Psilocybin-facilitated Treatment for Cocaine Use: A Pilot Study

Statistical Analysis Plan

SAP Version 1.0

Protocol Version 2.0, 09/28/2022

SAP Revision History		
Date	*Modification*	*Justification*
09/28/2022	Replaced mixed-model ANOVA (+ multiple comparisons) with MMRM for % abstinent days; added Firth's penalized logistic regression for complete abstinence	To model within-subject correlation and MAR missingness; address small-sample/(quasi) separation; align with best practices prior to SAP finalization.
09/28/2022	Data analytic plan finalization	Basis for SAP

Roles and Responsibility/SAP Contributors

- Peter S. Hendricks, PhD, Principal Investigator

- Melissa Bradley, MPH, Independent/blinded statistician and SAP author

- Otto Simonsson, PhD, Independent/blinded statistician

Table of Contents

Introduction

Background and Rationale

Psychedelics such as LSD and psilocybin are psychoactive substances with low dependence potential that can induce profound, meaningful subjective experiences. From the 1950s to 1970s, early clinical studies suggested transdiagnostic potential across multiple conditions, including several forms of addiction. A 2012 meta-analysis of LSD-assisted psychotherapy demonstrated significant reductions in alcohol misuse. Research was largely halted on account of Schedule I drug classification and an informal federal research funding moratorium despite limited scientific rationale for such restrictions.

Psychedelic research has reemerged in the past two decades, once again promising efficacy for multiple mental health and substance use disorders. Controlled trials have shown sustained positive changes in well-being and behavior, while pilot studies report reductions in alcohol and nicotine dependence. Observational data from over 25,000 drug-involved offenders further suggest that naturalistic hallucinogen use is associated with decreased supervision failure, contrasting with other illicit drugs.

Cocaine dependence remains a major public health challenge, contributing to infectious disease transmission, violence, and neonatal drug exposure. Approximately 1.1 million Americans meet criteria for cocaine abuse or dependence, and cocaine is a leading drug of treatment admission in Alabama. Given psilocybin's safety profile and preliminary evidence for substance use treatment, this trial will evaluate the feasibility and estimate the efficacy of psilocybin-assisted psychotherapy for cocaine dependence. The primary research question is whether psilocybin-facilitated treatment, compared with an active comparator, increases cocaine abstinence and prolongs time to cocaine lapse.

Objectives

Primary Objectives:

- Test psilocybin in conjunction with cognitive behavioral therapy in the treatment of Cocaine Use Disorder
    - *Hypothesis 1*: psilocybin, as compared to placebo, will yield a higher percentage of cocaine abstinent days as assessed with the Timeline Followback Interview (TLFB) and verified by urine toxicology
    - *Hypothesis 2*: psilocybin, as compared to placebo, will yield a greater likelihood of complete abstinence from cocaine as assessed with the TLFB and verified with urine toxicology
    - *Hypothesis 3*: psilocybin, as compared to placebo, will yield greater latency to first cocaine lapse through 180 days after the end of treatment as assessed with the TLFB and verified with urine toxicology

Secondary Objectives:

- Investigate the impact of psilocybin-assisted psychotherapy on:
    - Severity of cocaine dependence
    - Cocaine withdrawal, craving, abstinence-self-efficacy, and motivation to quit/remain abstinent
    - The use of other drugs (percentage of use/abstinence, complete abstinence), particularly alcohol
    - Indicators of quality of life
    - Psychological constructs
    - Mediators of treatment with an emphasis on subjective drug effects
    - Neurobiology (Using MRI/fMRI/MRS)

Primary objectives per study trial protocol are evaluated in this Statistical Analysis Plan and secondary trial outcomes will be formally evaluated in subsequent analyses.

Study Methods

Trial Design

This is a quadruple-blind, randomized controlled trial (N = 40; 1:1 allocation ratio) with a parallel-group design. Participants will be randomized into two groups: psilocybin (n = 20) or diphenhydramine (n = 20). Participants in the psilocybin group will receive 25 mg/70 kg of psilocybin. Participants in the diphenhydramine group will receive 100 mg of diphenhydramine. All participants will receive empirically validated and manualized cognitive behavioral treatment for cocaine dependence ("Cocaine-Specific Coping Skills Training").

Randomization

Randomization will be performed by UAB's Investigational Drug Service $\leq$ 48 hours prior to drug administration using a random number table. No minimization or stratification will occur. Allocation concealment will be maintained by Investigation Drug Service personnel, who will have no further involvement in the study.

Sample Size

Sample size calculations were not be performed due to an absence of prior data on the magnitude of the expected effect.

Framework

Superiority hypothesis testing will be performed. Treatment group comparisons of percentage of days abstinent from cocaine, complete abstinence from cocaine, and time to cocaine lapse will be presented using a superiority framework.

Statistical Interim Analyses and Stopping Guidance

No formal interim analysis is planned. The trial may be stopped early in the event of clear evidence of participant harm or other serious safety concerns, as determined by the Principal Investigator in consultation with the Data Safety and Monitoring Board and IRB.

Timing of Final Analysis

The final analysis of efficacy will be performed after enrollment of the last subject and final data collection and cleaning is performed. Preliminary data analysis may be conducted for regulatory reporting purposes, with presentation as appropriate.

Timing of Outcome Assessments

Percentage of abstinent days will be calculated across six periods: (1) in the past 90 days preceding the first prescreening visit (Baseline); (2) during the prescreening period (Baseline through the first preparation session); (3) during the preparation period; (4) during the integration period; (5) from the final integration session through Day 90; and (6) from Day 90 through Day 180. Complete abstinence will be assessed from the end of the all-day drug session through Day 180; participants lost to follow-up will be coded as non-abstinent. Time to first cocaine lapse will be measured as the number of days from the drug administration visit (considered Day 0) to first cocaine use through Day 180.

Statistical Principles

Confidence Intervals and *p*-values

Statistical significance levels of 0.05 and 95% confidence intervals will be reported. No adjustments will be made for multiplicity, however, exact *p*-values up to four decimal places will be reported so that readers of the final report can estimate a conservative Bonferroni-type correction of their choosing.

Adherence and Protocol Deviations

This trial is intended in part to evaluate the feasibility of the research protocol, and as such protocol deviations are not predefined or formally assessed. However, deviations from the written study protocol, such as rescheduling specified study visits, will be identified and documented. Adherence will be defined as attending study visits and abstaining from cocaine use for seven days prior to drug administration.

Analysis Populations

The intention-to-treat (ITT) population will include all randomized participants, regardless of adherence or protocol deviations, and will serve as the primary analysis set.

Trial Population

Screening Data

Screening data to be collected include plans for moving from the study area, age, cocaine use, cocaine preference, other substance use history, interest in quitting, reading or writing in English, availability of transportation to study sessions, current pregnancy and nursing status and plans, fMRI-relevant questions (history of brain trauma, metallic objects in body, fear of enclosed spaces), and questions related to mental and physical health history. Aside from a general eligibility/ineligibility indicator per CONSORT reporting guidelines, screening data will not be reported as individuals who do not consent to participate do not agree to have their personal or demographic data, particularly PHI, publicly reported.

Eligibility

This trial has the following inclusion criteria: adults 25 years of age or older who score at least 3 on the Severity of Dependence Scale, identify cocaine as their drug of choice, and use cocaine on at least 4 separate days in the past month. Participants must also indicate a desire to cease cocaine use via a goal of complete abstinence on the Thoughts About Abstinence questionnaire, have the ability to read and write in English, no prior hallucinogen use or at least 3 years since their last use of a hallucinogen, have transportation available following their drug administration session, the ability to attend two consecutive prescreening appointments, determined to be in good general health as assessed by a detailed medical history interview and physical examination, DSM-IV diagnosis of cocaine dependence on a structured interview with no additional substance dependence diagnoses other than tobacco/nicotine, and abstinence from cocaine for at least 7 days prior to experimental drug administration, as confirmed via urinalysis, and no signs of intoxication on other drugs.

Exclusion criteria include women who are pregnant or breast feeding, current psychiatric diagnoses other than cocaine dependence (or other substance use disorders per inclusion criteria), current hypertension, use of tricyclic antidepressants, lithium, SSRIs, MAOIs, haloperidol or other antipsychotic medications, mood stabilizers, St. John's wort, or medications with serotonin activity, personal or family history (first- or second-degree relatives) of any psychotic or bipolar disorders, current suicidal or homicidal ideation, plans to move from the study site area in the next 6 months, contraindications of fMRI (metallic objects in the body, claustrophobia, difficulty with prior fMRI, prior head trauma or loss of consciousness, or neurological disease), or contraindications of diphenhydramine (glaucoma, difficulty urinating due to an enlarged prostate, or breathing problems such as emphysema or chronic bronchitis).

Recruitment

The CONSORT diagram will report counts at each stage, including excluded/discontinued/lost to follow-up with reasons as applicable: telephone screen, prescreening, medical eligibility visit, enrolled (entered preparation), discontinued before randomization, randomized (1:1), received allocated intervention (ingested study drug), completed integration, completed Day 90 follow-up,

completed Day 180 follow-up, and analyzed (ITT; safety and per-protocol noted in the caption). Definitions and windows per protocol.

Withdrawal/Follow-Up

Intervention withdrawal will be identified and documented among participants who decline or do not attend the drug administration session after enrollment, and those who decline or do not attend all post-drug integration sessions. Loss to follow-up will be documented among participants who decline or fail to attend Day 90 and Day 180 visits after completing the drug administration session.

Baseline Participant Characteristics

Baseline sociodemographic characteristics will be summarized for the intention-to-treat population by randomized arm. Variables summarized will include gender, ethnic origin, age, sexual orientation, marital status, employment status, annual individual income, educational attainment, living situation, and U.S. Armed Forces veteran status. Drug use history variables will include preferred route of cocaine administration, age at first cocaine use, age of onset of regular cocaine use, number of prior cocaine quit attempts, lifetime history of treatment for cocaine use, longest duration of cocaine abstinence, daily tobacco use in the past 90 days, heavy drinking in the past 90 days, illicit drug (e.g., cannabis, methamphetamine, opioid) use in the past 90 days, and lifetime history of hallucinogen use. Categorical variables will be presented as n (%) and continuous variables will be presented as median (Q1, Q3). Missing/Unknown will be shown as its own category, and denominators will be the number randomized per arm. No hypothesis testing will be performed.

Analysis

Outcome Definitions

This trial has three co-primary outcomes, all assessed by retrospective calendar-based interviews (TLFB) with urine verification: (1) Percentage of cocaine abstinent days, calculated as 100 × (number of days with no cocaine use ÷ number of TLFB-observed days) and summarized across six periods (past 90 days at Baseline, prescreening, preparation, integration, final integration session through Day 90, and Day 90 through Day 180); (2) Complete cocaine abstinence from the end of the all-day drug session through Day 180, defined as no cocaine use whatsoever with participants lost to follow-up coded as non-abstinent; and (3) Time to cocaine lapse, measured in days from the all-day drug session to first cocaine use through Day 180, with observations without a lapse censored at last contact.

Analysis Methods

Baseline participant characteristics of the groups will be described, and as appropriate, compared with Wilcoxon rank-sum tests or independent samples t-tests for continuous variables and Fisher’s exact tests for categorical variables. Any variables that differ between groups will be included as covariates in supplemental sensitivity models, as appropriate.

Percentage of cocaine abstinent days will be compared between groups using mixed models for repeated measures (MMRMs) with first-order autoregressive covariance structures, first modeled across pre-randomization periods (Baseline, prescreening, preparation) and then across post-randomization periods (integration; final integration→Day 90; Day 90→Day 180) with a group × time-period interaction term for pairwise post-randomization comparisons. Estimated marginal means will be extracted to obtain between-group contrasts, and standardized effect sizes (Hedges' g) will be calculated. These models handle missing data via maximum likelihood using observed values and within-subject correlations without explicit imputation.

Fisher's exact test and Firth's penalized logistic regression will first compare groups on complete abstinence from the prescreening period through the preparation period. Fisher's exact test and Firth's penalized logistic regression will then compare groups on complete abstinence from the end of the all-day drug session to Day 180. Firth's penalized logistic regression will be used for complete abstinence analyses to mitigate small-sample bias and to address potential (quasi) complete separation in the outcome–predictor combination. We will report penalized-likelihood odds ratios with 95% confidence intervals and *p*-values. Fisher's exact tests are used for univariate comparisons when cell counts are small; Firth's logistic provides regression-based estimates when separation precludes ordinary logistic regression.

Time to cocaine lapse will be shown with Kaplan–Meier curves and a log-rank test, followed by a Cox proportional hazards model to estimate the hazard ratio assuming noninformative censoring. All randomized participants will be analyzed per intention-to-treat.

Per Stang and Baethge (2018), the rationale for testing groups on the primary outcomes prior to randomization is to confirm group comparability and detect potential sources of bias, ensuring valid causal inference.

Missing Data

Missing data will be reported by arm and time period as the number without outcome data at each time period, with reasons summarized when available (e.g., no-show, withdrawal). For percentage of cocaine abstinent days, the percentage will be computed using only days with TLFB entries within each prespecified period (i.e., 100 × abstinent days ÷ TLFB-observed days), so partial entries contribute to the denominator based on observed days. Lost to follow-up will be identified per visit window and reflected in the CONSORT flow diagram.

For analysis, MMRMs will use maximum likelihood with the observed data and within-subject correlations, without explicitly imputing values, for percentage of abstinent days across pre- and post-randomization periods. Complete abstinence through Day 180 will code participants lost to follow-up as non-abstinent in the primary analysis and will be compared using Fisher's exact test and Firth's penalized logistic regression. Time to cocaine lapse will be analyzed with Kaplan–Meier methods and a Cox model, assuming noninformative censoring with censoring at last contact.

Additional Analyses

Additional planned comparisons include: (1) group differences in the number of days from the first participant's drug administration date to each subsequent participant's drug administration date using a Wilcoxon rank-sum test or independent samples t-test; (2) group differences in the number of days taken to complete prescreening, preparation, and integration using Wilcoxon rank-sum tests or independent samples t-tests; and (3) group differences in Credibility/Expectancy Questionnaire scores using Wilcoxon rank-sum tests or independent samples t-tests. These variables will be included as covariates in sensitivity models if groups differ, as appropriate.

Harms

Harms will be summarized descriptively for all participants who receive a study drug (safety set). Adverse events (AEs) will be elicited at each visit, observed continuously during the all-day drug session, and specifically queried at the first integration session; any new or worsening events during/after dosing through Day 180 will be documented. Each AE will be classified and reported using (addendum: version 28.0) of the Medical Dictionary for Regulatory Activities (MedDRA), with Preferred Terms (PTs) applied for event categorization. AEs will then be further classified by relatedness, expectedness, and temporal proximity (i.e., during the all-day drug session or after the all-day drug session).

The primary safety analysis will be treatment-emergent AE incidence, defined as the number (%) of participants with ≥1 AE from drug administration through Day 180, summarized overall and by category. Complete listings of serious AEs and AEs leading to discontinuation will be provided. No formal hypothesis testing is planned for safety; results will be presented as an incidence table, with relatedness and expectedness tabulated.

Statistical Software

Analyses will be conducted in R (addendum: version 2024.04.02). For percentage of cocaine abstinent days, MMRMs will be fit using the *mmrm* package, with estimated marginal means and pairwise contrasts extracted with *emmeans* and the *lme4* package. Hedges' *g* estimates will be computed with *effsize.* For complete cocaine abstinence, Firth's penalized logistic regression will be fit with the *logistf* package. Time to cocaine lapse analyses will use the following packages: *survival* for Kaplan-Meier, log-rank, and Cox proportional hazards; *survminer* for Kaplan-Meier plotting. Model outputs will be organized with the *broom* and *tidyverse* packages.

References

1. The MMRM approach to handling missing data appears to outperform multiple imputation.

Siddiqui, O. (2009). MMRM versus MI in dealing with missing data—A comparison based on 25 NDA data sets. *Journal of Biopharmaceutical Statistics, 21*, 423-436.

2. Firth's penalized likelihood logistic regression is acknowledged as a non-standard method and will be used only to address small-sample bias and (quasi) complete separation when ordinary logistic regression is unstable.

Firth, D. (1993). Bias reduction of maximum likelihood estimates (1993). *Biometrika, 80*, 27–38.

Heinze, G. & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine, 21*, 2409–2419.

3. There are sound arguments for comparing the groups on the primary outcomes prior to randomization.

Stang, A. & Baethge, C. (2018). Imbalance *p* values for baseline covariates in randomized controlled trials: A last resort for the use of *p* values? A pro and contra debate. *Clinical Epidemiology, 10*, 531-535.

A data management plan was approved by the IRB. Data will be stored in a secure environment designed to maintain confidentiality and analyzed by independent and blinded statisticians. As there are no industry sponsors and this was an investigator-initiated trial, no drug master file (DMF) was referenced.