THE QUANTUM TRIP TRIAL (QTT)

STATISTICAL ANALYSIS PLAN

Can a single administration of psilocybin reduce alcohol intake in patients with alcohol use disorder? A randomized, double-blinded, placebo-controlled clinical trial.

Protocol Number #PSILO4ALCO
EU CT 2024-517497-17-01
Clinicaltrials.gov: NCT05416229

SAP date/status February 17, 2026/ first edition

Protocol version: February 14, 2024/Third edition

SAP contributors:

Investigator
Mathias Ebbesen Jensen, MD, PhD
Mental Health Center, Frederiksberg and Bispebjerg Hospital

Sponsor and Principal Investigator
Anders Fink-Jensen MD DMSc, Professor
Mental Health Center, Frederiksberg and Bispebjerg Hospital
University of Copenhagen, Denmark

Senior statistician
Claus Thorn Ekstrøm PhD, Professor
Section of Biostatistics, University of Copenhagen, Denmark

Signatures:

Anders Fink-Jensen

_____ Date: _____

Claus Thorn Ekstrøm

_____ Date: _____

Mathias Ebbesen Jensen

_____ Date: _____

INTRODUCTION

The aim of the Quantum Trip Trial (QTT) was to test whether a single dose of 25 mg psilocybin with psychotherapy could reduce alcohol intake in patients diagnosed with alcohol use disorder.

This statistical analysis plan prespecifies analyses for the primary and secondary outcomes of the trial and is intended to support multiple planned publications arising from the study.

Prior to analyses, the statistical analysis plan will be uploaded to ClinicalTrials.gov and the dataset will be locked. All analyses will be conducted under blinded conditions.

HYPOTHESES

- **Primary hypothesis:**
  - Participants receiving psilocybin will show a greater reduction in alcohol consumption as measured by percentage heavy drinking days the past 28 days using the Timeline Followback Method (TLFB) from baseline to Week 12 compared to participants receiving placebo.

- **Secondary hypotheses:**
  - Participants receiving psilocybin will show greater increases in neuroplasticity, measured in whole-blood levels of blood brain-derived neurotrophic factor (BDNF), from baseline to Week 1 compared with participants receiving placebo.
  - A mean-centering Partial Least Squares analysis will show a group (psilocybin vs placebo) × condition (alcohol vs neutral) interaction, in the anterior cingulate cortex, insula, striatum, and frontal gyrus where the neural differentiation between alcohol and neutral cues will differ between psilocybin and placebo groups, with reduced differentiation in the psilocybin group. Additionally, in-scan craving ratings following viewing pictures of alcohol (VAS 0-100) will be lower in participants randomized to psilocybin relative to placebo.
  - At Week 1, resting-state functional connectivity, defined as Fisher r to z-transformed correlations between BOLD time series extracted from the striatum (defined using the Tian subcortical atlas) and pre-specified regions of the salience network (defined using the Schaefer/Yeo cortical atlas), will differ between participants randomized to psilocybin and placebo.

- **Exploratory hypotheses:**
  - Lower Week-1 neural differentiation between alcohol and neutral cues (lower latent variable scores from a separate single-group Partial Least Squares analysis collapsing across treatment groups) will be associated with greater reductions in alcohol consumption from baseline to Week 12. Differences between psilocybin and placebo

groups in striatal functional connectivity (striatum-salience network and striatum-prefrontal) at Week 1 will be associated with changes in alcohol consumption from baseline to Week 12.

- Resting-state functional connectivity in the Control networks (defined using the Schaefer/Yeo cortical atlas) will be lower in the psilocybin group compared to the placebo group at Week 1 and will be associated with changes in alcohol consumption from baseline to Week 12.
- Greater acute subjective effects measured on the dosing day will be associated with greater reductions in alcohol consumption from baseline to Week 12.
- At Week 1, resting-state functional connectivity, defined as Fisher r to z-transformed correlations between BOLD time series extracted from the striatum (defined using the Tian subcortical atlas) and all brain regions (defined using the Schaefer/Yeo cortical atlas), will differ between participants randomized to psilocybin and placebo. Results will be presented as effect sizes i.e., Cohen's D, and inferential statistics i.e., p-values will not be reported.
- Task-based fMRI analyses will be repeated using a GLM in SPM to provide a map of task response as well as group-level effects on the task response.

STUDY METHODS

STUDY DESIGN

The QTT trial was a randomized, double-blinded, placebo-controlled, parallel-group, single-site clinical trial for 12 weeks (main trial) and a long-term follow-up 26 weeks after the intervention. The trial compared single-dose psilocybin therapy vs placebo therapy. The treatment involved two preparation sessions, a high-dose psilocybin session (25 mg) or placebo (inert substance) session, and three integration sessions. Baseline data was collected at screening visit and outcomes were assessed 1, 4, 8, 12 (primary endpoint) and 26 weeks after the psilocybin/placebo session. A sub-group of participants underwent fMRI scan 1 week after the intervention (no baseline scan).

RANDOMIZATION

The participants were randomized in two groups (1:1): psilocybin therapy or placebo therapy. The randomization was stratified in terms of age (20-45, 46-70 years), sex, and baseline alcohol consumption i.e. the number of heavy drinking days in the last 28 days before inclusion measured by the TLFB-method (5-15, 16-28 days). The randomization tool in REDCap was used. Patients, investigators, other caregivers performing assessments, and persons performing data analysis have been blinded from the time of randomization and until the time of database unlock.

SAMPLE SIZE

The study was initially designed to enroll 90 participants, based on an a priori power analysis informed by effect sizes reported in the open-label pilot study by Bogenschutz and colleagues (2015). Assuming a two-

sided alpha level of 0.05 and 90% power, the required sample size was estimated to be 27 participants per treatment arm. Anticipating a dropout rate of approximately 40%, the total planned enrollment was therefore set at 90 participants. Due to delays associated with the COVID-19 pandemic and lack of financial resources to continue the study, recruitment was terminated before reaching the intended enrollment target. Following the decision to stop recruitment, a revised power calculation—based on the originally specified primary endpoint and the same effect size assumptions as the initial power analysis—indicated that 22 participants per treatment arm would provide approximately 85% power, corresponding to a total of 44 completers. As of December 2025, the observed dropout rate was substantially lower than anticipated (13%), resulting in 59 enrolled participants, of whom 51 completed the primary endpoint assessment. The number of completers therefore exceeded the minimum sample size required according to the revised power calculation. Of these, 28 completed fMRI scans, however only those with complete behavioural data and adequate quality imaging data for analyses will be retained for analyses. Imaging data will be quality controlled using fMRIprep.

TRIAL POPULATION

Recruitment

Recruitment began September 2023 and ended September 2025. Last patient's last visit in the main trial i.e., the 12-week outcome assessment was on 22nd December 2025. Last patient's last visit concluding the entire study i.e., the 26-week follow-up will be end of March 2026. Participants were recruited in part from outpatient clinics, but predominantly from society via advertisement.

Eligibility

Treatment seeking individuals aged 20-70, weighing 60-95kg, diagnosed with AUD and alcohol dependence according to DSM-V and ICD-10, respectively, who had at least 5 heavy drinking days within the past 28 days of screening, and no co-occurring substance use disorders (except nicotine) and no concomitant pharmacotherapy for AUD.

OUTCOMES

OUTCOME CATEGORY	SOURCE	TIME POINTS	PRIMARY COMPARISON (BETWEEN-GROUP DIFFERENCE)
Primary			
1. Percentage of Heavy Drinking Days (PHDD)	TLFB/28 days	BL, W4, W8, W12	W12, adjusted for BL
Secondary			
2. <u>Alcohol consumption:</u>			
• % drinking days	TLFB/28 days	BL, W4, W8, W12	W12, adjusted for BL
• Drinks per day	TLFB/28 days	BL, W4, W8, W12	W12, adjusted for BL
• Drinks per drinking day	TLFB/28 days	BL, W4, W8, W12	W12, adjusted for BL
• Total alcohol (g/day)	TLFB/28 days	BL, W4, W8, W12	W12, adjusted for BL
• % abstinent days	TLFB/28 days	BL, W4, W8, W12	W12, adjusted for BL

• WHO risk levels	TLFB/28 days	BL, W4, W8, W12	W12, adjusted for BL
• PEth	Blood	BL, W4, W8, W12	W12, adjusted for BL
• GGT	Blood	BL, W12	W12, adjusted for BL
• ALAT	Blood	BL, W12	W12, adjusted for BL
• MCV	Blood	BL, W12	W12, adjusted for BL
3. Brain imaging and neuroplasticity			
• Alcohol cue–related BOLD response and in-scan craving ratings (alcohol > neutral)	fMRI	W1	W1 (post-treatment only)
• Resting-state functional connectivity between striatum and salience -network regions	fMRI	W1	W1 (post-treatment only)
• BDNF	Blood	BL, dosing, W1, W12	W1, adjusted for BL
4. Psychometrics			
• SETS	Questionnaire	BL	BL
• PACS	Questionnaire	BL, W1, W4, W8, W12	W12, adjusted for BL
• AUDIT	Questionnaire	BL, W12	W12, adjusted for BL
• DUDIT	Questionnaire	BL, W12	W12, adjusted for BL
• AASE	Questionnaire	BL, W1, W4, W8, W12	W12, adjusted for BL
• MDI	Questionnaire	BL, W1, W4, W8, W12	W12, adjusted for BL
• MAAS	Questionnaire	BL, W1, W4, W8, W12	W12, adjusted for BL
• AAQ	Questionnaire	BL, W1, W4, W8, W12	W12, adjusted for BL
• SF-36	Questionnaire	BL, W4, W12	W12, adjusted for BL
• NEO-PI	Questionnaire	BL, W4, W12	W12, adjusted for BL
• FTND	Questionnaire	BL, W4, W8, W12	W12, adjusted for BL
• PEQ	Questionnaire	W4, W12	W12
5. Acute effects			
• SDI	VAS (0-10)	Dosing (0–360 min)	Association with W12 alcohol outcomes, adjusted for baseline
• MEQ30 / 11D-ASC / EBI / EDI / AWE	Questionnaire	End of dosing day	Association with W12 alcohol outcomes, adjusted for baseline
6. Safety			
• Adverse events	AE reporting	Throughout study	Descriptive

Abbreviations: BL = baseline; W1, W4, W8, W12 = Week 1, Week 4, Week 8, and Week 12; PHDD = percentage of heavy drinking days; TLFB = Timeline Followback; WHO = World Health Organization; PEth = phosphatidylethanol; GGT = gamma-glutamyl transferase; ALAT = alanine aminotransferase; MCV = mean corpuscular volume; fMRI = functional magnetic resonance imaging; BOLD = blood oxygenation level–dependent; ROI = region of interest; BDNF = brain-derived neurotrophic factor; SETS = Stanford Expectancy Treatment Scale; PACS = Penn Alcohol Craving Scale; AUDIT = Alcohol Use Disorders Identification Test; DUDIT = Drug Use Disorders Identification Test; AASE = Alcohol Abstinence Self-Efficacy Scale; MDI = Major Depression Inventory; MAAS = Mindful Attention Awareness Scale; AAQ = Acceptance and Action Questionnaire; SF-36 = 36-Item Short Form Health Survey; NEO-PI = NEO Personality Inventory; FTND = Fagerström Test for Nicotine Dependence; PEQ = Persisting Effects Questionnaire; SDI = Subjective Drug Intensity; VAS = visual analogue scale; MEQ30 = Mystical Experience Questionnaire (30-item version); 11D-ASC = 11-Dimensional Altered States of Consciousness Rating Scale; EBI = Emotional Breakthrough Inventory; EDI = Ego Dissolution Inventory; AWE = Awe Experience Scale; AE = adverse event.

Blinding

Analyses will be carried out with the treatment groups still blinded and labeled as “group 1” and “group 2”. Before dividing participants, the statistical analysis plan will be completed, signed, and uploaded at clinicaltrials.gov, and the data set locked. The final unblinding of treatment groups (psilocybin or placebo), will not be carried out until all statistical analyses are performed. The last long-term follow up-visit (Week 26) is in April 2026, and this data will be analyzed unblinded.

ANALYSIS

Population

Analyses will follow the intention-to-treat principle i.e., analyzing participants who have been randomized and who have also received a dosing with psilocybin/placebo.

Descriptive statistics

All outcomes will be presented using descriptive statistics, normally distributed data by the mean and standard deviation (SD) and skewed distributions by the median and interquartile range (IQR). Binary and categorical variables will be presented using counts and percentage.

Efficacy analyses

For continuous outcomes, analyses will use a baseline-adjusted linear mixed model for repeated measures including all observed timepoints for the given outcome. The primary estimand is the between-group difference in mean change from baseline at Week 12 (unless otherwise specified). Model assumptions will be assessed using standard diagnostic methods. If substantial deviations from model assumptions are observed, appropriate transformations of the outcome or alternative modeling approaches will be considered as sensitivity analyses.

Categorical outcomes assessed at a single time point will be analyzed using chi-square or Fisher’s exact tests as appropriate. Repeated categorical outcomes will be analyzed using GEE or generalized linear mixed models.

Missing data

Missing data will be handled using full information maximum likelihood (FIML), which uses all available observed data to estimate model parameters under the missing at random (MAR) assumption. While the

MAR assumption may not be valid we will also consider sensitivity analyses by setting missing observations to extreme but realistic values to bound realistic effect size. See below.

Sensitivity analysis

Because the MAR assumption cannot be formally tested, pre-planned sensitivity analyses will be conducted under missing not at random (MNAR) assumptions. Specifically, missingness is assumed to potentially depend on factors related the unobserved Week 12 outcome values. The following conservative scenarios will be evaluated:

- Complete-case analysis: Analyses will be restricted to participants with observed values at both baseline and the Week 12 endpoint. In the case of neuroimaging data, only complete cases will be analysed.
- Single imputation (conservative MNAR scenario): For participants with missing Week 12 data, the continuous Week 12 outcome will be imputed as a 50% reduction from the individual's baseline value (i.e., Week 12 = 0.5 × baseline), and the primary analysis model will be re-estimated.

Per protocol analysis

A per-protocol (PP) analysis will be conducted including only participants who did not violate key protocol requirements through Week 12. The PP analysis is intended to estimate the effect of the intervention under full protocol adherence and will be interpreted cautiously due to the potential for selection bias.

Pre-specified mechanistic/exploratory analyses

Analyses will examine associations between selected secondary endpoints and the primary endpoint. These analyses are intended to assess relationships between outcomes and to inform hypotheses regarding potential mechanisms of action. Associations will be evaluated using correlation analyses and/or regression models as appropriate, without adjustment for multiplicity.

- Week 1 neural differentiation between alcohol and neutral cues, quantified using latent variable scores from a mean-centering Partial Least Squares (PLS) analysis of whole-brain blood oxygenation level-dependent (BOLD) responses, and associations with subsequent changes in alcohol consumption and brain-derived neurotrophic factor (BDNF) levels. Resting-state functional connectivity at Week 1 (striatum to salience-network, and within control-network connectivity) and their associations with changes in alcohol consumption and changes in BDNF levels.
- Acute subjective effects during the dosing session and their associations with changes in alcohol consumption and changes in BDNF levels.

Significance level

The primary endpoint will be tested using a one-sided significance level of 0.05 with corresponding 95% confidence intervals, reflecting an a priori directional hypothesis that psilocybin will reduce heavy drinking days compared with placebo. The direction of the primary hypothesis and the use of one-sided testing were prespecified prior to database lock and unblinding and were informed by consistent evidence from prior

clinical studies. Effects in the opposite direction, if observed, will be fully reported and interpreted descriptively. All secondary and exploratory analyses will be conducted using two-sided significance tests. For Partial Least Squares fMRI analyses, latent variable significance will be assessed via permutation testing (1,000 permutations, $p < 0.05$), with voxel reliability determined by bootstrap ratios $\geq |2.5|$ (200 bootstrap iterations). The PLS analysis will include group (psilocybin vs placebo) as a between-subjects factor, testing for group × condition (alcohol vs neutral) interactions in whole-brain BOLD responses. Additionally, a separate single-group PLS analysis (collapsing across psilocybin and placebo groups) will be performed to derive individual latent variable scores reflecting each participant's neural differentiation between alcohol and neutral cues. These individual LV scores will be correlated with changes in alcohol consumption from baseline to Week 12. No multiplicity adjustments will be performed, and no additional covariate adjustments will be applied except for the baseline value of the endpoint in question.

Statistical software

All statistical analysis will be performed with R. fMRI data will be preprocessed using fMRIprep v25.1.0 (https://fmriprep.org/en/stable/) and denoised using nilearn in Python. Partial Least Squares analyses are performed using the Rotman-Baycrest PLS Toolbox in MATLAB. Resting-state functional connectivity analyses will be performed using custom code in MATLAB.