**Statistical Analysis Plan**

**Study Title:** Improving Treatment Outcomes for Prescription Opioid Dependence

**NCT Number:** NCT02543944

Date last reviewed by IRB: 10/4/2021

**Data Analysis Methods:**

We will compare baseline characteristics of the two groups using the intent-to-treat sample with analyses of variance for continuous variables (e.g., age) and $Chi^2$ analyses for categorical variables (e.g., sex, race) to determine whether any significant baseline differences have accrued despite randomization. In addition, given that later phases (i.e., transition, depot NTX therapy) may have different sample configurations due to anticipated dropout, we will similarly compare the two groups on participant characteristics. If we find significant baseline differences in variables within a phase or at each subsequent phase relative to the one prior, these differences will be informative of factors associated with treatment outcomes. For those variables found to affect outcome, we will use contingency tables with stratification or add these factors to the appropriate random regression models to adjust for these differences.

Our primary outcome is changes in opioid positive urines over time during the detox phase. Other major outcomes are: percentage of detox phase completers, naltrexone phase initiators and vivitrol receivers.

Specific Aim 1: Efficacy and tolerability of GBP versus Placebo during and immediately following BUP detox. For data during the BUP taper (weeks 2-3), dependent variables obtained at several time points (e.g., urine sample results for opioids, opioid withdrawal symptoms scores) will be entered into random regression models, also known as hierarchical linear models (HLMs), to determine whether scores change differentially across treatment groups [154, 155]. Continuous data will be analyzed longitudinally with MIXREG, an HLM modeling program for continuous measures [156]. Dichotomous urine results (i.e., negative or positive) will be analyzed longitudinally using the SAS procedure GLIMMIX, which allows an HLM modeling program for ordinal outcome measures [157]. We will use all available data in our analyses and will make no attempt to interpolate missing data. HLM methodologies fit a regression line for each participant, effectively interpolating missing data, before deriving final estimates. This approach of modeling repeated measures is specifically designed for use in repeated measures designs with missing data, allowing for intra-participant serial correlation and unequal variance and covariance structures over time. These problems, common to clinical trial data, are solved by incorporating available trend data for each individual with information on the behavior of the group from which the participant is drawn. If there are any significant baseline differences, the variable will be added as a cofactor in the HLM analyses. We will examine differences in retention using Kaplan-Meier Survival Analysis. Analyses will be done on those who participated long enough to start the BUP taper, as well as during GBP induction to detect possible differential dropout before the taper. To determine whether treatment group differences occurred in 1) success of retention from the BUP detox to the NTX transition and 2) eligibility to undergo oral NTX induction, we will build 2 X 2 contingency tables using percentages of participants in the GBP vs. placebo groups who 1) return on day 1 of week 4 versus drop out after completing the taper, and 2) test negative versus positive for opioids. The type, severity, and frequency of adverse events will be compared across groups. All analyses will employ a significance level of $\alpha = 0.05$, and all tests are two-tailed.

Specific Aim 2: Acceptability and feasibility of outpatient transition to and maintenance on depot NTX therapy. Because there are two phases to NTX therapy transition (i.e., 4-day oral induction and up to 8-week depot NTX), data obtained during these phases will be analyzed separately. For the oral NTX induction, descriptive statistics will be used to report the percentage of abstinent participants who receive the 4-day oral dosing. Urine data collected each day during the 4-day induction will be entered into random regression models similar to those above to determine whether changes in illicit opioid or other drug use occurred over time. To obtain pilot data on the association between treatment group (GBP vs. placebo) and successful induction, data regarding the percentage of those receiving oral NTX and the NTX injection or not in the GBP vs. placebo groups will be entered into a 2 X 2 contingency table. The type, severity, and frequency of opioid withdrawal and/or adverse events will be compared across groups. Reasons for not receiving the NTX injection will be summarized.

For the NTX phase (weeks 4-12), dependent variables obtained at weekly time points (e.g., urine sample results for opioids and other drugs, side effect scores, opioid withdrawal scores, etc.) will be entered into random regression models (HLMs) to determine whether scores change differentially over time [154, 155]. Continuous data will be analyzed longitudinally with MIXREG [156]. Dichotomous urine results (i.e., negative or positive) will be analyzed longitudinally using the SAS procedure GLIMMIX [157]. We will examine differences in study dropout during this phase using Kaplan-Meier Survival Analysis. Our principal analyses will be done on those who participated long enough to start the depot NTX therapy. The type, severity, and frequency of adverse events will be recorded and summarized.

Specific Aim 3: Identify prognosticators of treatment outcomes. Predictors of treatment efficacy during each phase will be examined using baseline assessments (e.g., sex, level of opioid use, distress tolerance

scores, etc.). We will evaluate general predictors of successful outcomes (with success as defined above) regardless of treatment type and predictors of differential response to treatments. Depending on the distribution of participants with moderate versus severe opioid dependence, severity of opioid dependence will also be examined as a prognosticator. Our basic analytic strategy will be to dichotomize groups into the presence and absence of a dichotomous predictor variable (e.g., female or alcohol dependence diagnosis) and then to examine the relative predictive values on treatment outcomes by adding each factor to the log linear analyses for single time point outcomes and HLM for longitudinal measures such as frequency of self-reported drug use, opioid-negative urines and craving measures. Each continuous measure (e.g., HAM-A, distress tolerance scores) will be entered as a covariate in the above analyses.

Longer-Term Outcomes. Continuous data from the 16-week follow-up interview will be initially entered into repeated measures ANOVA's with group (GBP vs. placebo), successful completion of detox (yes vs. no), successful induction onto oral NTX therapy (yes vs. no), and successful transition to depot NTX therapy (yes vs. no) as factors. Dichotomous follow-up data will be entered into $Chi^2$ tables for each phase. Longitudinal data captured during each phase of the study and at follow-up will be entered into random regression models as above, and piece-wise comparisons will be made between each phase. For all analyses, a p value $<0.05$ will be used to infer statistical significance.

Sample Size. PASS (Power Analysis & Sample Size) Software was used to estimate power. Sample size for the intent-to-treat sample was determined based on having enough power during and immediately following the BUP taper to detect group differences in the primary outcome of illicit drug use and identify potential prognosticators. Based on the proportion of participants in the GBP versus placebo groups having opioid-free urines on the last day of the BUP taper from our pilot trial with GBP [31, 129], treating the one missing value as positive, a difference in the percentage of participants with opioid-free urines at the end of detox as well as immediately prior to initiating NTX therapy will be detected with >90% power with an analysis sample size of 106 total participants anticipated to remain for NTX transition. Thus, our sample size is adequate to detect differences for our primary outcome during and immediately following the detoxification. Meanwhile, there has been minimal, if any, study of sex (or other bivariate prognosticator) differences for the effect of an adjunct medication for opioid dependence, much less for prescription opioid dependence. This study will provide unique data to quantify a potential difference. Given our expected sample size of 150 participants entering the BUP detox phase, the study will have >80% power for an effect size of 0.29 (small effect size), and 106 participants eligible for NTX transition will have >80% power to detect an effect size of 0.35 (considered a "medium" effect size).