

# Statistical Analysis Plan

Official Title

A Randomized, Double-Blind Study of Single-Dose  
Psilocybin for Major Depressive Disorder (MDD)

Unique Protocol ID

**PSIL201**

NCT Number

**NCT03866174**

Document Date

**09-Aug-2022**

Usona Institute

2881 Woods Hollow Rd, Madison, WI 53711

Project Name: PSIL201

IND # 129532

Sponsored by: Usona Institute

## **Statistical Analysis Plan**

Prepared and Distributed by:  
The Emmes Company, LLC  
Rockville, Maryland

Version **3.0, 09AUG2022**

## TABLE OF CONTENTS

Table of Contents.....	ii
Version History.....	v
Table of Abbreviations .....	vi
1.0 Introduction .....	9
2.0 Study Objectives and Endpoints.....	10
2.1 Objectives .....	10
2.1.1 Primary Objective.....	10
2.1.2 Secondary Objectives .....	10
2.1.3 Exploratory Objectives .....	10
2.1.4 Exploratory Mechanism-Based Objectives.....	11
2.1.5 Safety Objectives .....	11
2.2 Study Endpoints .....	12
2.2.1 Primary Study Endpoint .....	12
2.2.2 Secondary Study Endpoints.....	12
2.2.3 Exploratory Study Endpoints.....	12
2.2.4 Exploratory Mechanism-Based Endpoints.....	13
2.2.5 Safety Endpoints.....	13
3.0 Study Design.....	14
3.1 Overall Study Plan and Design.....	14
3.2 Selection of Study Population.....	14
3.3 Efficacy and Safety Outcomes.....	15
3.3.1 Efficacy Outcomes .....	15
3.3.2 Safety Outcomes .....	15
Table 1. Efficacy Outcome Measures.....	16
Table 2. Scoring Methods for Efficacy Outcomes.....	19
Table 3. Safety Outcome Measures .....	21
4.0 General Statistical Considerations.....	25
4.1 General Principles .....	25
4.2 Timing of Analyses .....	25
4.3 Analysis Time Points .....	25
4.4 Analysis Populations .....	25
4.4.1 ITT Population .....	25

4.4.2	Per Protocol Population .....	26
4.4.3	Safety Sets .....	26
4.5	Randomization and Blinding .....	26
4.6	Sample Size and Power Considerations .....	27
Table 4. Summary of Power at Day 43 (Primary Endpoint) and Day 8 (Key Secondary Endpoint) for N = 100 .....		28
4.7	Early Termination and Dropouts .....	29
4.8	Missing Data .....	30
4.8.1	Missing Dates .....	30
4.9	Covariates .....	31
4.10	Subgroup Analyses .....	31
4.11	Significance Testing and Multiple Comparisons .....	31
4.12	Interim Analyses .....	32
4.13	Stopping Criteria .....	32
4.14	Data Monitoring .....	32
4.15	Software used for Analysis .....	32
4.16	Reporting Conventions .....	32
4.17	Derived Variable Definitions .....	32
4.17.1	Treatment-Resistant Depression .....	32
4.17.2	Sustained Depressive Symptom Response .....	33
4.17.3	Sustained Depressive Symptom Remission .....	33
5.0	Study Participants .....	34
5.1	Analysis of Participant Characteristics .....	34
5.2	Participant Disposition .....	34
Figure 1. PSIL201 CONSORT Diagram .....		36
5.3	Protocol Deviations .....	37
5.4	Measurement of Treatment Compliance .....	37
6.0	Analysis of Efficacy Endpoints .....	38
6.1	Efficacy Aims and Hypotheses .....	38
6.2	Efficacy Analyses .....	40
6.2.1	General Considerations .....	40
6.2.2	Primary Efficacy Analysis .....	40
Table 5. Sensitivity Analyses for Primary and Key Secondary Endpoints .....		44
6.3	Secondary Efficacy Analyses .....	45
6.4	Exploratory Outcome Analyses .....	45

Table 6: Syntax for LSMESTIMATE Statements to Test Secondary and Exploratory Outcome Hypotheses .....	47
6.5 Mechanism-Based Exploratory Analyses.....	48
Figure 2: Path with No Mediator .....	49
Figure 3: Path with Single Mediator, MEQ-30 or EBI.....	50
Figure 4: Path with Multiple Mediators, MEQ-30 and EBI .....	50
6.6 Supplemental Analyses for Efficacy Outcomes.....	51
Table 7. Supplemental Analyses .....	52
7.0 Analysis of Safety Endpoints .....	53
7.1 Adverse Events .....	53
7.1.1 Cardiovascular Proarrhythmic Adverse Events .....	54
7.1.2 Abuse Liability Adverse Events .....	54
7.1.3 Treatment Emergent Adverse Events .....	55
7.1.4 Solicited Adverse Events .....	55
7.1.5 Serious Adverse Events.....	55
7.2 Medical History.....	55
7.3 Physical Exam.....	55
7.4 Clinical Laboratory Evaluations .....	55
7.5 Vital Signs .....	56
7.6 Concomitant Medications .....	56
7.7 Abuse Monitoring .....	57
7.8 Pregnancies .....	57
8.0 Tables, Figures, and Listings.....	59
9.0 References.....	60
Appendix A: Plan Approval Signature Page .....	62

**VERSION HISTORY**

Version/Date	Section	Description	Author(s)
1.0, 04MAR2021	All	Initial Document	[REDACTED], [REDACTED]
2.0, 08JUN2021	Tables 2 and 5; Sections 4.3, 6.2 through 6.5, and 7.7	Updates and clarifications based on comments received from FDA	[REDACTED], [REDACTED], [REDACTED], [REDACTED]
3.0, 10FEB2022	Table 3, Sections 4.3, 4.10, 5.3, 6.2.1, 6.2.2.2, 6.3, 6.4, 7.1.4, 7.3, 7.4, 7.7, and 9.0	Updates regarding visual perceptual effects and MMRM model based on FDA and DSMB feedback, addition of subgroup analysis for Site, additional clarifications on analysis time points, deviations resulting in exclusion from per-protocol population, and summaries of abuse monitoring	[REDACTED], [REDACTED]

**TABLE OF ABBREVIATIONS**

<b>Abbreviation</b>	<b>Definition</b>
AD	Antidepressant
AE	Adverse Event
AIC	Akaike Information Criterion
ATRQ	Massachusetts General Hospital Antidepressant Treatment History Questionnaire
AUDIT	Alcohol Use Disorders Identification Test
BP	Blood Pressure
BPM	Beats Per Minute
CBC	Complete Blood Count
CGI-S	Clinical Global Impression—Severity
CI	Confidence Interval
CIOMS	Council for International Organizations of Medical Sciences
CMP	Complete Metabolic Panel
CRF	Case Report Form
C-SSRS	Columbia Suicide Severity Rating Scale
<i>d</i>	Cohen's <i>d</i>
DBP	diastolic blood pressure
DSM-5	Diagnostic and Statistical Manual of Mental Disorders - Fifth Edition
DSMB	Data and Safety Monitoring Board
DUDIT	Drug Use Disorders Identification Test
EBI	Emotional Breakthrough Inventory
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
Emmes	The Emmes Company, LLC

Abbreviation	Definition
HAM-A	Hamilton Anxiety Rating Scale
HR	Heart Rate
hs-CRP	High-Sensitivity C-Reactive Protein
ICC	Intraclass Correlation
IND	Investigational New Drug
ITT	Intent-to-Treat
LICQ	Lifetime Illness Characteristic Questionnaire
MADRS	Montgomery-Asberg Depression Rating Scale
max	maximum
MedDRA	Medical Dictionary for Regulatory Activities
MEQ-30	Mystical Experiences Questionnaire
MDD	Major Depressive Disorder
min	minimum
mmHg	Millimeters of Mercury
MMRM	Mixed-effect Model with Repeated Measures
ODQ	Oxford Depression Questionnaire
<i>p</i>	Probability
PI	Principal Investigator
PP	Per Protocol
PSIL201	A Randomized, Double-Blind Study of Single-Dose Psilocybin for Major Depressive Disorder (MDD)
PT	preferred term
Q-LES-Q	Quality of Life Enjoyment and Satisfaction Questionnaire
SAEs	Serious Adverse Events
SAP	Statistical Analysis Plan

Abbreviation	Definition
SaS	Set and Setting
SBP	Systolic Blood Pressure
SCID	Structured Clinical Interview for DSM-5
SCID-PD	Structured Clinical Interview for DSM-5 – Personality Disorder
SCID-CT	Structured Clinical Interview for DSM-5 – Clinical Trials Version
SD	Standard Deviation
SDS	Sheehan Disability Scale
SMDDS	Symptoms of Major Depressive Disorder Scale
SOC	System Organ Class
SUSAR	Suspected Unexpected Serious Adverse Reaction
TEAE	Treatment Emergent Adverse Event
TSH	Thyroid Stimulating Hormone
WOCBP	Woman of Childbearing Potential

## 1.0 INTRODUCTION

Study PSIL201, “A Randomized, Double-Blind Study of Single-Dose Psilocybin for Major Depressive Disorder (MDD)”, is a two-arm double blind randomized clinical trial designed to evaluate the potential efficacy and safety of a single 25 mg oral dose of psilocybin for MDD compared to an active placebo (niacin) in otherwise medically healthy participants between the ages of 21 and 65.

This statistical analysis plan (SAP) details analyses to be performed in support of Protocol PSIL201 and the final clinical study report for Study PSIL201, and includes a description of study design, study outcomes assessments and safety measures, participants’ characteristics, general statistical considerations and planned statistical analysis for efficacy and safety outcomes.

## 2.0 STUDY OBJECTIVES AND ENDPOINTS

### 2.1 OBJECTIVES

#### 2.1.1 PRIMARY OBJECTIVE

The primary objective of this study is to evaluate the potential efficacy of a single 25 mg oral dose of psilocybin for MDD compared to an active placebo (niacin) in otherwise medically healthy participants ages 21 to 65, assessed as the difference between groups in changes in Montgomery-Asberg Depression Rating Scale (MADRS)-assessed depressive symptoms from Baseline to Day 43 post-dose.

#### 2.1.2 SECONDARY OBJECTIVES

The key secondary objective of this study is to evaluate the potential efficacy of a single 25 mg dose of psilocybin for MDD compared to an active placebo in medically healthy participants ages 21 to 65, measured as between-group difference in change in MADRS-assessed depressive symptoms from Baseline to Day 8.

Additional secondary objectives are to evaluate between-group differences in:

- Change in Sheehan Disability Scale (SDS)-assessed functional disability status from Baseline to Day 43 post-dose
- Sustained depressive symptom response following dosing
- Sustained depressive symptom remission following dosing

#### 2.1.3 EXPLORATORY OBJECTIVES

Exploratory objectives of this study are to evaluate the potential efficacy of a single 25 mg oral dose of psilocybin for MDD compared to an active placebo in medically healthy participants between the ages of 21 and 65, measured as between-group differences in:

- Change in MADRS-assessed depressive symptoms from Baseline to post-dose Day 2, 15 and 29
- Change in Clinical Global Impression—Severity (CGI-S) score from Baseline to post-dose Day 8, 15, 29 and 43
- Change in Hamilton Anxiety Rating Scale (HAM-A)-assessed anxiety symptoms from Baseline to post-dose Day 8, 15, 29 and 43
- Change in Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q)-assessed health-related quality of life from Baseline to post-dose Day 8, 15, 29 and 43
- Change in SDS-assessed functional disability to post-dose Day 8, 15 and 29
- Degree of concordance between central-rated and computer-administered MADRS measures of depressive symptoms at Baseline and post-dose time points
- Change in Symptoms of Major Depressive Disorder Scale (SMDDS)-assessed depressive symptoms from Baseline to post-dose Day 8 and 43
- Change in SMDDS-assessed depressive symptoms from pre-dose Day 1 to post-dose Day 8 and 43

- Change in Oxford Depression Questionnaire (ODQ)-assessed depressive symptoms from Baseline to post-dose Day 43

#### **2.1.4 EXPLORATORY MECHANISM-BASED OBJECTIVES**

Exploratory mechanism-based objectives of this study are to evaluate the potential impact of acute psychological effects of a single 25 mg dose of psilocybin on longer-term antidepressant effects in MDD compared to an active placebo in medically healthy participants ages 21 to 65, measured as between group differences in:

- 30-item Mystical Experience Questionnaire (MEQ-30) score reported post-dose
- Emotional Breakthrough Inventory (EBI) score reported post-dose
- Association between MEQ-30-assessed mystical-type experiences reported post-dose and longer-term changes in MADRS-assessed depressive symptoms
- Association between EBI-assessed emotional break through experiences reported post-dose and longer-term changes in MADRS-assessed depressive symptoms

#### **2.1.5 SAFETY OBJECTIVES**

The overall safety objective of this study is to evaluate a single 25 mg oral dose of psilocybin compared to an active placebo in incidence, severity and frequency of Adverse Events (AEs), Treatment Emergent AEs (TEAEs), Solicited AEs, and Serious Adverse Events (SAEs) before, during and after the dosing session and at all follow-up visits.

- Period 1 (Screening Period): Screening through end of Baseline;
- Period 2 (Preparation Period): Preparation until Randomization;
- Period 3 (Dosing Period): Randomization through Day 9; and
- Period 4 (Follow-up Period): Day 10 through Day 43

Specific safety monitoring objectives in each of the study periods are as follows:

##### **Period 1 and Period 2:**

- Incidence of AEs by severity
- Incidence of AEs requiring medical attention
- Incidence of AEs requiring psychiatric attention
- Incidence of AEs leading to withdrawal from study
- Incidence of SAEs
- Incidence of new concomitant medications
- Incidence of self-report or urine toxicology identified illicit and non-prescribed drug use

##### **Period 3 and Period 4:**

Differences between the psilocybin and active placebo groups in:

- Incidence of AEs by severity

- Incidence of AEs requiring medical attention
- Incidence of AEs requiring psychiatric attention
- Incidence of AEs leading to termination from the study
- Incidence of TEAEs
- Incidence of TEAEs by severity
- Incidence of solicited AEs
- Incidence of solicited AEs by severity
- Incidence of SAEs
- Incidence of new concomitant medications
- Incidence of psychiatric concomitant medications
- Incidence of self-report or urine toxicology identified illicit and non-prescribed drug use
- Incidence of clinically significant abnormalities on safety laboratory assessments or physical examination

## **2.2 STUDY ENDPOINTS**

### **2.2.1 PRIMARY STUDY ENDPOINT**

1. Between-group difference in mean change of central rater Montgomery-Asberg Depression Rating Scale (MADRS) score from Baseline to post-dose Day 43

### **2.2.2 SECONDARY STUDY ENDPOINTS**

1. Between-group difference in change in central rater MADRS score from Baseline to post-dose Day 8 (Key Secondary)
2. Between-group difference in change in site rater administered SDS score from Baseline to post-dose Day 43
3. Between-group difference in sustained depressive symptom response defined as a  $\geq$  50% reduction from Baseline central rater MADRS score at the following post-dose assessments: Day 8, 15, 29 and 43
4. Between-group difference in sustained depressive symptom remission defined as a central rater MADRS total score  $\leq$  10 at the following post-dose assessments: Day 8, 15, 29 and 43

### **2.2.3 EXPLORATORY STUDY ENDPOINTS**

1. Between-group difference in change in central rater MADRS score from Baseline to Day 2, 15 and 29 post-dose
2. Between-group difference in change in central rater CGI-S score from Baseline to Day 8, 15, 29 and 43 post-dose
3. Between-group difference in change in HAM-A score from Baseline to Day 8, 15, 29 and 43 post-dose
4. Between-group difference in change in Q-LES-Q score from Baseline to Day 8, 15, 29 and 43 post-dose
5. Between-group difference in change in site rater administered SDS score from Baseline to Day 8, 15, and 29 post-dose
6. Degree of concordance between change in central rater MADRS score from Baseline to all post-dose assessments and computer-administered MADRS score over the same assessment period.

7. Between-group difference in change in SMDDS score from Baseline to Day 8 and Day 43
8. Between-group difference in change in SMDDS score from pre-dose Day 1 to post-dose Day 8 and 43
9. Between-group difference in change in ODQ score from Baseline to Day 43

#### **2.2.4 EXPLORATORY MECHANISM-BASED ENDPOINTS**

1. Between-group difference in post-dose 30-item MEQ-30 score
2. Between-group difference in post-dose EBI score

#### **2.2.5 SAFETY ENDPOINTS**

1. Relative incidence of AEs by severity
2. Relative incidence of TEAEs by severity
3. Relative incidence of SAEs by severity
4. Relative incidence of solicited AEs by severity
5. Relative incidence of concomitant medication use
6. Relative incidence of clinically significant abnormalities and changes from Baseline in clinical laboratory values
7. Relative incidence of clinically significant abnormalities and changes from Baseline in vital signs
8. Relative incidence of clinically significant physical examination findings
9. Relative incidence of self-reported or urine toxicology identified illicit and non-prescribed drug use

### 3.0 STUDY DESIGN

#### 3.1 OVERALL STUDY PLAN AND DESIGN

PSIL201 is a randomized two-arm double-blind Phase II study designed to evaluate potential efficacy of a single 25 mg oral dose of psilocybin for MDD compared to a single 100 mg oral dose of active placebo (niacin). Approximately 100 participants (males and females) ages 21 to 65 who, at Screening, meet DSM-5 criteria for MDD with a current depressive episode of at least a 60-day duration with a Screening MADRS score  $\geq 28$  and who meet all other inclusion/exclusion criteria at Baseline will be randomized with a 1:1 allocation under double-blind conditions to receive a single 25 mg oral dose of psilocybin or a single 100 mg oral dose of niacin.

Participants deemed eligible following successful completion of all screening assessments will complete central rater, site rater and self-report measures at Baseline for a final eligibility determination. Eligible participants at Baseline will undergo preparation sessions and be eligible for randomization on Dosing Day to receive either psilocybin or niacin active-placebo and will complete follow-up visits and assessments on study Day 2, 8, 15, 29 and 43 (within corresponding visit windows). Study outcome measures will assess depressive symptoms, clinical global functioning, functional disability, anxiety symptoms and health-related quality of life. Safety outcome measures will be collected at all assessment time points from the time of consent through the end of study.

To enhance participant safety, the current study does not propose to test psilocybin as a “context-less” pharmacological agent, but rather within a “set and setting” (SaS) protocol similar to the protocol that has been used in all modern studies of psilocybin in both diseased and normal healthy populations. The SaS protocol for this study includes: 1) a period of preparation with session Facilitators prior to dosing; 2) administration of study medications in an aesthetically pleasing room under the supervision of two Facilitators who are present throughout the session (with the exception of short, temporary allowances for facilitator breaks; e.g. bathroom breaks); and 3) three post-dose integration sessions during which participants are encouraged to discuss their intervention experience with the Facilitators. The SaS will be identical for those randomized to psilocybin or niacin active placebo.

A diagram of the study design and schedule of assessments are provided in the study protocol.

#### 3.2 SELECTION OF STUDY POPULATION

The planned study population for this clinical trial is approximately 100 otherwise medically healthy adults ages 21 to 65 who meet DSM-5 criteria for a diagnosis of MDD and are currently experiencing a major depressive episode of at least a 60-day duration at the time of the Screening.

Only participants who meet depressive symptom severity criteria and who do not show an unacceptably large degree of symptom improvement between the Screening visit and the Baseline assessment (indexed by  $\leq 30\%$  improvement on both the central rater and computer MADRS assessments) will be eligible for randomization. In addition, at Screening and Baseline,

only participants with a score difference  $\leq 7$  between the central rater and computer MADRS assessments will be eligible for randomization.

Potential participants who meet all other entry criteria at Screening but who are taking an antidepressant or an antidepressant plus an augmenting agent (e.g., a second antidepressant, an atypical antipsychotic, lithium) will be eligible for continuation in the study but will enter a medication taper during which current psychotropic medications will be withdrawn under the supervision of a study psychiatric medical provider. Participants will be eligible to undergo the Baseline assessment at least 2 weeks after the last dose of the applicable medication.

Detailed inclusion and exclusion criteria are provided in the study protocol.

### **3.3 EFFICACY AND SAFETY OUTCOMES**

#### **3.3.1 EFFICACY OUTCOMES**

Table 1 provides a summary of the efficacy outcome variables and associated assessment tool.

Table 2 describes the scoring method for each of the efficacy outcome measures. For additional details regarding assessment administration refer to the study protocol.

#### **3.3.2 SAFETY OUTCOMES**

Safety outcomes are presented in Table 3.

Adverse events, serious adverse events, treatment emergent adverse events, and Suspected Unexpected Serious Adverse Reactions are defined in the study protocol. AEs will be graded for intensity and relationship to study product. Elevated blood pressure and heart rate will only be reported as an AE should medication be needed; substance and alcohol abuse/misuse and substance and alcohol use disorder will only be reported should they result in a SAE.

**TABLE 1. EFFICACY OUTCOME MEASURES**

Outcomes/Endpoints	Measure/Score	Study Days	Administration
<b>Primary</b>			
Difference between groups in changes in central rater MADRS-assessed depressive symptoms from Baseline to Day 43 post-dose	MADRS	Baseline and post-dose Day 43	Independent central rating by telephone
<b>Secondary</b>			
Change in central-rater MADRS-assessed depressive symptoms from Baseline to Day 8 post-dose (Key Secondary)	MADRS	Baseline and post-dose Day 8	Independent central rating by telephone
Change in functional disability status from Baseline to Day 43 post-dose	SDS	Baseline and post-dose Day 43	Site rater administration
Sustained depressive symptom response following dosing	MADRS	Baseline and post-dose Day 8, 15, 29 and 43	Independent central rating by telephone
Sustained depressive symptom remission following dosing	MADRS	Baseline and post-dose Day 8, 15, 29 and 43	Independent central rating by telephone
<b>Exploratory</b>			
Change in MADRS-assessed depressive symptoms from Baseline to post-dose Day 15 and 29	MADRS	Baseline and post-dose Day 15 and 29	Independent central rating by telephone
Change in clinical global impression from Baseline to post-dose Day 8, 15, 29 and 43	CGI-S	Baseline and post-dose Day 8, 15, 29 and 43	Independent central rating by telephone
Change in anxiety symptoms from Baseline to post-dose Day 8, 15, 29 and 43	HAM-A	Baseline and post-dose Day 8, 15, 29 and 43	Site rater administration
Change in health-related quality of life from Baseline to post-dose Day 8, 15, 29 and 43	Q-LES-Q	Baseline and post-dose Day 8, 15, 29 and 43	Self-report collected electronically at the site

Outcomes/Endpoints	Measure/Score	Study Days	Administration
Change in functional disability to post-dose Day 8, 15 and 29	SDS	Baseline and post-dose Day 8, 15, and 29	Site rater administration
Degree of concordance between central-rated and computer-administered MADRS measures of depressive symptoms at Baseline and post-dose time points	MADRS	Baseline and post-dose Day 8, 15, 29 and 43	Independent central rating by telephone and computer-administered
Change in Symptoms of Major Depressive Disorder Scale (SMDDS)-assessed depressive symptoms from Baseline to post-dose Day 8 and 43	SMDDS	Baseline and post-dose Day 8 and Day 43	Self-report collected electronically at the site
Change in Symptoms of Major Depressive Disorder Scale (SMDDS)-assessed depressive symptoms from Baseline to post-dose Day 8 and 43	SMDDS	Pre-Dose Day 1 and post-dose Day 8 and Day 43	Self-report collected electronically at the site
Change in Oxford Depression Questionnaire (ODQ)-assessed depressive symptoms from Baseline to post-dose Day 43	ODQ	Baseline and post-dose Day 43	Self-report collected electronically at the site
<b>Exploratory Mechanism-Based Objectives</b>			
Mystical-type experiences reported post-dose	MEQ-30	Post-dose Day 1	Self-report collected electronically at the site
Emotional breakthrough experiences reported post-dose	EBI	Post-dose Day 1	Self-report collected electronically at the site

Outcomes/Endpoints	Measure/Score	Study Days	Administration
Association between mystical-type experiences reported post-dose and longer-term changes in MADRS-assessed depressive symptoms	MEQ-30; MADRS	Post-dose Day 1 (MEQ-30); Baseline and post-dose Day 8, 15, 29, and 43 (MADRS)	Self-report collected electronically at the site (MEQ-30); independent central rating by telephone (MADRS)
Association between emotional break through experiences reported post-dose and longer-term changes in MADRS-assessed depressive symptoms	EBI; MADRS	Post-dose Day 1 (MEQ-30); Baseline and post-dose Day 8, 15, 29, and 43 (MADRS)	Self-report collected electronically at the site (EBI); independent central rating by telephone (MADRS)

**TABLE 2. SCORING METHODS FOR EFFICACY OUTCOMES**

Assessment Tool	Scoring Method <sup>1</sup>	Range of Possible Values
MADRS	Total score: Summed score of all 10 items	0 - 60
SDS <sup>2</sup>	<p>Mean score: Mean of work/school, social life, and family/home responsibilities items. For each of the three items, scores range from 0 through 10. The number most representative of how much each area was disrupted by symptoms is marked along the line from 0 = not at all, to 10 = extremely. For the work/school item, if the participant has not worked for reasons <i>unrelated</i> to the disorder, a checkbox with a response "I have not worked/studied at all during the past week for reasons unrelated to this disorder" will be selected and no score for that item is provided.</p> <p>When item scores for all items are present, the Mean score will be calculated as the mean of all three items. If the checkbox is selected for the work/school item, the Mean score will be calculated as the mean of the social life and family/home responsibilities items.</p>	0 - 10
HAM-A	Total score: Summed score of the 14 items that assess anxious mood, tension, fear, insomnia, intellectual (cognitive) symptoms, depressed mood, somatic (sensory), cardiovascular, respiratory, gastrointestinal, genitourinary, autonomic and somatic (muscular) symptoms, and behavior at interview. Each item is scored using a 0-4 Likert scale. (Hamilton 1959)	0 - 56
CGI-S	Rater assessed single 7-point scale	1 - 7
Q-LES-Q	Total score: Summed score of the first 14 items on the scale, with the last two items serving as stand-alone queries	14 - 70
SMDDS	Total score: Summed score of all 16 items	0 - 60
ODQ	<p>Total score: Summed score of all 26 items</p> <p>Antidepressant as Cause (AC) dimension score: Sum of questions 21 - 26</p>	<p>Total score: 26 - 130</p> <p>AC dimension score: 6 - 30</p>

Assessment Tool	Scoring Method <sup>1</sup>	Range of Possible Values
MEQ-30	<p>Total score: Sum of 30 items, 4 component scales are the average of items contributing that component:</p> <ul style="list-style-type: none"> <li>• Mystical (items 4, 5, 6, 9, 14, 15, 16, 18, 20, 21, 23-26, 28)</li> <li>• Positive mood (items 2, 8, 12, 17, 27, 30)</li> <li>• Transcendence of time and space (items 1, 7, 11, 13, 19, 22)</li> <li>• Ineffability (items 3, 10, 29)</li> </ul>	Total score: 0 to 150
EBI	Total score: Each question will be asked using a Visual Analog Scale (VAS) scale of 0-100, and the total score will be the average of the six questions.	Total score: 0 - 100

<sup>1</sup>Where applicable, component or dimension scores will be reported as secondary exploratory analyses in addition to total scores.

<sup>2</sup>The functionality of the Signant Health tablet used for collection of electronic patient reported outcomes requires a response to either the work/school item score OR the checkbox indicating that the participant did not work/go to school for reasons unrelated to the disorder; therefore, missing data on this item is not expected. If participants have not worked/gone to school for reasons *related* to the disorder, they are instructed to provide a score for the work/school item.

**TABLE 3. SAFETY OUTCOME MEASURES**

Outcome	Definition	Study Period/ Study Day
<b>Adverse Events</b>		
AE	Any untoward medical occurrence in a research participant, whether or not considered drug related which occurs during the conduct of the clinical trial. Any change in clinical status, ECGs, routine labs, physical examinations, etc., that is considered clinically significant by the PI is considered an AE.	Periods 1-4
SAE	<p>An SAE or serious suspected adverse reaction or serious adverse reaction as determined by the PI or the Sponsor is any event that results in any of the following outcomes:</p> <ol style="list-style-type: none"> <li>1. Death</li> <li>2. Life-threatening AE (Life-threatening means that the study participant was, in the opinion of the PI or Sponsor, at immediate risk of death from the reaction as it occurred.)</li> <li>3. Inpatient hospitalization or prolongation of existing hospitalization</li> <li>4. Persistent or significant incapacity or substantial disruption of the ability to conduct normal life functions</li> <li>5. Congenital abnormality or birth defect</li> <li>6. Important medical event that may not result in one of the above outcomes but may jeopardize the health of the study participant or require medical or surgical intervention to prevent one of the outcomes listed in the above definition of serious event.</li> </ol>	Periods 1-4
TEAE	Any AE not present prior to the initiation of treatment or any event already present that worsens in either intensity or frequency following exposure to the study drug/treatment.	Periods 3 and 4
Suspected Unexpected Serious Adverse Reaction (SUSAR)	An AE that is serious, there is a certain degree of probability that the event is a harmful and an undesirable reaction to the medicinal product under investigation, regardless of the administered dose, and the adverse reaction is unexpected.	Periods 3 and 4

Outcome	Definition	Study Period/ Study Day
Abuse Liability AEs	<p>AEs with the following MedDRA Preferred Terms:</p> <ul style="list-style-type: none"> <li>Euphoria-related terms: euphoric mood, elevated mood, feeling abnormal, feeling drunk, feeling of relaxation, dizziness, thinking abnormal, hallucination, inappropriate affect</li> <li>Terms of impaired attention, cognition, and mood: somnolence, mood disorders, and disturbances</li> <li>Dissociative/psychotic terms: psychosis, aggression, confusion, and disorientation</li> </ul>	Periods 1-4
Cardiovascular Proarrhythmic AEs	<p>AEs with the following MedDRA Preferred Terms:</p> <ul style="list-style-type: none"> <li>Torsade de pointes</li> <li>Sudden death</li> <li>Ventricular tachycardia</li> <li>Ventricular fibrillation and flutter</li> <li>Syncope</li> <li>Seizures</li> </ul>	Periods 1-4
<b>Solicited Adverse Events</b>		
Visual perceptual effects	<ol style="list-style-type: none"> <li>Uncontrolled or disturbing study drug effects</li> <li>Visual distortions (e.g. geometric hallucinations, false perceptions of movement in the peripheral visual fields, flashes of color, intensified color, trails for images of moving objects, positive after-images, halos around objects)</li> </ol> <p>Visual perceptual effects are solicited on the day of dosing for effects occurring during the dosing session, and at post-dose visits for return of VP effects since the dosing session. The specific questions asked on the day of dosing following the dosing session and at post-dose study visits are provided in the study protocol.</p>	Periods 3 and 4
Active suicidal Ideation	C-SSRS and/or MADRS and verified by clinical assessment	Period 3 and 4 <sup>a</sup>
Headache	Participants will be asked if they have or experienced a headache. These will be classified as a migraine, tension headache, or other headache.	Period 3

Outcome	Definition	Study Period/ Study Day
Nausea		Period 3
Elevated blood pressure (BP)	Systolic blood pressure (SBP) >140 mmHg or diastolic blood pressure (DBP) >90 mmHg on three separate readings and requiring medication	Period 3 and 4 <sup>a</sup>
Elevated heart rate (HR)	HR >100 beats per minute (BPM) and requiring medication	Period 3 and 4 <sup>a</sup>
Drug overdose with suicidal intent	Drug overdose with intention of suicide	Period 3 and 4
<b>Alcohol and Substance Abuse Monitoring</b>		
Alcohol abuse/misuse	Total score on AUDIT of $\geq 8$ for men or $\geq 6$ for women	Screening and post-dose Day 43 <sup>b</sup>
Drug abuse/misuse	Total score on DUDIT of $\geq 8$ for men or $\geq 6$ for women	Screening and post-dose Day 43 <sup>b</sup>
Substance and alcohol use disorders	Based on SCID-CT Alcohol use disorder: answer of Yes to "At least two alcohol use items coded + and items occurred during the last 12 months" Non-alcohol substance use disorder: answer of Yes to "At least two substance use items are coded + and items occurred during the last 12 months"	Screening and post-dose Day 43 <sup>b</sup>
<b>Additional Safety Assessments</b>		
Medical history	Self-reported medical history and psychiatric history collected via SCID-PD and SCID-CT to confirm eligibility	Screening
Physical examination	Physical exams will be conducted at Screening and Baseline to confirm eligibility for the study and at Day 8 and Day 43 visits to examine whether persistent health related physical changes are observed post-dosing. Additional physical exams may be administered depending on participant's signs and symptoms at the discretion of the investigator.	Complete physical exam at Screening, Baseline, Day 8, and Day 43
Laboratory assessments	Complete Metabolic Panel (CMP), Complete Blood Count with Differential (CBC w/ Diff), Thyroid Stimulating Hormone (TSH), high-sensitivity C-Reactive Protein (hs-CRP), INR, Urinalysis, Urine drug screen, Urine pregnancy test.	Screening, Baseline, and post-dose Day 2, Day 8 and Day 43

Outcome	Definition	Study Period/ Study Day
Vital signs	BP, HR, and Temp	Screening, Baseline, Day 1 prior to dosing, post-dose Day 8, 15, 29 and Day 43 BP and HR only: post- dose 30, 60, 90, 120 minutes, 4, 6, and 7 hours after drug administration
Concomitant medications	Self-reported medication use collected at all study visits	Periods 1 - 4

<sup>a</sup> Recorded for Periods 1-4; may result in screen failure or early termination in Periods 1-2.

<sup>b</sup> Evidence of any substance or alcohol use will not be reported as an AE. Should any substance or alcohol use be considered a SAE, it will be reported on a SAE form in addition to being captured through the assessments described above.

## 4.0 GENERAL STATISTICAL CONSIDERATIONS

### 4.1 GENERAL PRINCIPLES

Unless otherwise noted, continuous variables will be summarized using the following descriptive statistics: N (non-missing sample size), mean, standard deviation (SD), median, maximum (max) and minimum (min). Categorical variables will be summarized using counts/frequencies and percentages. Denominators used for percentages will be clearly indicated in the table. For a percentage within each group, an exact two-sided 95% CI will be provided using Clopper-Pearson method, unless noted otherwise.

Normality assumptions for continuous variables will be examined visually using QQ-plots and residual plots. Variable transformation using natural logarithm or rank transformation will be applied in case of nonnormality. Should normality not be obtained through transformation of the data, alternative non-parametric methods (such as Wilcoxon rank sum test) will be applied to the non-transformed data.

### 4.2 TIMING OF ANALYSES

All primary, secondary and exploratory analyses will be performed after database lock. The DSMB will review accumulating data as described in Section 4.14 and the DSMB Charter.

### 4.3 ANALYSIS TIME POINTS

Baseline refers to measures and assessments collected at the Baseline visit (Visit number 00B or 00BZ for repeat baseline visits). For participants who completed assessments at both the initial and repeat baseline visit, scores for the assessments from the repeat baseline visit will be used for analyses.

For participants who completed assessments at both the initial and repeat screening visits, scores and results from the repeat screening will be used for all summaries. Scores and results from both initial and repeat screening and baseline visits will be included in the Listings.

The following Periods will be used for Safety summaries:

- Period 1 (Screening Period): Screening through end of Baseline;
- Period 2 (Preparation Period): Preparation until Randomization;
- Period 3 (Dosing Period): Randomization through Day 9;
- Period 4 (Follow-up Period): Day 10 through Day 43.

### 4.4 ANALYSIS POPULATIONS

#### 4.4.1 ITT POPULATION

Analyses of primary, secondary and exploratory efficacy outcomes will be conducted on an intent-to-treat (ITT) population that will include all randomized participants. Participants will be analyzed according to the treatment to which they were randomized.

#### 4.4.2 PER PROTOCOL POPULATION

The Per Protocol (PP) population will be a subset of the ITT population who meet the following additional criteria:

- Received investigational product
- Completed post-dose Day 8 and Day 43 central-rater MADRS assessment
- No major protocol deviations that may affect the primary or key secondary outcome measures (see Section 5.3 for definition of major deviations)

Participants will be analyzed according to the arm to which they were randomized.

The decision to exclude a participant from per protocol population or a participant's data from the per protocol analyses will be made by blinded members of the study team.

#### 4.4.3 SAFETY SETS

Populations for analysis of safety data will be defined as follows:

- **All Enrolled Population:** All participants who have signed written informed consent. The All Enrolled population will be used for summaries of safety data collected during Study Periods 1 and 2.
- **Preparation Phase Termination Population:** Participants who discontinue study participation between the end of the Baseline assessments and up to randomization. Adverse events will be analyzed among those who go through the screening, baseline and preparatory phase but are ultimately found not to be eligible to investigate whether being discontinued in the study immediately prior to randomization is associated with an increase in depression-related adverse events.
- **Safety Population:** Participants who were randomized and received study product. The safety population will be used for analysis of safety data collected during Study Periods 3 and 4. Participants will be analyzed according to the intervention they received.

#### 4.5 RANDOMIZATION AND BLINDING

Eligible participants will be randomly assigned in a 1:1 ratio to the psilocybin group or the active control group. Emmes unblinded statisticians will generate treatment assignments using randomized blocks with varying block sizes per the Randomization Plan. Treatment assignments will be stratified by site. Treatment allocation will be conducted centrally by Emmes via eClinical.

Participants and investigators along with the entire study staff (i.e. Usona and all vendors to whom responsibilities have been delegated) will be blinded to treatment assignments, with the exception of the specific unblinded study staff as indicated in the PSIL201 Blinding Management Plan. To minimize chances of site personnel guessing treatment assignments and thereby potentially biasing outcome measures, the following measures will be taken: The study's primary outcome assessment (MADRS) will be conducted by remote blinded central raters who will have no interactions with either participants or study personnel other than conducting the MADRS assessments, which will be conducted by telephone, as is standard practice when central raters are used in pharmaceutical trials. These raters will have no information regarding

specifics of the study design and will not know where in the study protocol any given participant is when assessed. In addition, CGI-S will be assessed by the same blinded central raters who assessed MADRS. The intervention session facilitators who will be in the room with the participants throughout the intervention session will not have contact with the remote central raters.

Full blinding of study personnel, the Sponsor and participants (except as noted in the Blinding Management Plan) will be maintained until data lock at the conclusion of the study. Exceptions on an individual subject basis will be made in cases in which it is determined by site PIs, the Medical Monitor or the Sponsor that unblinding of a participant's intervention assignment is required for participant safety. Details of study blinding are provided in the PSIL201 Study Blinding Plan.

## 4.6 SAMPLE SIZE AND POWER CONSIDERATIONS

This study has been powered to evaluate the clinical efficacy of psilocybin for the primary and key secondary objectives, i.e. to test the difference between psilocybin and niacin in the change in central rater MADRS score from Baseline to Day 43 and Day 8 respectively, in the ITT analysis set.

Approximately 100 participants are planned to be randomized into this study. This sample size was selected based on the following assumptions:

- Psilocybin group MADRS mean change from baseline:
  - Day 8 = 18
  - Day 43: 17
- Placebo group MADRS mean change from baseline for all time points = 10
- MADRS change from baseline SD: 10 in both groups at each time point
- Dropout of 5% at Day 8 and additional 7.5% at Day 43
- $\alpha = 0.05$  for the primary and key secondary endpoints

The rationale for these sample size assumptions is discussed in Protocol Section 9.2.1.

Simulations were generated to estimate power for a primary endpoint of MADRS change from baseline to Day 43 and key secondary endpoint of MADRS change from baseline to Day 8 assuming a MMRM model as specified in Section 6.2.2 using the above model assumptions. A sample size of 100 participants would result in 92% power for primary Day 43 endpoint and 98% power for the key secondary Day 8 endpoint.

To explore the impact of deviations from the assumptions above, the estimated power at Day 43 and Day 8 assuming a sample size of  $N = 100$  was generated for the following scenarios:

- Psilocybin group MADRS mean change from baseline:
  - Day 8 = 18
  - Day 43: Considered range of 16-18
- Placebo group MADRS mean change from baseline for all time points = 10
- MADRS change from baseline SD: 9 - 12 in both groups at each time point

- Correlation among repeated measures  $\rho = 0.5$  across all time points
- Dropout of 5% at Day 8 and additional 2.5%, 5%, or 7.5% at Day 43
- $\alpha = 0.05$  for the primary and key secondary endpoints

Table 4 below summarizes the estimated Power for the Day 43 endpoint and Day 8 endpoint for N = 100 and assumptions described above.

**TABLE 4. SUMMARY OF POWER AT DAY 43 (PRIMARY ENDPOINT) AND DAY 8 (KEY SECONDARY ENDPOINT) FOR N = 100**

Model Parameters					
Psilocybin Grp Mean MADRS Change from BL				Actual Power	
Day 8	Day 43	MADRS Change from BL SD	Drop out by Day 43 <sup>a</sup>	Day 43	Day 8 <sup>b</sup>
18	16	9	2.5%	0.89	1.00
18	16	9	5.0%	0.89	1.00
18	16	9	7.5%	0.88	1.00
18	16	10	2.5%	0.82	0.99
18	16	10	5.0%	0.82	0.99
18	16	10	7.5%	0.81	0.99
18	16	11	2.5%	0.76	0.97
18	16	11	5.0%	0.75	0.98
18	16	11	7.5%	0.75	0.98
18	16	12	2.5%	0.69	0.94
18	16	12	5.0%	0.69	0.95
18	16	12	7.5%	0.68	0.95
18	17	9	2.5%	0.97	0.99
18	17	9	5.0%	0.96	0.99
18	17	9	7.5%	0.96	0.99
18	17	10	2.5%	0.93	0.98
18	17	10	5.0%	0.92	0.98
18	17	10	7.5%	0.92	0.98
18	17	11	2.5%	0.86	0.96
18	17	11	5.0%	0.86	0.96
18	17	11	7.5%	0.85	0.97
18	17	12	2.5%	0.80	0.93
18	17	12	5.0%	0.79	0.93

Model Parameters					
Psilocybin Grp Mean MADRS Change from BL				Actual Power	
Day 8	Day 43	MADRS Change from BL SD	Drop out by Day 43 <sup>a</sup>	Day 43	Day 8 <sup>b</sup>
18	17	12	7.5%	0.78	0.93
18	18	9	2.5%	0.99	0.99
18	18	9	5.0%	0.99	0.99
18	18	9	7.5%	0.99	0.99
18	18	10	2.5%	0.96	0.98
18	18	10	5.0%	0.96	0.98
18	18	10	7.5%	0.96	0.98
18	18	11	2.5%	0.93	0.95
18	18	11	5.0%	0.93	0.95
18	18	11	7.5%	0.93	0.95
18	18	12	2.5%	0.89	0.93
18	18	12	5.0%	0.88	0.93
18	18	12	7.5%	0.88	0.93

For all scenarios explored, power at Day 43 is less than power at Day 8, so it is sufficient to ensure that the Day 43 power for the primary endpoint meets the desired threshold. The power at Day 43 for the above scenarios ranges from 68% for a change from baseline at day 43 of 16, SD of 12 and drop out at Day 43 of 7.5% to 99% for a change from baseline at day 43 of 18, SD of 9 and drop out at Day 43 of 2.5%.

## 4.7 EARLY TERMINATION AND DROPOUTS

Participants who discontinue study participation after the Baseline visit but prior to randomization will be considered Preparation Phase Terminations and will complete an End of Study visit within 14 days. These participants will not be included in the ITT population but will be included in the safety All Enrolled and Preparation Phase Termination populations.

If a participant is randomized but chooses to not receive treatment, he/she will be encouraged to continue with all scheduled follow-up assessments. Data from these participants will be included in the ITT population and All Enrolled safety population. These participants will not be replaced. The frequency of this event occurring is expected to be minimal given that randomization occurs on the day of dosing.

Upon consultation with the study Sponsor as needed, if a site PI feels that for any reason, continuing participation would negatively impact safety should a participant remain in the study,

he/she will be asked to complete the End of Study visit within 14 days. Efforts will be made to obtain information about AE outcomes, if deemed necessary by the PI, Medical Monitor and/or Sponsor.

For any other issue post-dosing not judged to impact participant safety, participants will be encouraged to remain in the study and continue with all follow up assessments, given that no additional exposure to the study medication is required following the dosing day. If a participant does not agree to remain in the study, or if a participant chooses to withdraw from the study for any other reason post-dosing, he/she will be asked to complete the End of Study visit assessments within 43 days of their dosing day, so that the maximal time between dosing and assessment is no longer than the time period for participants who complete the study.

Dropouts are participants who withdraw consent to participate in the study or are lost to follow-up. If they withdraw consent through the Baseline visit, they are considered screen failures. If they withdraw consent after Baseline but prior to randomization, they are considered preparation phase terminations. If they withdraw consent following randomization, they are considered early terminations.

A participant will be considered lost to follow-up if he or she repeatedly fails to return for scheduled visits post dosing and is unable to be contacted by the study site. Specific actions for attempting to contact the participants as specified in the study protocol. Should the participant continue to be unreachable, he/she will be considered to have withdrawn from the study.

## **4.8 MISSING DATA**

Every effort will be made to minimize incomplete data. For example, for primary and secondary endpoints, the Signant Health Rater Station is configured such that items within an assessment are generally not allowed to be skipped, limiting the possibility of missing items that would preclude calculation of total and component/dimension scores.

To minimize the impact of post-randomization early termination, participants who receive a study intervention but miss the Day 8 post-dose assessment for any reason other than study discontinuation will be encouraged to continue in the study and all other subsequent assessment points will be collected to allow the participant's data to be included in analyses of secondary endpoints, as well as the safety database.

In alignment with ICH E9 (R1), a treatment policy strategy as defined in Section 6.2.2.1 the primary analysis will be carried out on the ITT population, without imputation of missing data points. Missing data will be handled using a maximum likelihood approach assuming data are missing at random (MAR). Sensitivity and supplementary analyses as specified in Section 6.2.2.3, including both those that assume MAR and those that assume missing not at random (MNAR) will be performed to evaluate robustness of the results of the primary analysis model.

### **4.8.1 MISSING DATES**

Conventions for imputation of missing dates will be consistent with the methods described in the Advantage eClinical® eCRF completion guidelines.

## 4.9 COVARIATES

Repeated measures analyses of primary, secondary and exploratory outcomes will be adjusted for baseline central-rater MADRS total score, study site, sex, and treatment-resistant depression status.

## 4.10 SUBGROUP ANALYSES

Several exploratory subgroup analyses are planned to investigate whether there is initial evidence that certain subgroups respond better to psilocybin intervention than others and to refine the study population of interest for future studies. Specifically, subgroup analyses will be conducted comparing treatment effects for:

- 1) Participants with TRD at enrollment vs. those without
- 2) Participants who completed a medication taper prior to study enrollment vs. those who did not
- 3) Age (<45 vs. ≥45)
- 4) Baseline MADRS (≤ median vs. > median)
- 5) Site

Subgroup analyses will be conducted for primary, secondary, and safety outcomes including related TEAEs, SAEs, and solicited AEs.

For continuous endpoints, a treatment by subgroup interaction term will be added to primary and secondary MMRM to test differences in treatment effect by subgroup (e.g. TRD participants vs. no TRD participants or medication taper participants vs. no medication taper participants), and treatment effects within each subgroup will be estimated using the LSMESTIMATE statement in PROC MIXED.

For binary endpoints, a Breslow-Day test will be used to test homogeneity of the treatment effect between subgroups.

Subgroup analyses for primary, secondary efficacy outcomes will be performed on the ITT population and safety outcomes will use the Safety Population. Testing will be performed without adjustment to the alpha level given the exploratory nature of the analyses.

The treatment effect for the primary and key secondary outcomes stratified by site will be visualized using forest plots.

## 4.11 SIGNIFICANCE TESTING AND MULTIPLE COMPARISONS

All tests will be two-tailed with an alpha level set at  $p < 0.05$ . P-values will be rounded to 3 decimal places.

To mitigate the risk of alpha inflation, a sequential significance testing procedure will be employed for primary and secondary hypotheses as specified in Section 6.2.2.2.

## 4.12 INTERIM ANALYSES

No formal statistical interim analyses are planned for this study protocol. Interim reports to review safety outcomes will be presented at the regular DSMB meetings or unscheduled times per the DSMB's request. Efficacy data will only be included if requested by the DSMB and would be presented only during the closed DSMB sessions. Reports including efficacy data will be prepared and reviewed by unblinded staff and only shared with DSMB members. Any efficacy summaries will be presented without any formal statistical testing.

## 4.13 STOPPING CRITERIA

No formal stopping criteria are planned for this study protocol.

## 4.14 DATA MONITORING

Study progress and safety reports will be reviewed regularly by a DSMB. The DSMB will review actual and projected accrual, patient demographics, and safety outcomes including a summary of the number and seriousness of AEs, cumulative reports of SAEs requiring expedited reporting, new SAEs, and other safety outcomes as specified in Table 3. Additional details are specified in the DSMB Charter.

## 4.15 SOFTWARE USED FOR ANALYSIS

Unless otherwise specified, all analyses will be performed using SAS version 9.4 or above.

## 4.16 REPORTING CONVENTIONS

P-values  $\geq 0.001$  and  $\leq 0.999$  will be reported to 3 decimal places; p-values less than 0.001 will be reported as " $<0.001$ " and p-values greater than 0.999 will be reported as " $>0.999$ ." The mean, standard deviation, and other statistics will be reported to 1 decimal place greater than the original data. The minimum and maximum will use the same number of decimal places as the original data. Proportions will be presented as 2 decimal places; values greater than zero but  $<0.01$  will be presented as " $<0.01$ ". Percentages will be reported to the nearest whole number; values greater than zero but  $<1\%$  will be presented as " $<1\%$ "; values greater than 99% but less than 100% will be reported as  $>99\%$ . Estimated parameters, not on the same scale as raw observations (e.g. regression coefficients) will be reported to 3 significant figures.

## 4.17 DERIVED VARIABLE DEFINITIONS

### 4.17.1 TREATMENT-RESISTANT DEPRESSION

TRD will be assessed for the current depressive episode based on participant response to the Massachusetts General Hospital Antidepressant Treatment History Questionnaire collected at the Screening visit. TRD will be defined by meeting all the following criteria: 1) participant report of receiving treatment with at least two antidepressant medications (or one antidepressant with at least one augmenting agent) for this current depressive episode for at least 8 weeks, 2) dose of medication is equal to or greater than minimally adequate dose, and 3) response to these medications is  $<50\%$  improvement.

#### **4.17.2 SUSTAINED DEPRESSIVE SYMPTOM RESPONSE**

Sustained depressive symptom response will be defined as present if a  $\geq 50\%$  reduction from Baseline central rater MADRS score is observed at the following post-dose assessments: Day 8, 15, 29 and 43. Non-response will be defined as  $< 50\%$  reduction from Baseline central rater MADRS score observed at post-dose assessments at Day 8, 15, 29, or 43. Subjects with missing central rater MADRS scores at any post-dose assessment will be excluded from the main analysis of this endpoint. A supplemental analysis including participants with missing data due to lack of efficacy as a non-response will also be performed. See Section 6.6 for additional details.

#### **4.17.3 SUSTAINED DEPRESSIVE SYMPTOM REMISSION**

Sustained depressive symptom remission will be defined as a central rater MADRS total score  $\leq 10$  at the following post-dose assessments: Day 8, 15, 29 and 43. Non-remission will be defined as central rater MADRS score  $> 10$  at any post-dose assessment at Day 8, 15, 29 or 43. Subjects with missing central rater MADRS scores at any post-dose assessment will be excluded from the main analysis of this endpoint. A supplemental analysis including participants with missing data due to lack of efficacy as a non-remission will also be performed. See Section 6.6 for additional details.

## 5.0 STUDY PARTICIPANTS

### 5.1 ANALYSIS OF PARTICIPANT CHARACTERISTICS

Participant demographic information including sex, gender, ethnicity, race, education, income, marital status, and employment status will be summarized with counts and percentages and continuous variables including age (years), height (cm), weight (kg) and BMI (kg/m<sup>2</sup>) at screening, will be summarized by n (non-missing sample size), mean, standard deviation, median, maximum and minimum will be presented by treatment group for each of the study populations.

Sex will be categorized as male, female, unknown, and undifferentiated, and gender will be categorized as female, male, unknown, unspecified, not reported, or other. Ethnicity will be categorized as Hispanic or Latino, not Hispanic or Latino, not reported, and unknown. Race categories collected include American Indian or Alaskan Native, Asian, Native Hawaiian or other Pacific Islander, Black or African American, and White. Participants may self-designate as belonging to more than one race or may refuse to identify a race. Additional baseline characteristics including vital signs, medical history, concomitant medications, and physical exam (see Section 7) will be summarized by treatment group for each of the study populations.

### 5.2 PARTICIPANT DISPOSITION

A CONSORT diagram displaying the disposition of study participants will be included and will present the number of participants screened, enrolled, randomized, lost to follow-up, and analyzed (Figure 1).

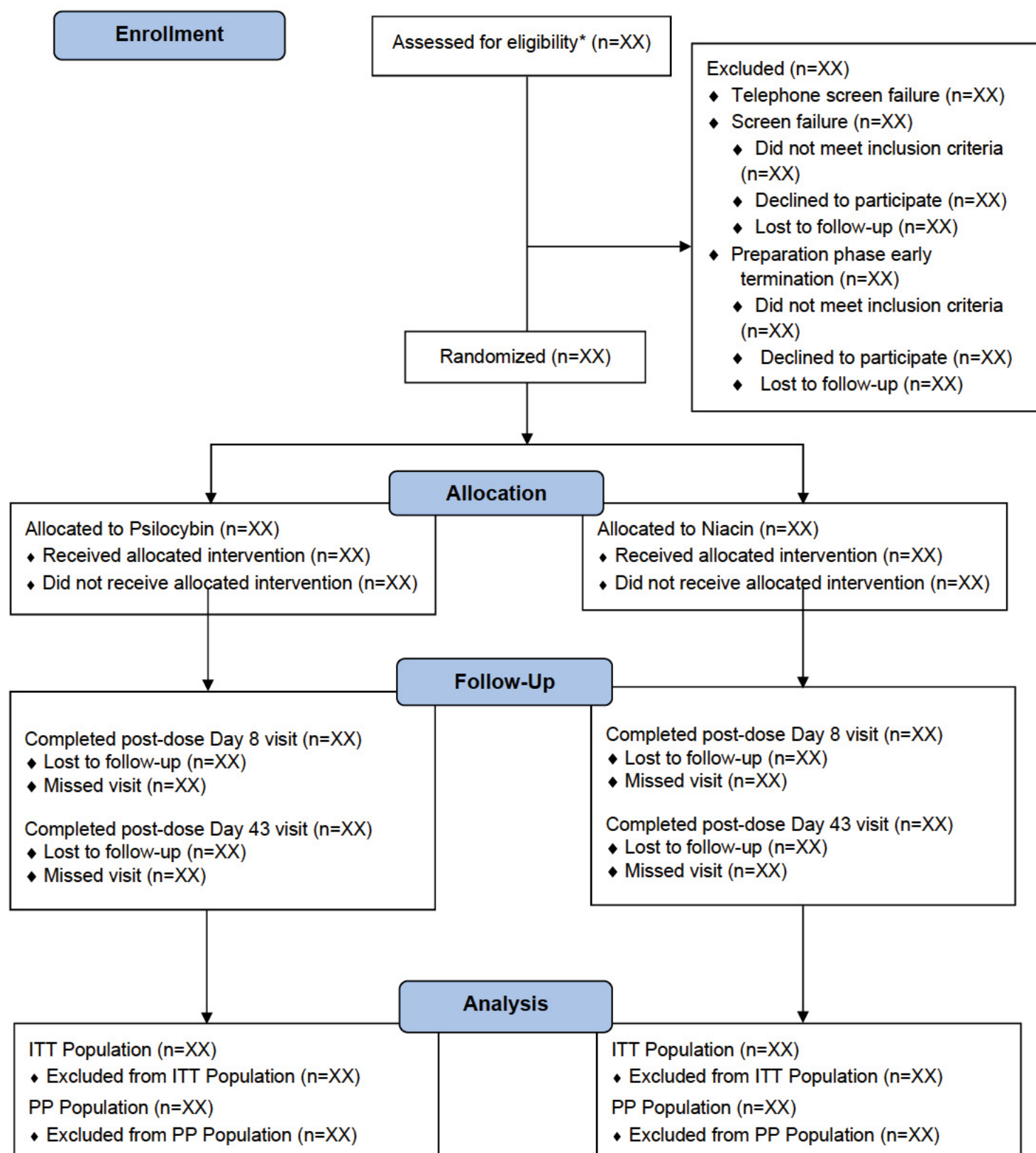
The number of participants screened, enrolled, completed baseline assessments, eligible for randomization, randomized, dosed, and completed each of the post-dose visits (Day 2, 5, 8, 9, 15, 16, 29, 43) will be tabulated and summarized by treatment and site. A visit will be defined as completed if the subject successfully completes both the computer-administered and central rated MADRS assessments.

The composition of the safety, ITT and PP populations, including reasons for participant exclusion will presented and participants who were excluded from any analysis population will be listed along with reason for exclusion.

Additionally, the number of participants with assessments or visit schedules impacted as results of the shutdown of facilities due to COVID-19 will be summarized and any possible impact to study results will be included in the body of the Clinical Study Report per FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic.

Subjects with missed visits or deviations due to COVID-19 will be listed. Additionally, Protocol v9 was updated to allow for assessments, including the MADRS, to be completed remotely if necessary. Since the primary study endpoint uses the central-rated MADRS assessments which is conducted using a remote central rater, the impact of not completing this assessment at the testing site is expected to be minimal. However, to assess any potential impact to the primary analysis, the number of participants completing the MADRS assessments at the site versus off-

site will be summarized. Depending on the proportion of in-person vs off-site assessments, additional sensitivity analyses for the primary analysis may be run.

**FIGURE 1. PSIL201 CONSORT DIAGRAM**

\*Includes all potential participants from the point of telephone screening.

### 5.3 PROTOCOL DEVIATIONS

Protocol deviations will be summarized by classification of major or minor deviation, deviation category, and treatment group.

Major protocol deviations will be classified as those deviations that have a major impact on participant rights, safety or well-being or those that have a major impact on the accuracy or completeness of study data. Deviations that do not meet these criteria will be classified as minor deviations.

The subset of major deviations that would exclude a participant from the PP population will include:

- 1) participants were randomized but did not meet eligibility criteria,
- 2) participants reported illicit use of psilocybin or other psychedelics following randomization,
- 3) participants reported use of anti-depressants or starting psychotherapy following randomization, as defined by:
  - Participants who restarted psychotherapy defined as >1 session following dosing and prior to study completion, or
  - Participants who take >1 dose of any medication listed in Appendix A on 2 consecutive days
- 4) participants received study drug not corresponding to their treatment assignment.

Protocol deviations will be reviewed by the Study Sponsor and SCC throughout the study to determine whether they should be classified as a major protocol deviation and whether the major deviations result in exclusion from the PP population.

The number of participants with deviations as well as frequencies for the types of protocol deviations will be presented by treatment group. A listing of all participant-level and site-level protocol deviations including whether the deviation was due to COVID-19 will be provided.

### 5.4 MEASUREMENT OF TREATMENT COMPLIANCE

All randomized participants will receive a single dose of either psilocybin or active placebo based on their study treatment assignment. Dosing will be administered by site personnel on the same day as randomization, therefore participant compliance is not expected to be an issue. It is anticipated that approximately 4 participants may withdraw following randomization. The number of randomized participants dosed will be presented by treatment group. Randomized participants who did not receive study drug will be listed with reason for treatment discontinuation if available. Additionally, a listing will be provided indicating those who received actual study drug not corresponding to their treatment assignment.

## 6.0 ANALYSIS OF EFFICACY ENDPOINTS

### 6.1 EFFICACY AIMS AND HYPOTHESES

**Primary Study Objective:** To evaluate the impact of a single administration of psilocybin on depressive symptoms in patients with MDD when compared to an active niacin placebo

**Hypothesis 1.1 (PRIMARY STUDY HYPOTHESIS):** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in a significant reduction in central rater MADRS scores from Baseline to Day 43 post intervention (primary study outcome).

The null ( $H_0$ ) and alternative hypotheses ( $H_A$ ) are written as follows:

$$H_0: \mu_{n, \text{Day 43}} = \mu_{p, \text{Day 43}}$$

$$H_A: \mu_{n, \text{Day 43}} \neq \mu_{p, \text{Day 43}}$$

Where  $\mu_{n, \text{Day 43}}$  and  $\mu_{p, \text{Day 43}}$  are the mean change in central rater MADRS from baseline in Day 43 post intervention in the niacin placebo and psilocybin treatment groups, respectively.

**Hypothesis 1.2 (KEY SECONDARY STUDY HYPOTHESIS):** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in a significant reduction in central rater MADRS scores from Baseline to Day 8 post-dose (key secondary study outcome).

The null ( $H_0$ ) and alternative hypotheses ( $H_A$ ) are writing as follows:

$$H_0: \mu_{n, \text{Day 8}} = \mu_{p, \text{Day 8}}$$

$$H_A: \mu_{n, \text{Day 8}} \neq \mu_{p, \text{Day 8}}$$

Where  $\mu_{n, \text{Day 8}}$  and  $\mu_{p, \text{Day 8}}$  are the mean change in central rater MADRS from baselined in Day 8 post intervention in the niacin placebo and psilocybin treatment groups, respectively.

**Hypothesis 1.3:** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in significantly higher rates of sustained response, defined as a  $\geq 50\%$  reduction from Baseline central rater MADRS scores to Day 8, 15, 29 and 43 post-dose (secondary study outcome).

**Hypothesis 1.4:** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in significantly higher rates of sustained remission, defined as a central rater MADRS scores  $\leq 10$  at Day 8, 15, 29 and 43 post-dose (secondary study outcome).

**Hypothesis 1.5:** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in a significant reduction in central rater MADRS scores from Baseline to Day 2, Day 15, and 29 post-dose (exploratory study outcome).

**Hypothesis 1.6:** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in a significant reduction in SMDDS scores from Baseline to Day 8 and 43 post-dose (exploratory study outcome).

**Hypothesis 1.7:** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in a significant reduction in SMDDS scores from Day 1 (pre-dose) to Day 8 and 43 post-dose (exploratory study outcome).

**Hypothesis 1.8:** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in a significant reduction in ODQ scores from Baseline to Day 43 post-dose (exploratory study outcome).

**Specific Aim 2: To evaluate the impact of a single administration of psilocybin on functional disability in patients with MDD when compared to an active niacin placebo**

**Hypothesis 2.1:** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in a significant reduction in SDS scores from Baseline to Day 43 post-intervention (secondary study outcome).

**Hypothesis 2.2:** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in a significant reduction in SDS scores from Baseline to Day 8, 15, and 29 post-dose (exploratory study outcome).

**Specific Aim 3: To evaluate the impact of a single administration of psilocybin on anxiety symptoms in patients with MDD when compared to an active niacin placebo**

**Hypothesis 3.1:** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in a significant reduction in HAM-A scores from Baseline to Day 8, 15, 29 and 43 post-dose (exploratory study outcome).

**Specific Aim 4: To evaluate the impact of a single administration of psilocybin on quality of life in patients with MDD when compared to an active niacin placebo**

**Hypothesis 4.1:** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in a significant increase in the Q-LES-Q scores from Baseline to Day 8, 15, 29 and 43 post-dose (exploratory study outcome).

**Hypothesis 4.2:** When compared to a single administration of niacin placebo, a single administration of psilocybin will result in a significant decrease in CGI-S scores from Baseline to Day 8, 15, 29 and 43 post-dose (exploratory study outcome).

**Specific Aim 5: To examine whether the occurrence of mystical-type or emotional breakthrough-type experiences in response to study medications predicts subsequent antidepressant responses**

**Hypothesis 5.1:** Scores on MEQ-30 will be significantly higher post-dose in participants receiving a single administration of psilocybin than in those receiving niacin placebo (exploratory study outcome).

**Hypothesis 5.2:** Scores on the EBI will be significantly higher post-dose in participants receiving a single administration of psilocybin than in those receiving niacin placebo (exploratory study outcome).

**Hypothesis 5.3:** A significant indirect effect through post-dose MEQ-30 scores of the effect of treatment assignment, psilocybin compared to niacin, on change in central rater MADRS scores will be observed for the post-dose time point at which MADRS score differences between the psilocybin and niacin placebo groups are maximal (exploratory study outcome).

**Hypothesis 5.4:** A significant indirect effect through post-dose EBI scores of the effect of treatment assignment, psilocybin compared to niacin, on change in central rater MADRS scores will be observed for the post-dose time point at which MADRS score differences between the psilocybin and niacin placebo groups are maximal. (exploratory study outcome).

**Specific Aim 6: To examine the relationship between central-rated and computer-administered MADRS measures of depressive symptoms**

**Hypothesis 6.1:** Good agreement will be observed when comparing central-rated and computer-administered versions of MADRS scores at all time points (exploratory study outcome).

## 6.2 EFFICACY ANALYSES

### 6.2.1 GENERAL CONSIDERATIONS

Continuous variables related to efficacy outcomes will be summarized using the following descriptive statistics: n (non-missing sample size), mean, standard deviation, median, maximum and minimum stratified by treatment group and study time point. Categorical efficacy outcomes will be summarized with counts and percentages. Data will be summarized for Baseline, pre-dose Day 1, and all post-dose visit as applicable for each assessment. All data will be listed, included any assessments collected at screening.

Analyses for primary and secondary efficacy endpoints will be presented for the ITT and PP populations and exploratory efficacy endpoints will be presented for the ITT population.

### 6.2.2 PRIMARY EFFICACY ANALYSIS

#### 6.2.2.1 SPECIFICATION OF PRIMARY ESTIMAND

The primary efficacy analysis will use a treatment efficacy *de facto* estimand defined in alignment with the treatment policy strategy specified in ICH E9 (R1):

- **Treatment:** Psilocybin vs. niacin
- **Population:** All randomized subjects (ITT population)
- **Variable:** Change in central rater MADRS scores from Baseline to post-dose Day 43
- **Population-level summary:** Difference in Least square means between treatment groups using the MMRM model specified in Section 6.2.2.2.

#### 6.2.2.2 PRIMARY ANALYSIS MODEL

A mixed effect model for repeated measures (MMRM) with an unstructured covariance matrix, with Baseline questionnaire score, site, sex, and TRD as covariates and Treatment (TRT) as a fixed effect will be employed to test the primary study endpoint and this same model will be

employed to test key secondary and other continuous secondary and exploratory study endpoints. The Kenward and Roger (KR) method for approximating the denominator degrees of freedom will be used in the analyses. If the analysis doesn't converge using KR, the residual method will be used instead. If the model using the unstructured covariance matrix fails to converge, a compound symmetry covariance structure with sandwich estimator will be employed, and degrees of freedom will be approximated using the containment method. The covariance structure used for each model will be specified in the table footnotes. Time will be included in the model as a categorical variable with levels for each of the 5 post-dose study visits (Day 2, 8, 15, 29 and 43).

Least squares (LS) mean change from baseline values for each treatment group will be presented with their associated standard errors. Estimated treatment difference along with corresponding 95% confidence intervals (CI) and p-values for relevant visits, will be presented for treatment comparisons.

Missing outcome data will be estimated by the MMRM model which assumes data are Missing at Random (MAR). For the primary and key secondary endpoint, sensitivity analyses to test the departure from the MAR assumptions will be conducted as specified in Table 5.

Hypothesis 1.1 (primary study endpoint) and Hypotheses 1.2 (key secondary endpoint) and 2.1 (secondary study endpoint) will be tested by a priori contrasts within the MMRM analysis regardless of whether the overall model is statistically significant.

The MMRM to be used for analysis will be set up as follows:

$$Y_{ij} = \beta_0 + b_{0i} + \beta_1 \text{Time}_{ij} + \beta_2 \text{TRT}_i + \beta_3 (\text{Time}_{ij} * \text{TRT}_i) + \beta_4 \text{baseline\_score}_i + \beta_5 \text{site}_i + \beta_6 \text{SEX}_i + \beta_7 \text{TRD}_i + \epsilon_{ij}$$

For  $i \in \{1, \dots, n\}$ , where  $n$  = number of participants and  $j \in \{1, 2, 3, 4, 5\}$  corresponding to the post dose time points at Day 2, 8, 15, 29, and 43

$Y_{ij}$  = change from baseline score in outcome measure (i.e.  $\text{Score}_{(\text{Time} = X)} - \text{Score}_{(\text{Time} = \text{Baseline})}$ )

$b_{0i} \sim N(0, \sigma_b^2)$  = random intercept for the  $i^{\text{th}}$  participant

$\epsilon_{ij} \sim N(0, \sigma_e^2)$  = error term

A sequential significance testing procedure will be employed. The primary and secondary hypotheses will proceed in the following pre-specified order with testing being halted upon the first confirmation of the null hypothesis (Hypothesis 1.1; 1.2; 2.1; 1.3; 1.4). Hypotheses associated with exploratory study endpoints will be tested with no adjustment made to the alpha level given the pre-specified exploratory nature of these hypotheses.

The primary analysis endpoint (Hypothesis 1.1) will be estimated using the *de facto* estimand as specified in Section 6.2.2.1, which will include all randomized participants. A supplemental analysis of the primary endpoint using the PP population and equivalent analysis approach will also be performed.

The MMRM model specified in this section will be used with  $Y_{ij}$  equal to the centrally rated MADRS total score change from Baseline. Least square means (LSMEANS) will

be used to compare the change from Baseline to Day 8 post-dose between the treatment groups and will be tested by a priori contrasts within the MMRM analysis (1). If the model assuming the unstructured covariance matrix fails to converge, the alternative MMRM model with compound symmetry covariance structure with sandwich estimator will be employed (2). This test will use the LSMESTIMATE statement in the SAS PROC MIXED procedure as specified below (1.1). The type III tests of fixed effects and covariance parameter estimates from the model will be presented.

**Equation 1:** MMRM SAS Syntax For Model with Unstructured Covariance

```
PROC MIXED DATA = dataset METHOD=REML DDFM=KR;
  CLASS site time trt;
  MODEL y = trt time trt*time baseline_score site sex TRD;
  REPEATED time/SUBJECT = id TYPE =UN;
(1)
```

\* Hypothesis 1.1 (Primary Endpoint);  
LSMESTIMATE trt\*time "Change in MADRS, Baseline to Day 43" 0 0 0 0 0 0 0 0 1 -1/; (1.1)

RUN;

Where: *dataset*: is the analysis dataset, *y*: outcome of interest (i.e.: change in MADRS score from Baseline), *trt*: indicator variable for the treatment group (1: psilocybin, 2: niacin), *time*: a categorical factor representing Study Day, *baseline\_score*: baseline score variable *site*: categorical site variable, *id*: patient ID

**Equation 2:** MMRM SAS Syntax for Alternate Model with Compound Symmetry Covariance Structure and Sandwich Estimator

```
PROC MIXED DATA = dataset METHOD=REML DDFM= CONTAIN EMPIRICAL;
  CLASS site time trt;
  MODEL y = trt time trt*time baseline_score site sex TRD;
  REPEATED time/SUBJECT = id TYPE =CS;
(2)
```

\* Hypothesis 1.1 (Primary Endpoint);  
LSMESTIMATE trt\*time "Change in MADRS, Baseline to Day 43" 0 0 0 0 0 0 0 0 1 -1/; (1.1)

RUN;

Where: *dataset*: is the analysis dataset, *y*: outcome of interest (i.e.: change in MADRS score from Baseline), *trt*: indicator variable for the treatment group (1: psilocybin, 2: niacin), *time*: a categorical factor representing Study Day, *baseline\_score*: baseline score variable *site*: categorical site variable, *id*: patient ID

Distributional assumptions underlying the model used for analysis will be examined by obtaining a normal probability plot of the residuals and a plot of the residuals versus the fitted values (i.e.

checking the normality assumption and constant variance assumption of the model respectively) to gain confidence that the model assumptions are reasonable.

#### **6.2.2.3 SENSITIVITY ANALYSES FOR PRIMARY AND KEY SECONDARY ENDPOINTS**

For the MMRM model used for the primary and key secondary analyses, missing data is assumed to be MAR. Table 5 provides a summary of sensitivity analyses to be performed to evaluate the robustness of this assumption.

**TABLE 5. SENSITIVITY ANALYSES FOR PRIMARY AND KEY SECONDARY ENDPOINTS**

Sensitivity Analysis Method	Purpose
MAR Multiple Imputation <sup>1</sup>	MAR approach evaluated as an alternative to maximum likelihood MMRM as mechanism of missing data handling
Reference-based Multiple Imputation (Jump to Reference (J2R))	MNAR approach to evaluate deviations from MAR assumption; statistically principled alternative to single imputation approach such as last observation carried forward. <sup>2,3</sup>  Data from participants with missing data due to lack of efficacy of the study in either intervention group will be imputed as though they were in the active placebo group.
Tipping Point Delta-Adjustment Multiple Imputation <sup>3</sup>	Justification of MAR assumptions can be performed through Tipping Point approach. MNAR approach to evaluate deviations from MAR assumption.

<sup>1</sup> Little and Rubin 2019, Molenberghs and Kenward 2007

<sup>2</sup> Bell et al 2019

<sup>3</sup> Mallinckrodt et al 2019

For sensitivity analyses involving multiple imputation, SAS PROC MI (specifying BY treatment group, using MCMC statement, with the option chain=multiple, including baseline and post baseline scheduled visits), will be used to generate 50 complete datasets. The imputation model will use the same covariates specified for the MMRM model to impute MADRS scores for any missing post-baseline timepoints. Change from baseline will be calculated using the imputed scores.

For the first sensitivity analysis for MAR Multiple Imputation, all missing data for any reason will be imputed using the MCMC method to generate 50 complete datasets.

For J2R and tipping point delta-adjustment multiple imputation, non-monotone missing due to reasons other than lack of efficacy will be first imputed using the MCMC method and MAR assumption to generate 50 datasets with monotone missing pattern prior to performing the J2R or delta-adjustment multiple imputation.

For performing the J2R analysis, data from participants with monotone missing data due to lack of efficacy of the study in either intervention group will then be imputed as though they were in the active placebo group using the PROC MI Monotone method to generate 50 complete datasets.

For the tipping point delta-adjustment multiple imputation, using the dataset with monotone missing pattern, a delta value will be added to the imputed MADRS Score at Day 43 for all participants in the Psilocybin treatment group with missing data due to lack of efficacy of the study when imputing the monotone missing using PROC MI with the Monotone method. This

process will be repeated for increasing values of delta until the value for which treatment effect for the primary outcome is no longer significant ( $<0.05$ ). A similar process will be repeated with the imputed MADRS Score at Day 8.

For all three sensitivity analyses, analysis of the imputed data will be conducted using the same MMRM model specified for the primary analysis. PROC MIANALYZE will be used to combine the estimates for the LS Means across the imputed datasets and the overall inference will be obtained by applying Rubin's Rule (Rubin, 1987).

To explore the patterns of missingness and assess the MAR assumption, baseline demographics specified in Section 5.1 will be summarized by treatment group and stratified by completion of the Day 43 and Day 8 MADRS assessment status and separately by participants with missing data due to withdraw due to lack of efficacy.

### 6.3 SECONDARY EFFICACY ANALYSES

Secondary efficacy outcomes will be assessed using the ITT and PP populations. Analyses using the ITT population will use the *de facto* estimand.

The PROC MIXED syntax as in (1) in Section 6.2.2.2 will also be employed to test the secondary endpoint associated with Hypothesis 1.2 and will use LSMESTIMATE syntax as specified in Table 6.

An MMRM model using similar SAS syntax as (1) or (2), depending on the appropriate covariance structure, with  $Y_{ij}$  equal to the change from Baseline in SDS total score will be employed for assessment of Hypothesis 2.1. The LSMESTIMATE syntax used to compare change in SDS total score from Baseline to post-dose Day 43 between treatment groups is specified in Table 6.

Sustained depressive symptom response (Hypothesis 1.3) and sustained depressive symptom remission (Hypothesis 1.4) are defined in Section 4.17. The number of participants in each treatment group meeting the definition for sustained depressive symptom response and separately sustained depressive symptom remission will be summed and presented with counts and frequencies. Treatment group differences in sustained depressive symptom response and sustained depressive symptom remission will be assessed using logistic regression model with logit link function to model the probability of sustained remission or response. Models will be adjusted for site, sex, and TRD. Odds ratios with the niacin placebo group as the reference group will be summarized with 95% CI and p-values computed using the Wald test.

### 6.4 EXPLORATORY OUTCOME ANALYSES

Analysis for exploratory outcomes for Hypotheses 1.5, 1.6, 1.7, 1.8, 2.2, 3.1, 4.1, and 4.2 will use the ITT population and mirror the analysis approach specified in Section 6.2.2.2. MMRM will be used for comparisons between treatment groups for continuous outcomes using the SAS PROC MIXED syntax. The appropriate correlation structure (unstructured or compound symmetry) will be assessed for each endpoint individually using the same approach specified in Section 6.2.2.2. All models will use the same PROC MIXED syntax as in Equation (1) or (2),

depending on the appropriate covariance structure, and will use LSMESTIMATE syntax as specified in Table 6.

For the ODQ, the Antidepressant as Cause (AC) dimension score will be summarized in addition to the total score with summary statistics by group for each time point along with change from baseline.

Hypothesis 6.1 will be evaluated using intraclass correlation (ICC) and will be presented with 95% CIs. An ICC between 0.75 and 0.90 will be classified as “good” and an ICC above 0.90 will be classified as “excellent”. Additionally, agreement between the two measures will be assessed visually with the use of scatter plots and Bland-Altman plots. The scatter plots will include the Pearson correlation coefficient and associated p-value for the correlation between the two measures. If the data is non-normal even after transformation the spearman correlation will be used. The Bland-Altman plots will show the average of the central-rated and computer-administered on the x-axis and the difference (central-rated - computer-administered) on the y-axis. If required, coefficient of repeatability will also be calculated.

**TABLE 6: SYNTAX FOR LSMESTIMATE STATEMENTS TO TEST SECONDARY AND EXPLORATORY OUTCOME HYPOTHESES**

Hypothesis	Outcome (Y <sub>ij</sub> )	Syntax for LSMESTIMATE Statement
1.2	Change from Baseline in central rater MADRS	LSMESTIMATE trt*time "Change in MADRS, Baseline to Day 8" 0 0 1 -1 0 0 0 0 0/;
1.5	Change from Baseline in central rater MADRS	LSMESTIMATE trt*time "Change in MADRS, Baseline to Day 15" 0 0 0 0 1 -1 0 0 0 0/; LSMESTIMATE trt*time "Change in MADRS, Baseline to Day 29" 0 0 0 0 0 0 1 -1 0 0/;
1.6	Change from Baseline in SMDDS	LSMESTIMATE trt*time "Change in SMDDS, Baseline to Day 8" 1 -1 0 0/; LSMESTIMATE trt*time "Change in SMDDS, Baseline to Day 43" 0 0 1 -1/;
1.7	Change from Day 1 (Pre-Dose) in SMDDS	LSMESTIMATE trt*time "Change in SMDDS, Day 1 (pre-dose) to Day 8" 1 -1 0 0/; LSMESTIMATE trt*time "Change in SMDDS, Day 1 (pre-dose) to Day 43" 0 0 1 -1/;
1.8	Change from Baseline in ODQ	LSMESTIMATE trt*time "Change in QDS, Baseline to Day 43" 1 -1/;
2.1	Change from Baseline in SDS	LSMESTIMATE trt*time "Change in SDS, Baseline to Day 43" 0 0 0 0 0 0 1 -1/;
2.2	Change from Baseline in SDS	LSMESTIMATE trt*time "Change in SDS, Baseline to Day 8" 1 -1 0 0 0 0 0 0/; LSMESTIMATE trt*time "Change in SDS, Baseline to Day 15" 0 0 1 -1 0 0 0 0/; LSMESTIMATE trt*time "Change in SDS, Baseline to Day 29" 0 0 0 0 1 -1 0 0/;

Hypothesis	Outcome (Y <sub>ij</sub> )	Syntax for LSMESTIMATE Statement
3.1	Change from Baseline in HAM-A	LSMESTIMATE trt*time "Change in HAM-A, Baseline to Day 8" 1 -1 0 0 0 0 0;  LSMESTIMATE trt*time "Change in HAM-A, Baseline to Day 15" 0 0 1 -1 0 0 0;  LSMESTIMATE trt*time "Change in HAM-A, Baseline to Day 29" 0 0 0 0 1 -1 0 0;  LSMESTIMATE trt*time "Change in HAM-A, Baseline to Day 43" 0 0 0 0 0 0 1 -1;
4.1	Change from Baseline in Q-LES-Q	LSMESTIMATE trt*time "Change in Q-LES-Q, Baseline to Day 8" 1 -1 0 0 0 0 0;  LSMESTIMATE trt*time "Change in Q-LES-Q, Baseline to Day 15" 0 0 1 -1 0 0 0;  LSMESTIMATE trt*time "Change in Q-LES-Q, Baseline to Day 29" 0 0 0 0 1 -1 0 0;  LSMESTIMATE trt*time "Change in Q-LES-Q, Baseline to Day 43" 0 0 0 0 0 0 1 -1;
4.2	Change from Baseline in CGI-S	LSMESTIMATE trt*time "Change in CGI-S, Baseline to Day 8" 1 -1 0 0 0 0 0;  LSMESTIMATE trt*time "Change in CGI-S, Baseline to Day 15" 0 0 1 -1 0 0 0;  LSMESTIMATE trt*time "Change in CGI-S, Baseline to Day 29" 0 0 0 0 1 -1 0 0;  LSMESTIMATE trt*time "Change in CGI-S, Baseline to Day 43" 0 0 0 0 0 0 1 -1;

## 6.5 MECHANISM-BASED EXPLORATORY ANALYSES

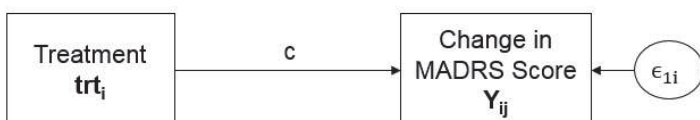
To conduct an exploratory analysis of the impact of mystical-type and/or emotional breakthrough-type experiences during the dosing sessions on subsequent longer-term antidepressant responses (Hypotheses 5.1 and 5.2), post-dose MEQ-30 and EBI total scores will be compared between intervention groups using a two-sample t-test. If data are not normally distributed appropriate non-parametric techniques will be used. For the MEQ-30, summary

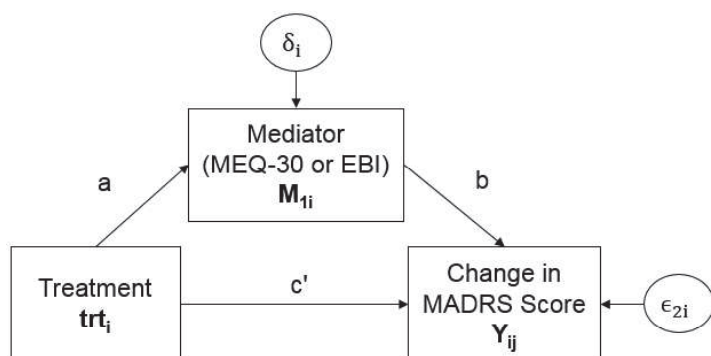
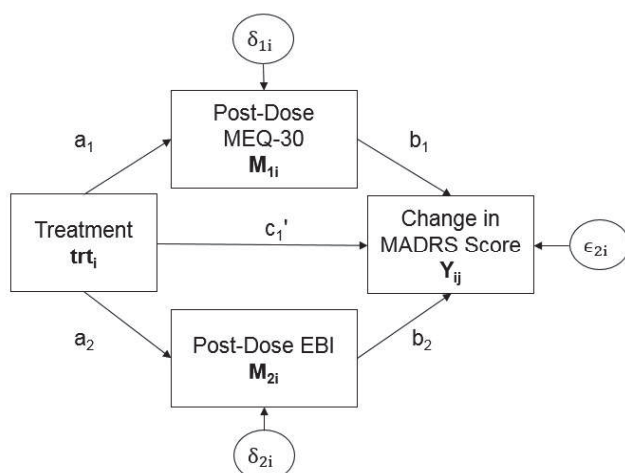
statistics by group will be provided for the 4 component scores (mystical, positive mood, transcendence of time and space, and ineffability) in addition to the total score.

Path analysis will be used to provide estimates of the magnitude and significance of hypothesized causal connections between MEQ-30 and EBI total score post-dose and subsequent antidepressant response at the post-dose time point at which MADRS score differences between the psilocybin and niacin placebo groups are maximal. Separate analyses will be performed to assess the mediation effect of MEQ-30 and EBI on change in MADRS score individually. A third analysis will be performed including both MEQ-30 and EBI as mediators in a single path diagram to compare the relative sizes of the indirect effects for each mediator in the same model.

The path diagram with no mediation is shown in Figure 2 below where  $c$  represents the total effect of treatment on change in MADRS score. The path diagram with one mediator (MEQ-30 or EBI) is shown in Figure 3, where  $a$  represents the effect of the treatment on the mediator,  $b$  is the effect of the mediator on the change in MADRS score, controlling for treatment and  $c'$  is the direct effect of treatment, controlling for the mediator. The indirect effect of treatment on change in MADRS score through the mediator can be calculated as the product of  $ab$  using the product of coefficients approach. The path diagram including both MEQ-30 and EBI is shown in Figure 4. Similar to Figure 3,  $a_1$  and  $a_2$  represent the effect of treatment on the mediators (MEQ-30 and EBI),  $b_1$ , and  $b_2$ , are the effect of the mediators on the change in MADRS score, controlling for treatment and other mediator and  $c_1'$  is the direct effect of treatment, controlling for both mediators. The indirect effect of treatment on change in MADRS score through the mediators is represented by the products  $a_1b_1$  for MEQ-30 and  $a_2b_2$  for EBI.

**FIGURE 2: PATH WITH NO MEDIATOR**



**FIGURE 3: PATH WITH SINGLE MEDIATOR, MEQ-30 OR EBI****FIGURE 4: PATH WITH MULTIPLE MEDIATORS, MEQ-30 AND EBI**

The following regression models will be used for the single mediator path analyses:

$$Y_{ij} = \alpha_{10} + c \text{TRT}_i + \epsilon_{1i} \quad (2.1)$$

$$M_i = \alpha_{20} + a \text{TRT}_i + \delta_i \quad (2.2)$$

$$Y_{ij} = \alpha_{30} + c' \text{TRT}_i + b M_i + \epsilon_{2i} \quad (2.3)$$

For  $i \in \{1, \dots, n\}$ , where  $n$  = number of participants

$Y_{ij}$  = change from baseline score in MADRS (i.e.  $\text{Score}_{(\text{Time} = X)} - \text{Score}_{(\text{Time} = \text{Baseline})}$ ) at time point  $j$  at which MADRS score differences between the psilocybin and niacin placebo groups are maximal

$M_i$  = Mediator (MEQ-30 or EBI)

$\epsilon_{1i} \sim N(0, \sigma_e^2)$ ,  $\epsilon_{2i} \sim N(0, \sigma_e^2)$ , and  $\delta_i \sim N(0, \sigma_e^2)$  = error terms

Prior to performing the path analysis, the association between treatment group and change in MADRS score will be assessed using the regression model as shown in equation (2.1). If the association between treatment and change in MADRS score is significant at the 0.05 level of

significance ( $p < 0.05$ ) then the path analysis will proceed as described below. If there are no post-dose visits for which the treatment assignment is significantly associated with the change in MADRS score at the 0.05 level of significance, then no further path analyses will be performed.

The results from equations (2.2) and (2.3) will be summarized for MEQ-30 and EBI. The indirect effect of treatment on change in MADRS score through MEQ-30 or EBI (ab from Figure 2), will be estimated by the product of a from model x.2 and b from model (3.3) (ab). The standard errors and 95% confidence interval terms will be calculated using bootstrap methods.

To assess the effects of MEQ-30 and EBI in a single path diagram, the following regression models will be used:

$$M_{1i} = \beta_{10} + a_1 \text{TRT}_i + \delta_{1i} \quad (3.1)$$

$$M_{2i} = \beta_{20} + a_2 \text{TRT}_i + \delta_{2i} \quad (3.2)$$

$$Y_{ij} = \beta_{30} + c'_1 \text{TRT}_i + b_1 M_{1i} + b_2 M_{2i} + \epsilon_{2i} \quad (3.3)$$

For  $i \in \{1, \dots, n\}$ , where  $n$  = number of participants

$Y_{ij}$  = change from baseline score in MADRS (i.e.  $\text{Score}_{(\text{Time} = X)} - \text{Score}_{(\text{Time} = \text{Baseline})}$ ) at time point  $j$  at which MADRS score differences between the psilocybin and niacin placebo groups are maximal

$M_{1i}$  = score on MEQ-30

$M_{2i}$  = score on EBI

$\epsilon_{1i} \sim N(0, \sigma_e^2)$ ,  $\delta_{1i} \sim N(0, \sigma_e^2)$ , and  $\delta_{2i} \sim N(0, \sigma_e^2)$  = error terms

The indirect effect of treatment on change in MADRS score through MEQ-30 controlling for the effect of EBI ( $a_1 b_1$  from Figure 4), will be estimated by the product of  $a_1$  from model (3.1) and the  $b_1$  term from model (3.3) ( $a_1 b_1$ ). Similarly, the indirect effect of treatment on change in MADRS score through EBI controlling for the effect of MEQ-30 ( $a_2 b_2$  from Figure 3), will be estimated by the product of  $a_2$  term from model (3.2) and the  $b_2$  term from model (3.3) ( $a_2 b_2$ ). The standard errors and 95% confidence interval for both terms will be calculated using bootstrap methods.

To assess if the indirect effect of treatment through MEQ-30 is different from EBI, the following contrast will be used:  $f_c = a_1 b_1 - a_2 b_2$  (Preacher and Hayes 2008). Bootstrap methods will be used to obtain the 95% confidence interval for the contrast.

## 6.6 SUPPLEMENTAL ANALYSES FOR EFFICACY OUTCOMES

Table 7 presents analyses that will be conducted to supplement information provided in primary, secondary, and exploratory analysis.

**TABLE 7. SUPPLEMENTAL ANALYSES**

<b>Modification to Primary/Secondary/Exploratory Analysis</b>	<b>Applicable Hypotheses</b>	<b>Purpose</b>
Replace central-rater MADRS total score with computer-administered MADRS total score	Hypotheses 1.1, 1.2, 1.3, 1.4, and 1.5	Evaluate consistency in results using computer-administered vs. central-rater MADRS to inform selection of endpoints for future studies
Fit MMRM for central-rater MADRS score with study day (i.e. time) treated as a continuous variable	Hypotheses 1.1, 1.2	Investigate possible temporal effect of depressive symptom response to psilocybin vs. niacin
Add baseline central-rater MADRS total score by treatment group interaction term to primary and secondary MADRS analyses	Hypotheses 1.1, 1.2, 1.3, 1.4	Adjust for potential treatment group imbalance in distribution of baseline central-rater MADRS score.
Define participants with missing central rater MADRS due to lack of efficacy as non-response or non-remission	Hypotheses 1.3, 1.4	Account for missing data due lack efficacy when defining non-response and non-remission

## 7.0 ANALYSIS OF SAFETY ENDPOINTS

Safety summaries will be presented for the following Study Periods:

- Period 1 (Screening Period): Screening through end of Baseline;
- Period 2 (Preparation Period): Preparation until Randomization;
- Period 3 (Dosing Period): Randomization through Day 9; and
- Period 4 (Follow-up Period): Day 10 through Day 43.

In general, adverse events will be summarized for the All Enrolled population (Periods 1 and 2), the Preparation Phase Early Termination Safety population (Period 2 only), and the Safety population (Periods 3 and 4). Safety summaries will be presented overall and stratified by treatment group for Periods 3 and 4.

Other safety endpoints will be summarized only for the Safety population unless otherwise specified.

### 7.1 ADVERSE EVENTS

AEs will be tabulated by body system and classified by severity and relationship to study product. When calculating the incidence of AEs including solicited AEs, TEAEs, SAEs, and SUSARs (i.e., on a per participant basis), each participant will be counted once at the highest severity and/or relationship reported within a given time period. Multiple instances of the same AE within a participant for that time period will be counted as separate events. AEs that are not resolved within a Study Period and continue into a subsequent period at the same severity level will only be counted in the period in which the AE began and included in the in subsequent periods only if the severity or frequency of the AE increases or if the initial event was fully resolved prior to onset of a second occurrence of the same AE.

TEAEs are defined as AEs with an onset after study drug administration or an already present event that worsens either in intensity or frequency following the treatment. TEAEs will be identified based on the date and time of onset or worsening relative the date and time of study drug administration.

At each level of summarization, the total number of AEs experienced will be reported as counts and number of participants with at least one AE will be summarized with counts and percentages. The denominator will be the number of participants in the safety population enrolled in the study during that Study Period and will be clearly indicated in the table and 95% CIs (Clopper Person Exact) will be presented for percentages.

Relative incidence of AEs comparing the psilocybin and active placebo groups will be calculated by dividing the percentage of participants experiencing an event in the psilocybin treatment group by the percentage of participant experiencing an event in the active placebo group and will be presented with 95% CIs (Wilson).

All AEs including unsolicited and solicited adverse events will be recorded and tabulated overall and for each Study Period. AEs will be summarized with counts and percentages as described above and will be presented by MedDRA system organ class (SOC) sorted by preferred term (PT). As applicable, AEs will be presented by Study Period, treatment group, severity, action taken (requiring medical attention, psychiatric attention), relationship to study product, and Council for International Organizations of Medical Sciences (CIOMS) frequency parameter.

AEs leading to discontinuation from the study will also be listed.

In addition to general AE summaries, AEs classified as cardiovascular proarrhythmic and abuse liability events will be summarized separately for Periods 3 and 4 only as specified in Sections 7.1.1 and 7.1.2.

### **7.1.1 CARDIOVASCULAR PROARRHYTHMIC ADVERSE EVENTS**

The following MedDRA Preferred Terms will be used to classify proarrhythmic AEs:

- Torsade de pointes
- Sudden death
- Ventricular tachycardia
- Ventricular fibrillation and flutter
- Syncope
- Seizures

AEs classified as proarrhythmic AEs based on the above criteria will be summarized by preferred term, study period and treatment group. At each level of summarization, the total number of AEs experienced will be reported as counts and number of participants with at least one AE will be summarized with counts and percentages. The denominator will be the number of participants in the safety population enrolled in the study during that Study Period and will be clearly indicated in the table.

### **7.1.2 ABUSE LIABILITY ADVERSE EVENTS**

The following MedDRA Preferred Terms will be used to classify abuse liability AEs:

- Euphoria-related terms: euphoric mood, elevated mood, feeling abnormal, feeling drunk, feeling of relaxation, dizziness, thinking abnormal, hallucination, inappropriate affect
- Terms of impaired attention, cognition, and mood: somnolence, mood disorders, and disturbances
- Dissociative/psychotic terms: psychosis, aggression, confusion, and disorientation

AEs classified as Abuse liability AEs based on the above criteria will be summarized by term category (Euphoria-related terms, Terms of impaired attention, cognition, and mood, and Dissociative/psychotic terms), preferred term, study period and treatment group. At each level of summarization, the total number of AEs experienced will be reported as counts and number of participants with at least one AE will be summarized with counts and percentages. The denominator will be the number of participants in the safety population enrolled in the study during that Study Period and will be clearly indicated in the table.

### **7.1.3 TREATMENT EMERGENT ADVERSE EVENTS**

TEAEs will be recorded and tabulated for Periods 3 and 4 for the participants in the Safety population who were enrolled in the study during that period.

### **7.1.4 SOLICITED ADVERSE EVENTS**

The following solicited events will be reported as AEs: visual perceptual effects, active suicidal ideation verified by clinical assessment, headache, nausea, elevated BP requiring medication, elevated HR requiring medication, and drug overdose with suicidal intent. These solicited AEs will be presented by Study Period, treatment group, severity and relationship to study product. Visual perceptual effects reported on the day of dosing and at post-dose follow up visits will be summarized separately.

Solicited AEs observed in randomized participants who received psilocybin with an incidence of 5% or greater and at least twice that of active placebo will be reported separately.

### **7.1.5 SERIOUS ADVERSE EVENTS**

SAEs will be summarized for all Study Periods. Serious TEAEs and SUSARs will be presented separately for Periods 3 and 4.

Individual data listings of deaths and serious AEs will be provided. The listing will include: participant ID, treatment group, AE description, duration, reason reported as an SAE, severity, relationship to treatment, alternate etiology if not related, action taken, whether participant discontinued due to AE, outcome, and MedDRA SOC and PT.

## **7.2 MEDICAL HISTORY**

A summary of baseline abnormalities will be provided by treatment group based on self-reported medical history collected at the Screening visit. Baseline psychiatric history will be assessed via the SCID-CT and SCID-PD and will be summarized along with other medical history data.

## **7.3 PHYSICAL EXAM**

A summary of abnormalities by body system and treatment group will be provided based on the complete physical exam conducted at Screening and Baseline. In addition, incidence of new clinically significant abnormalities identified post-dose will be presented by treatment group based on the physical exam conducted at post-dose Day 8 and Day 43. The relative incidence of clinically significant physical examination findings comparing the psilocybin and active placebo groups will be summarized as specified in section 7.1.

## **7.4 CLINICAL LABORATORY EVALUATIONS**

The following laboratory measurements will be collected at Screening, Baseline, Day 2, Day 8 and Day 43: Complete Metabolic Panel (CMP), Complete Blood Count with Differential (CBC w/ Diff), Thyroid Stimulating Hormone (TSH), high-sensitivity C-Reactive Protein (hs-CRP), urinalysis, INR (screening only), urine drug test, and urine pregnancy test.

Laboratory results will be summarized by Study Day, laboratory parameter, clinical significance and treatment group. Laboratory measurements will be summarized by treatment group using the same units as specified on the case report form (CRF). Continuous laboratory parameters will be summarized using descriptive statistics at baseline, each follow-up time point, and change from baseline and will be presented by treatment group. Categorical urinalysis variables will be summarized as negative and non-negative with counts and percentages. Additionally, laboratory parameters classified as normal, abnormal and clinically significant, and abnormal and non-clinically significant will be summarized by these classifications at each time point. Urine drug test results will be reported as either positive or negative at each time point.

Change from baseline in clinically significant abnormal labs will be summarized using a shift table, summarizing the proportion of subjects who had normal or abnormal non-clinically significant laboratory results at Baseline and abnormal clinically significant results post-dose by study time point. Proportions will be presented along with 95% CIs and relative incidence comparing the psilocybin and active placebo groups as specified in section 7.1.

## 7.5 VITAL SIGNS

Vital sign measurements include systolic blood pressure, diastolic blood pressure, heart rate and temperature. Vital signs are assessed at screening, baseline, prior to dosing on Study Day 1, throughout the dosing period and post dose on Study Days 2, 8, 15, 29 and 43. Descriptive statistics including mean, standard deviation, median, min and max values will be presented by study time point and treatment group.

BP and HR collected on post-dose Day 1 at 30, 60, 90, 120 minutes, 4, 6, and 7 hours after drug administration will be summarized in tabular format and graphically with box-plots showing the change from baseline (pre-dose Day 1) at each time point by treatment group.

The proportion of participants with BP or HR classified as an AE will be presented along with 95% CIs and relative incidence comparing the psilocybin and active placebo groups as specified in section 7.1.

For BP and HR monitoring at the pre-dose and day of dosing time points, the study protocol allows for the BP or HR measures to be repeated if the initial measurements are out of normal range. For timepoints with replicated measurement, the last recorded value for that time point will be used for analysis. All recorded values will be listed.

## 7.6 CONCOMITANT MEDICATIONS

Concomitant medications will be coded using the most recent WHODrug dictionary and will be presented by ATC4 drug class.

For Study Periods 1 and 2, summary tables tabulating the incidence and frequency of new concomitant medications will be presented by study period for all participants.

For study periods 3 and 4, summary tables tabulating the incidence and frequency of new concomitant medications will be presented by study period for all participants and stratified by

treatment group along with the relative incidence of new concomitant medications. Additionally, the incidence and frequency of use of new psychiatric concomitant medications will also be presented by study period for all participants, stratified by treatment group and with the relative incidence of new psychiatric concomitant medications.

New concomitant medications for a particular Study Period will be defined as any instance of either a new medication or a new dose of a medication with a start date that falls within that particular Study Period.

Psychiatric concomitant medications will be defined as medications with WHODrug Therapeutic subgroup (ATC2) codes of N05 – Psycholeptics or N06 – Psychoanaleptics.

## **7.7 ABUSE MONITORING**

In addition to abuse liability AEs to be reported per Section 7.1.2, analysis of urine drug screens and the AUDIT, DUDIT, SCID-CT, and Timeline Follow-back will be used to assess drug and alcohol use following dosing.

Drug abuse/misuse based on the DUDIT will be defined as a score of 6 or higher for men and 2 or higher for women. Alcohol abuse/misuse based on the AUDIT will be defined as a score of 8 or higher for men and 6 or higher for women. The number and percentage of participants meeting the specified cut point for drug or alcohol abuse/misuse at baseline and the incidence of new drug or alcohol abuse/misuse at post-dose Day 43 will be summarized by treatment group for each assessment tool. The incidence of new substance or alcohol use disorders as defined by the SCID-CT at Day 43 will be summarized by treatment group.

Incidence of any self-reported use via the TLFB or positive urine toxicology identified on the urine drug screen for illicit and non-prescribed drug use will be summarized by group and presented with relative incidence rates and 95% CIs (Wilson).

Any use based on the results of post-dose urine drug screens or indicated in the TLFB will be presented for alcohol use and for each drug class by treatment group using frequencies and percentages. All self-identified use from the TLFB and positive post-dose urine drug screens will be listed.

Differences between treatment groups in frequency of post-dose usage will be tested for each drug group using chi-squared or Fisher's exact tests.

False-positive results for the urine drug screen that have been verified by the site PI will not be counted as a positive result in the summary tables, but will be included in the listing along with the comment provided by the site PI.

## **7.8 PREGNANCIES**

For any participants in the safety population who became pregnant during the study, every attempt will be made to follow these participants to completion of pregnancy and live births will be followed for a minimum of 30 days post-delivery or to the first well-baby visit in order to document the outcome, including information regarding any complications with pregnancy,

delivery or congenital abnormalities/birth defects. Pregnancy, in and of itself, and elective abortion procedures, without complications, are not considered as AEs. A listing of pregnancies and outcomes will be presented.

## **8.0 TABLES, FIGURES, AND LISTINGS**

Mock tables, figures, and listings will be presented in the final SAP.

## 9.0 REFERENCES

- Bell ML, Floden L, Rabe BA, et al. Analytical approaches and estimands to take account of missing patient-reported data in longitudinal studies. *Patient Relat Outcome Meas*. 2019 Apr 16;10:129-140. doi: 10.2147/PROM.S178963. PMID: 31114411; PMCID: PMC6489631.
- Carhart-Harris RL, Bolstridge M, Rucker J, et al. Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *Lancet Psychiatry*. 2016 Jul;3(7):619-627. doi:10.1016/S2215-0366(16)30065-7.
- Gurka MJ, Edwards LJ, Muller KE. Avoiding bias in mixed model inference for fixed effects. *Stat Med*. 2011 Sep 30;30(22):2696-707. doi: 10.1002/sim.4293. Epub 2011 Jul 12. PMID: 21751227; PMCID: PMC3396027.
- Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol*. 1959;32(1):50-5. doi: 10.1111/j.2044-8341.1959.tb00467.x. PMID: 13638508.
- International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH Harmonised Guideline E9(R1): Estimands and Sensitivity Analysis in Clinical Trials. Current Step 1 version dated 16 June 2017.
- Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016 Jun;15(2):155-63. doi: 10.1016/j.jcm.2016.02.012.
- Little R, Rubin D. *Statistical Analysis with Missing Data*, Second Edition. John Wiley & Sons, 2019. doi: 10.1002/9781119482260.
- Mallinckrodt CH, Bell J, Liu G, et al. Aligning Estimators With Estimands in Clinical Trials: Putting the ICH E9(R1) Guidelines Into Practice. *Ther Innov Regul Sci*. 2020 Mar;54(2):353-364. doi: 10.1007/s43441-019-00063-9. Epub 2020 Jan 6. PMID: 32072593.
- Mallinckrodt C, Molenberghs G, Lipkovich I, Ratitch B. *Estimands, Estimators and Sensitivity Analysis in Clinical Trials* (1st ed.). CRC Press, 2019. Doi: 10.1201/9780429488825
- Molenberghs G, Kenward MG. *Missing Data in Clinical Studies*. John Wiley; New York, 20017-01. doi: 10.1002/9780470510445.
- Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav Res Methods*. 2008 Aug;40(3):879-891.
- Ross S, Bossis A, Guss J, et al. Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: a randomized controlled trial. *J Psychopharmacol*. 2016 Dec;30(12):1165-1180.
- Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons, Inc, 1987.









Sobel ME. Asymptotic confidence intervals for indirect effects in structural equations models. *Sociological Methodology*. 1982;13:290-312. DOI: 10.2307/27072.

United States Department of Health and Human Services Food and Drug Administration (FDA). Conduct of Clinical Trials of Medical Products During the COVID-19 Public Health Emergency: Guidance for Industry, Investigators, and Institutional Review Boards. March 2020, Updated on 04 December 2020.


**APPENDIX A: PLAN APPROVAL SIGNATURE PAGE**

The undersigned acknowledge they have reviewed the PSIL201 Statistical Analysis Plan and agree with the approach provided. Changes to this plan will be coordinated with and approved by the undersigned or their designated representatives.

Prepared by:

			
Signature:	I am approving this document.	Date:	09/Aug/2022 03:06 PM EDT
Print Name:	 		
Title:	Usona SCC Principal Investigator		
			
Signature:	I am approving this document.	Date:	09/Aug/2022 03:46 PM EDT
Print Name:	 		
Title:	Usona SCC Statistician		

Reviewed by:

	<i>Charles Raison</i>		
Signature:	I am approving this document.	Date:	17/Aug/2022 11:47 AM EDT
Print Name:	Charles Raison, MD    Charles Raison		
Title:	Usona Institute Director of Clinical and Translational Research		