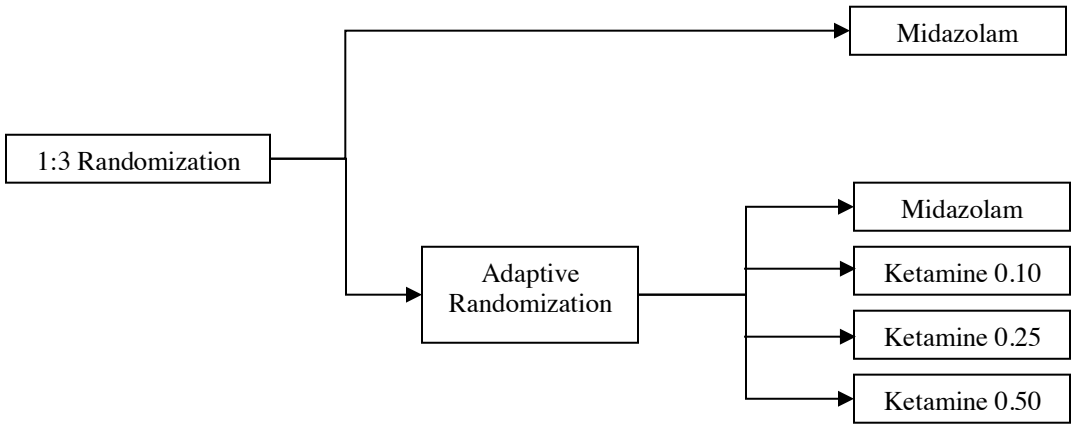
Updated Trial Design and Analysis Plan

The study protocol captures depression severity via the MADRS on a regular basis following administration of the pharmacotherapy; however, the MADRS is not obtained daily over the initial seven day post-infusion period. This renders the use of time to relapse problematic as both an adaptation criterion and for the final analysis. Given that we are accruing participants at ~ 1-2 per month, have currently randomized n = 22 patients, and anticipate initiating the adaptive component of the protocol at n = 25, we have elected to update the adaptation and analysis criteria to make the conduct of the study logistically feasible.

We will define the response, on which we will base adaptive randomization, as well as the final decision-making, as the probability that a condition results in response (i.e. $\geq$ 50% decrease from the baseline MADRS score) at day seven, following the infusion. The initial design planned to allocate a maximum of N = 72 participants to k = 4 conditions (i.e. midazolam, ketamine 0.10 mg/kg, 0.25 mg/kg and 0.5 mg/kg). An initial randomization was to assign participants in a 1:3 ratio to receive midazolam or further, adaptive randomization. The 75% of participants allocated to adaptive randomization will receive midazolam, ketamine 0.10 mg/kg, 0.25 mg/kg or 0.50 mg/kg (Figure AAA). The rationale for the initial 1:3 split, which remains in place, was that if investigators expectations held true, the adaptive algorithm would rapidly decrease the proportion of participants adaptively randomized to midazolam and increase those allocated to ketamine. The 1:3 split will guarantee a sufficient control group (i.e. midazolam) with which to credibly compare the best performing ketamine condition. Based on our initial estimates, the maximum number of participants initially randomized to midazolam or further, adaptive randomization would be n = 18 and n = 54 respectively. Among participants allocated to adaptive randomization (n = 54), the minimum number assigned to conditions (i.e. if all conditions stop for utility/futility) will be n = 20. Allocation of these initial n = 20 participants will utilize equal, random assignment probabilities prior to invocation of any adaptive decision rules.

All of the above procedures remain in place, with the exception that the alteration in the adaptation plan will permit a slight reduction in the sample size required for adaptive randomization from n = 54 to n = 49 resulting in an overall sample size of N = 67, slightly lower than the initially planned N = 72.

Figure AAA. Trial Design.



While we previously focused on the decision-rules that would maximize the time to relapse we now define these decision rules to maximize the probability of responding on day seven, following the infusion. Discussion of prior specification for the survival model is detailed in the previous version of the protocol. For the updated adaptation/analysis plan we derive the priors as follows. The anticipated times-to relapse were 2, 3, 4 and 7 days for midazolam, ketamine 0.10 mg/kg, 0.25 mg/kg and 0.50 mg/kg respectively. Assuming that the survival model curve follows an Exponential Distribution, the probability of remaining a responder at day seven is 0.09, 0.20, 0.30 and 0.50 for midazolam, ketamine 0.10 mg/kg, 0.25 mg/kg and 0.50 mg/kg respectively. Based on extant data and clinical experience we stipulated that the best and worst case

probabilities of seven-day response for each of these conditions were (0.1, 0.25), (0.10,0.30), (0.20,0.50) and (0.30,0.70) for midazolam, ketamine 0.10 mg/kg, 0.25 mg/kg and 0.50 mg/kg respectively. We use these probabilities to define the 2.5^{th} and 97.5^{th} percentiles of the respective prior distributions. Since the use of a binary decision-rule implies a binomial process, we will use a Beta-Binomial model for adaptation and decision-making. Assuming that the probabilities above constitute the 2.5^{th} and 97.5^{th} percentiles of a Beta Distribution for each treatment, we can calculate the Beta prior distributions as ~Beta(a = 1.797, b = 17.743), ~Beta(a = 10.849, b = 46.329), ~Beta(a = 12.607, b = 24.268) and ~Beta(a = 11.260, b = 11.260). These priors countenance a range of credible probability values encompassing both optimistic and pessimistic estimates of the probability of response at day seven for each treatment.

The randomization and adaptation rules remain similar with a number of exceptions, which we will provide in **bold** text. Following equal allocation of the first n = 20 participants to the four conditions, the adaptive decision-rules will take effect. If the response rate of the best performing condition is large enough that the posterior probability is >0.975 that it is better than the next best condition, this will trigger a stop to the study, and declaration of superiority for the highest-performing condition. Suspension of accrual will occur for any condition that demonstrates a response rate so small that the posterior probability is <0.025 that it is better than the best performing condition. Further, if data for any condition indicates **that a 10% chance of responding at day seven has a posterior probability < 0.05, that condition will be stopped for futility**. Finally, after allocation of all participants, subject to the preceding decision-rules, a condition will be declared superior, and a candidate for further outpatient testing, clinical trials if the posterior probability that its response rate exceeds that of the next best performing condition is **> 0.75**. For the purposes of adaptive randomization, any participant failing to demonstrate an initial response will be treated as **a non-responder at day seven**.

Updated Simulations and Operating Characteristics. Changes are provided in **bold** text. **The anticipated probability of remaining a responder at day seven is 0.09, 0.20, 0.30 and 0.50 for midazolam, ketamine 0.10 mg/kg, 0.25 mg/kg and 0.50 mg/kg respectively. Assuming these effects as well as the previously specified prior distributions and decision-rules, K = 10,000 Monte Carlo simulations indicated that the preceding decision rules would identify the best condition (i.e. ketamine 0.50 mg/kg) 95% percent of the time. This estimate corresponds to conventional power and exceeds the widely-used value of 80%. While the maximum planned sample size for adaptive allocation is N = 49, on average, the adaptive design required allocation of n = 42 participants to reach these conclusions. Specifically, over 10,000 Monte Carlo simulations, an average n = 5, 5, 8 and 24 participants were allocated to Midazolam, Ketamine 0.10 mg/kg, 0.25 mg/kg and 0.50 mg/kg by adaptive randomization. Clearly, n = 5 adaptively allocated to Midazolam fails to provide a credible comparison group for the best performing Ketamine 0.05 mg/kg condition (n = 24); hence the required initial randomization step which will augment the Midazolam groups with approximately n = 18 participants, and provide a credible comparison group (n = 23). Evaluation of the Type I Error rate requires ascertainment of the chance that the adaptive randomization design will identify a treatment as best when, in fact, the null hypothesis obtains. Setting the probability of seven-day response to 0.09 for all conditions simulates this scenario. K = 10000 Monte Carlo simulations indicate that the preceding decision-rules identify a condition as best 4.21% of the time. This compares favorably to the 5% Type I Error Rate usually implemented in conventional trials.**

Comparing these simulation results to the sample sizes required for a conventional, between groups approach permits determination of the degree to which to Bayesian adaptive approach confers benefit in decision-making about the most promising compound/dose for further investigation. Detection of a difference between probabilities of responding of 0.20, 0.30 and 0.50 at seven days post infusion (i.e. Ketamine 0.10 mg/kg, 0.25 mg/kg and 0.50 mg/kg) relative to a comparison group with a probability of responding of 0.09 (i.e. Midazolam) would require samples of n = 344, 124, and 44 respectively to achieve 80% power with a

two-sided alpha = 0.05. The Bayesian adaptive randomization approach permits flexibility in exploring multiple doses while maximizing allocation to more promising conditions.

Final Data Analysis

Preliminary data analyses will inspect baseline, group differences and compliance variables for correlations with specified outcomes. Variables demonstrating baseline group differences, for which there is a correlation with outcomes, will be treated as potential confounders.[1,2] Two sets of analyses will determine the degree to which any group differences might confound conclusions regarding treatment: one including, and one excluding the relevant variable as a covariate.

Given recommendations that all clinical trials should be analyzed in conventional, Frequentist and Bayesian fashion, parallel analyses will implement each approach for evaluation of Aims 1 and 2.[3] Frequentist results yield the probability of the observed data, or data more extreme, given that the null hypothesis holds. Bayesian results yield to probability that the governing parameter for an observed process equals some value or range of values. This permits statements regarding the probability that treatment confers benefit of some magnitude; a critical issue in treatment development. Statistical analyses will use R.[4]

Broadly, the analytic strategy will use generalized linear modeling. Continuous, dichotomous and time-to-event data will utilize linear, logistic, and proportional hazards regression respectively (Proc GENMOD and Proc PHREG; SAS v. 9.3). Longitudinal analyses will employ generalized linear mixed models (Proc GLIMMIX; SAS 9.3). Intention-to-treat analyses will evaluate time to relapse as a function of drug condition, collapsing participants initially randomized to midazolam with those assigned to the same condition via adaptive randomization. For intention-to-treat purposes, participants failing to demonstrate a response to treatment will counted as having relapsed on the first day of follow-up. *Per protocol* analyses will subsequently estimate effects among only those participants who demonstrated an initial response to treatment. Multiple imputation and maximum likelihood solutions, which are robust under assumptions of missing at random, will address missing data in cross-sectional and longitudinal analyses respectively.[5] Sensitivity analyses will evaluate robustness of analytic conclusions to missing data. Non-ignorable missing data patterns will be addressed through pattern-mixture modeling methods.[6]

For Bayesian analyses, unless otherwise indicated in the data analytic plan, priors will be neutral and diffuse. For linear, Poisson, and logistic and Cox Proportional Hazards regression, priors for coeficients will take the form ~N(mean = 0, var = 1×10^6) in the linear, log, log(odds) and log(hazard) scales respectively. **Evaluation of proportions will use beta-binomial models with both the previously stated Beta prior as well as with ~Beta(1,1) priors for the purposes of reporting. Priors for error or dispersion terms will use ~Uniform(1,100), a ~Half-Normal ($\mu = 0, \sigma = 100$) or a ~Folded T-Distribution(df = 3, $\mu = 0, \sigma = 100$).** Sensitivity analysis using optimistic and pessimistic, skeptical priors will evaluate prior assumptions.[7] Assessing the convergence of Bayesian analyses on the posterior distributions via Monte-Carlo Markov chain (MCMC) will use graphical (Trace Plot, Autocorrelation Plot) and quantitative (Geweke Diagnostics, Gelman-Rubin Diagnostics, and Heidelberger-Welsh Diagnostics) evidence. Evaluation of posterior distributions will permit statements regarding the probability that effects of varying magnitudes exist, given the data.

---

[1] Pocock, S. J. & et al. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat.Med, 21,* 2917-2930.

[2]Assman, S. F. & et al. (2009). Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet, 355,* 1067-1069.

[3] WIjeysundera DN AP, Hux JE, Beattie WS, Laupacis A. Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. J Clin Epidemiol. 2009;62(1).

[4] http://www.r-project.org .

[5] Enders, C. (2010) *Applied Missing Data Analysis*. Guilford Press.

[6] Fitzmaurice, G.M. & Laird, N.M. (2000). Generalized linear mixture models for handling nonignorable dropouts in longitudinal studies. *Biostatistics*. 1(2): 141-156.

[7] Spiegelhalter, D.J., Abrams, K.R. & Myles, J.P. (2004) *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons Ltd. West Sussex, England.