all randomized patients. Cannabis consumption and cannabis abstinence, are assessed repeatedly over time and we will try to collect them at all study assessment points. We will account for unobserved

data by examining the primary outcome variables using longitudinal mixed effects models (MEM) (Brown & Prescott, 1999; Diggle, et al., 2002) using PROC GLIMMIX in SAS®. MEMs do not require complete measurements data to estimate the outcome variable. The inferences from analyses with missing data are valid provided that they are "missing at random" (Little & Rubin, 2002). 'Missing at random' (i.e the missing mechanism does not depend on the value of the unobserved outcome) is untestable in most medical research and in our study as well. One can assume either parametric or semi-parametric models for the missingness that does depend on the unobserved outcome value and do the analysis (Diggle & Kenward, 1994; Kenward, 1998; Liu, et al., 1999; Rotnitzky, et al., 1998; Scharfstein, et al., 1999). Comparison of the inferences from assuming various models for the missingness provides a measure of the validity of the efficacy estimate from the model that assumes missing 'at random'. One can also compute a local sensitivity index which measures the change in the estimated treatment effect in a neighborhood of the 'missing at random' model for missingness (Rotnitzky, et al., 2001). We plan to perform a sensitivity analysis based on these two approaches to assess the effect of the assumption of missing 'at random' on the inference.