## STATISTICAL ANALYSES

**Data Analysis Plan:** The distributions of all continuous variables will be checked for normality, and transformations will be employed (if necessary) before applying specific parametric techniques. Before performing specific analyses, we will examine all outcomes for outliers. All statistical comparisons will be two-sided and the critical level for rejection of the null hypothesis will be considered to be a p-value of $< 0.05$. Where needed, False Discovery Rate Analysis will be used to adjust critical p values to control for multiple comparisons. Any comparison where homogeneity is violated will not be considered significant. Means, standard deviations, proportions, and 95% confidence intervals will be used to describe all continuous variables.

**Dropouts and Missing Data:** As this is an inpatient laboratory study within a controlled setting, we expect only minimal intermittent missing data, though we do expect to have some subjects drop out without completing all of the experimental conditions. We will use all available data from subjects prior to drop-out. We will account for missing data directly using mixed-effects models (MEM) (Diggle et al. 2002). MEMs do not require complete measurements to estimate effects and inferences are valid provided that data 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). 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. 1998). We will perform sensitivity analyses based on these two approaches to assess the effect of the assumption of missing 'at random' on the inference.

**Randomization:** Once a participant is enrolled, a randomization sequence table will be generated by a blinded statistician. The table will determine the order of testing of the various dose combinations of dronabinol (.00, .35, and .70 mg), and naloxone (0.0, 0.2 and 0.4 mg).

### Aims, Outcome Measures, and Hypotheses

**Primary Aim:** Assess the ability of inhaled dronabinol to alter the severity of naloxone-precipitated withdrawal.

- The primary dependent variable (DV) for this aim will be the total opioid withdrawal score from the testing session. Repeated-measures analysis of variance (ANOVA) will be used to assess differences among the various dose conditions.

- We hypothesize that there will be a dose-dependent effect of dronabinol to attenuate the severity of naloxone-precipitated withdrawal.

**Secondary Aim**: Assess the safety of inhaled dronabinol in combination with naloxone.

- The primary DVs are the SAFTEE (AEs not related to withdrawal) and physiological parameters. Generalized mixed-effects models with participant as a random effect and an autoregressive correlation structure for the within-participant observations will be used to analyze the total number of AEs and their severity.

- We hypothesize that dronabinol, in combination with naloxone, will produce more AEs than placebo but less than naloxone alone.

**Secondary Aim**: Assess the impact of inhaled dronabinol on naloxone's effects on pupil diameter (i.e., μ-opioid receptor blockage).

- The primary DV will be pre-to-post- medication change in pupil diameter from each testing session. Repeated-measures ANOVA will be used to assess differences among the various dose conditions.

- We hypothesize that dronabinol will not significantly alter naloxone's effects on pupil diameter (in comparison to placebo).

**Secondary Aim**: Assess the abuse potential of inhaled dronabinol.

- The primary DV will be pre-to-post- medication change in positive subjective drug ratings such as drug "Liking" and "Good Effect."

- We hypothesize that self-reported dronabinol "Liking" will not be significantly greater than placebo.

## Power

To approximate a suitable number of subjects for a traditional hypothesis-testing approach, a power analysis was conducted using G* Power 3.1.7. Estimates of the magnitude of the medication effect on withdrawal were estimated using a previous study that assessed naloxone-precipitated withdrawal (Jones et al., 2016). Based on these data we calculated a Cohen's D effect size of .75, which constituted a 15-point difference (mean SD of 20) in total withdrawal score (0-125) between any two dose conditions. Applying this to the current proposal, a total sample size of 16 completers should provide 90% power to detect this difference at a p of < 0.05. Sixteen completers will provide 88% power to detect a 10-mm difference in positive subjective effects, e.g. "Liking" on a 100-mm VAS scale (alpha of 0.05). This analysis assumes a SD of 12 mm and a Cohen's D effect size of 0.84 based on previous visual analog scale assessments of abuse liability of opioids and cannabinoids (Cooper et al., 2018, Jones et al., 2017).

Due to the exploratory nature of this study, effect size estimates will be the main research outcome for the secondary aims, along with the interpretation of confidence interval lengths and degree of overlap to indicate precision and degree of difference (i.e., threshold of clinical importance) between dose conditions (respectively). A sample size of 16 completers was determined to be sufficient to obtain precise 95% confidence intervals with a margin of error of at most 25%. Thus, with regard to the detection of differences in all the study aims, these calculations suggest that this study is suitably powered.

## Sex as a Biological Variable

Sex-specific comparisons will be performed to provide an indicator of a Sex x Treatment interaction. The within-subject nature of the study's design should control for potential sex differences. Though not powered to specifically detect sex differences, we are still in compliance with NIH policies to enroll males and females. Furthermore, sex as a biological variable will be examined and reported in accordance with NIH policy (https://orwh.od.nih.gov/sites/orwh/files/docs/NOT-OD-15-102_Guidance.pdf).