STATISTICAL ANALYSIS PLAN

RANDOMIZED, DOUBLE-BLIND, PLACEBO CONTROL STUDY OF A SINGLE VERSUS REPEATED INTRAVENOUS SUB-ANESTHETIC KETAMINE TREATMENT IN REFRACTORY DEPRESSION

NCT02360280

LAST IRB APPROVED VERSION: 06/05/2018

Data Analysis Plan

Sample Size Determination

The power analysis was performed for MADRS score using nQquery Advisor 4 (Statistical Solutions, 2000) under the following assumptions: (1) analyses of covariance (ANCOVA) with the main effects of treatment (a single vs. six infusions) and time (0 and Day 13), and the treatment by time interaction; (2) compound symmetric covariance matrix, and; (3) 5% significance level. In our pilot study, we observed a mean MADRS change of 18.8 (SD=11.2) in a ketamine infusion group. Assuming the same effect size and variability in the ketamine group, we could detect a statistically significant effect when compared to an 8.8 point (or smaller) MADRS change in the midazolam group. The estimates for the mean, standard deviation, and intra-subject correlation obtained from **a sample size of 21 patients per group** will be required to detect a 10-point difference between the two groups in "change in MADRS score from baseline to post-infusion" with 80% power at the p=0.05 two-tailed significance level (t-test). The estimated minimum clinically important difference (MCID) for the MADRS ranged from 1.6 to 1.9 (Duru, Fantino 2008). We expected to enroll about two subjects per month treatment period (*see timeline below*). Considering an attrition rate of 30%, we plan to recruit a total of approximately 28 subjects per intervention or 56 subjects participants total.

Analysis Plan

Each subject will be assessed for the primary outcome variable (MADRS score) at baseline (1-2 weeks prior to starting the intervention), at 24-hour post infusion (Day 2, 4, 6, 9, 11, and 13), and weekly for 4 weeks, biweekly for 8 weeks and monthly for 3 months post-treatment. For each study subject, the expected probability of receiving the intervention will be estimated by propensity scores computed by logistic regression model. Potentially important confounding variables used to compute propensity scores will include age, pain intensity, length of current MDE, and level of anxiety. For **Aim #1**, (e.g., the analysis of the efficacy of a single vs. six infusions in change from baseline in MADRS score over 12 days of treatment), we will use ANCOVA to compare both groups at Day 13 with baseline MADRS and propensity score included as adjustment covariates . For secondary outcomes (Response defined as ›50% decrease in MADRS baseline score, and Remission defined as MADRS score ≤9), logistic regression models will be used to compare two treatment groups at Day 13 with baseline MADRS and propensity score included as adjustment covariates in both analyses. For **Aim #2** (e.g., the analysis of durability of antidepressant effect assessed during 6-month follow-up among post-infusion responders), Cox proportional hazards regression will be used to compare groups on the elapsed time from post-infusion response to occurrence of relapse (defined as ‹50% of baseline MADRS score) at each follow-up visit. Patients who do not relapse will be censored at the last evaluation visit. Post-infusion MADRS and propensity score will be included as adjustment covariates. For the **exploratory aim #3** (e.g., trajectory of change in depression severity among patients who receive six IV ketamine infusions over a 12-days treatment), a linear mixed model will be used to compare change in MADRS score with baseline MADRS score and propensity score included as adjustment covariates. For exploratory aim #4, brain activity as measured by the blood-oxygen-level-dependent (BOLD signal) during an emotional task will be measured within the amygdala and hippocampus. Correlation of the time series of BOLD signal fluctuation between brain regions represents functional connectivity within networks. Paired *t*-tests or ANOVA will examine pre-post sedative-related changes over time in outcome measures (e.g. brain functional activation and connectivity, cognitive performance,

peripheral biomarkers). Pearson's correlations will be used to measure the relationship between change in secondary outcomes (activation and connectivity) with change in clinical symptoms.

Missing Data.

We will perform intent-to-treat analysis, meaning that all subjects randomized into the two treatment groups will be included in the analysis regardless of the extent of compliance with the treatment or withdrawals during the trial. This will create incomplete data because some subjects will withdraw from the study during the course of the treatment with their responses missing after withdrawal. If dropout process is related to the outcome measure (i.e., symptom level), this will present a challenge in the analysis. The majority of the currently available statistical methods assume that data are missing at random. However, in this study, it is plausible that the likelihood of dropout is related to the level of or change in symptom level (e.g., those with increase or no reduction in symptom level may be more likely to drop out of the study). When subjects drop out of the study, we will obtain data on reasons for dropout and analyze whether the dropout process is at random with respect to the outcome measures. If the dropout process is related to the outcome measures, we will utilize models that incorporate a nonrandom dropout mechanism (Little, Yau 1996). These analyses will be carried out as the secondary analyses for hypothesis-generating purposes.