MAPP2 Statistical Analysis Plan

Version 1: 1 May 2019
Version 2: 22 Jan 2020
Version 3: 10 Jun 2020
Version 4: 06 Sep 2022

SPONSOR	Multidisciplinary Association for Psychedelic Studies (MAPS) 3141 Stevens Creek Blvd #40563 San Jose, CA, 95117
SPONSOR REPRESENTATIVE	Rick Doblin, Ph.D. Executive Director, MAPS
SPONSOR DELEGATE & TRIAL ORGANIZER	MAPS Public Benefit Corporation (MPBC) 3141 Stevens Creek Blvd #40547 San Jose, CA, 95117
SPONSOR DESIGNEE	Amy Emerson Chief Executive Officer, MPBC Email: amy@mapsbcorp.com
MEDICAL MONITOR(S)	Sara Garcia Velaszquez, M.D.,
STATISTICIAN(S)	Scott Hamilton, Ph.D.
PREPARED BY	Scott Hamilton, Ph.D. Claire Le Lait, M.S.
USE	In conjunction with relevant FDA, HC, MOH guidance

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACE	Adverse Childhood Experiences Questionnaire
AE	Adverse Event
AESI	Adverse Event of Special Interest
ATC	Anatomical Therapeutic Chemical
AUDIT	Alcohol Use Disorders Identification Test
BDI-II	Beck Depression Inventory-II
BMI	Body Mass Index
BP	Blood Pressure
CAPS-4	Clinician Administered PTSD Scale for DSM-4
CAPS-5	Clinician Administered PTSD Scale for DSM-5
CPGS	Chronic Pain Grade Scale
C-SSRS	Columbia Suicide Severity Rating Scale
DDIS	Dissociative Disorders Interview Schedule
DMC	Data Monitoring Committee
DSM-4	Diagnostic and Statistical Manual of Mental Disorders, 4th edition
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th edition
DSPI	Dissociative Subtype of PTSD Interview
DUDIT	Drug Abuse Disorders Identification Test
EAT-26	Eating Attitudes Test
EQ-5D-5L	EuroQol Five Dimensions-Five Levels Questionnaire
GCP	Good Clinical Practice
HPQSF	Health and Work Performance Absenteeism and Presenteeism Short Form
IASC	Inventory of Altered Self-Capacities
IP	Investigational Product
IPF	Inventory of Psychological Functioning
IR	Independent Rater
IWRS	Interactive Web Randomization System
LEC-5	Life Events Checklist
LS	Least Squares
MAPS	Multidisciplinary Association for Psychedelic Studies
MAR	Missing at Random
MDMA	3,4-methylenedioxymethamphetamine
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
mITT	Modified Intent-to-Treat
MMRM	Mixed Model Repeated Measure
MNAR	Missing Not at Random
MPBC	MAPS Public Benefit Corporation
NPP	Not Per Protocol
PCL-5	PTSD Checklist for DSM-5
PP	Per Protocol
PSQI	Pittsburgh Sleep Quality Index
PT	Preferred Terms
PTSD	Posttraumatic Stress Disorder
REML	Restricted Maximum Likelihood
SAE	Serious Adverse Event

SAP	Statistical Analysis Plan
SCS	Self-compassion Scale
SDS	Sheehan Disability Scale
SOC	System Organ Class
SRNU	Self-reported Nicotine Use
TEAE	Treatment Emergent Adverse Event
TAS-20	Toronto Alexithymia Scale
UFEC	Utilization of Facility-based and Emergent Care
WHO DDE	WHO Drug Dictionary EnhancedTM

1.0 Definitions of Terms

Categorical data: refers to discrete (indivisible) variables, such as gender or ethnicity; data will be presented as total numbers of each category as needed to describe the sample

Completers: are defined as participants who complete all three planned Experimental Sessions and the CAPS-5 outcome assessment 18 weeks after randomization (Visit 19)

Descriptive data: includes mean, median, standard deviation, minimum, and maximum of numerical data used as needed to describe the sample

Difference scores: consist of scores computed by subtracting one value from another, as subtracting Baseline from follow-up score, used to test for differences between and within groups to determine change as a function of experimental treatment over time

Dropouts: are defined as participants who withdraw consent due to any reason after randomization and no longer participate in the study, i.e., no further contact with investigators or site staff

Efficacy: type of analysis used to assess therapeutic effects or benefits

Exploratory analyses: inferential or descriptive analysis of the data to determine trends that might lead to hypotheses for further study

Frequency listing: tabular listing of numbers and/or percentages of events used as needed to describe the sample or data characteristics

Outcome measures: primary and secondary study measures that are used to test the study hypotheses

Post-randomization Early Terminators: are defined as participants who discontinue study treatment but continue to participate in study evaluations and outcome assessments

Pre-randomization Early Terminators: are defined as participants who discontinue participation after enrollment but before randomization during the Preparatory Period and never receive Investigational Product

Process measures: study measures or qualitative observations collected during the study that may increase depth of understanding of the condition and treatment, although not necessarily related to safety or efficacy

Protocol deviation: event that represents significant divergence from the intended study design as described in the protocol

Safety: assessment of indicators of potential risks and adverse events

Safety measures: study measures that assess safety of the Investigational Product (IP), such as heart rate monitoring, blood pressure, body temperature

Study design: all elements of a research project that define the study question, experimental methods, study procedures including randomization and blinding, measurement techniques, data workflow, and statistical analysis

Tabular listing: list of each variable or item for each individual participant either in total or by treatment group in a table format

2.0 Introduction

This document contains a Statistical Analysis Plan (SAP) for the study, "A Randomized, Double-Blind, Placebo-Controlled, Multi-Site Phase 3 Study of the Efficacy and Safety of Manualized MDMA-Assisted Psychotherapy for the Treatment of Posttraumatic Stress Disorder of Moderate or Greater Severity."

The Multidisciplinary Association for Psychedelic Studies (MAPS) is a non-profit research and educational organization working to obtain approval for the prescription use of 3,4-methylenedioxymethamphetamine (MDMA) in conjunction with therapy in persons with posttraumatic stress disorder (PTSD). MAPS has delegated trial organization activities to its wholly owned subsidiary, MAPS Public Benefit Corporation (MPBC).

To confirm the efficacy and safety of this treatment for participants with moderate or greater severity of PTSD, the sponsor is conducting this Phase 3 randomized, placebo-controlled, double-blind, multi-site study with three monthly Experimental Sessions of therapy combined with either MDMA or placebo, along with 12 non-drug therapy sessions including Preparatory and Integrative sessions. The Primary Outcome measure, the Clinician Administered PTSD Scale (CAPS-5), evaluates PTSD symptom severity and is assessed by a blinded centralized Independent Rater (IR) pool.

3.0 Study Objectives

3.1 Primary Objective

The primary objective of this study is to evaluate the *de jure* efficacy of MDMA-assisted therapy for PTSD compared to identical therapy with inactive placebo, as measured by

the reduction in CAPS-5 Total Severity Score from Visit 3 (Baseline) to Visit 19 (18 weeks post Baseline).

3.2 Secondary Objective

The key secondary objective of this study is to evaluate the efficacy of MDMA-assisted therapy for PTSD compared to identical therapy with inactive placebo in clinician-rated functional impairment, as measured by the mean change in Sheehan Disability Scale (SDS) item scores from Visit 3 (Baseline) to Visit 19 (18 weeks post Baseline).

3.3 Safety Objectives

The overall safety objective is to assess differences between groups in severity, incidence and frequency of Adverse Events (AEs), Treatment Emergent AEs (TEAEs), AEs of Special Interest (AESIs), and Serious Adverse Events (SAEs), concomitant medication use, suicidal ideation and behavior, and vital signs to support the package insert for MDMA-assisted therapy. The following safety objectives will evaluate the safety of MDMA-assisted therapy compared to identical therapy with inactive placebo:

1. Compare relative incidence of AEs during Experimental Sessions that may be indicative of a medical complication of the Investigational Product (IP) such as clinical signs and symptoms of chest pain, shortness of breath, or neurological symptoms or any other signs or symptoms that prompt additional vital sign measurements
2. Compare relative incidence of AEs by severity
3. Compare relative incidence of TEAEs, to determine relationship to the IP based on relative incidence in the MDMA group
4. Compare relative incidence of TEAEs by severity reported during an Experimental Session, 1 and 2 days after IP administration
5. Compare relative incidence of AESIs, defined as AEs specified in the protocol related to cardiac function and abuse liability
6. Compare relative incidence of AEs by severity categorized as leading to discontinuation of IP, resulting in death or hospitalization, and continuing at Study Termination
7. Compare relative incidence of SAEs
8. Compare relative incidence of concomitant medications taken during an Experimental Session, 1 and 2 days after IP administration
9. Compare relative incidence of psychiatric concomitant medications taken during the Treatment Period
10. Compare relative incidence of positive or serious ideation and suicidal behavior assessed with the Columbia Suicide Severity Rating Scale (C-SSRS) by treatment group
11. Compare median changes in blood pressure, heart rate, and body temperature from pre-IP administration to end of each Experimental Session by treatment group

3.4 Exploratory Objectives

The study protocol describes additional objectives that may be explored to characterize participants receiving MDMA-AT in comparison to those receiving therapy with inactive placebo to support the primary objective. These objectives may be assessed at future timepoints for post hoc analyses or exploratory publications and therefore will not be included in the present analysis plan.

4.0 Measures

Table 1: Protocol Objectives and Assessment Tools

Objectives	Measure	Measure Type	Administration
Eligibility			
Assess Axis 1 psychiatric disorders	MINI	Eligibility	Site
Assess Axis 2 personality disorders	SCID-5-PD with SCID-5-SPQ	Eligibility	Telemedicine/ self-report measure at Site
Confirm PTSD diagnosis and severity	PCL-5 with LEC-5	Eligibility	Site
Identify dissociative disorders	Dissociative Disorders Interview Schedule (DDIS)	Eligibility	Site
Primary			
Assess changes in PTSD symptom severity from Visit 3 to Visit 19 compared between groups	CAPS-5	Outcome	Telemedicine
Key Secondary			
Assess mean changes in clinician-rated functional impairment from Visit 3 to Visit 19 compared between groups	SDS Mean item scores	Outcome	Site
Safety			
Compare relative incidence of positive or serious ideation and suicidal behavior between groups	C-SSRS	Safety	Site

5.0 Study Design

Table 2: Phase 3 Dose Regimen of MDMA or Placebo

Experimental Session	Initial Dose	Supplemental Dose*	Min-Max Cumulative Dose
1	80 milligrams (mg)	40 mg	80 mg to 120 mg
2	80 or 120 mg	40 or 60 mg	80 mg to 180 mg
3	80 or 120 mg	40 or 60 mg	80 mg to 180 mg
Total Cumulative Dose			240 mg to 480 mg

* Unless tolerability issues emerge or the participant declines.

Table 3: Schedule of Events

Study Procedures	**Screening Period** (2-6 weeks)				**Preparatory Period w/ Enrollment Confirmation** (1-11 weeks)			
	Screening			**Enrollment**	**Preparatory**		**Baseline CAPS-5 T1**	**Baseline & Enrollment Confirmation**
Visit	Phone Screening	Screening	IR Screening	V0	V1	V2	V3	V4
Visit Description	Phone Calls	In-person Visits & Labs	Telemedicine	Enrollment	Prep. 1	Prep. 2	Telemedicine	Prep. 3 & Enrollment Confirmation
Visit Timing	Prior to Initial Screening	Over 3 wks (-2/+1 wks)	2 days post Screen (+7 days)	2 days post SCID (+5 days)	Within 1 wk of V0 (±5 days)	Within 3 wks of V1	Post V2 & Taper	Within 6 days of V3
Initial Phone Screen	✓							
Informed Consent	Send Copy	✓						
Follow-up Phone Screen	✓							
Assess Eligibility	✓	✓		✓	✓	✓		✓
Medical/Psychiatric History	✓ [A]	✓			✓	✓		✓
Past/Current Medication & Adherence	✓	✓			✓	✓		✓
Weight, Resting Vitals		✓						
Physical Exam		✓						
ECG & Rhythm Strip		✓						
Clinical Lab Tests, BAC, HIV Test		✓						
Drug Screen		✓						✓
Pregnancy Screen		✓						✓
Enter Participant in eCRF [B]		✓						
Record			✓		✓	✓	✓	✓
Medication Taper				✓	✓	✓		
Study Enrollment				✓				✓ Confirmed
All AEs [C]				✓	✓	✓		✓
90-min Preparatory Session					✓	✓		✓
Phone Call Follow-up [D]						✓		

[A] At Screening, collect data on previous hospitalizations and healthcare utilization. Request participants to obtain medical/psychiatric records to bring to the in-person screening.

[B] Participants will be entered into the eCRF after the IR visit is scheduled

[C] All Adverse Events (AEs) includes collecting Serious Adverse Events, AEs of Special Interest, AEs of Psychiatric Status, AEs requiring medical advice or attention, AEs that indicate withdrawal of a participant, and all other AEs

[D] If needed, call participant to confirm medication tapering and stabilization is complete prior to Visit 3

	Treatment Period ~12 weeks (±3 weeks)														**Follow-up Period & Study Termination** 18 weeks (±3 wks) post Baseline	
	Treatment 1					**Treatment 2**					**Treatment 3**				**Primary Outcome**	**Study Termination**
Visit	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
Visit Description	Exp. 1	Int. 1.1	Int. 1.2	CAPS-5 T2: Tele-medicine	Int. 1.3	Exp. 2	Int. 2.1	Int. 2.2	CAPS-5 T3: Tele-medicine	Int. 2.3	Exp. 3	Int. 3.1	Int. 3.2	Int. 3.3	CAPS-5 T4 Outcome: Tele-medicine [I]	Study Termination
Visit Timing	Within 1 wk of V3	Morning after V5	Within 2 wks of V5	Within 4 wks of V5	1-7 days before V10	Within 4 wks of V5 (±1 wk)	Morning after V10	Within 2 wks of V10	Within 4 wks of V10	1-7 days before V15	Within 4 wks of V10 (±1 wk)	Morning after V15	Within 2 wks of V15	Within 4 wks of V15	8 wks post V15 (±1 wk)	2 days post V19 (-1/+7 days)
Past/Current Medication & Adherence	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓		✓
Drug Screen	✓					✓					✓					
Pregnancy Screen	✓					✓					✓					
Record	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
All AEs [C]	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓		✓
Randomization [E]	✓															
Container Assignment [E]	✓					✓					✓					
Administer IP	✓					✓					✓					
8-hour Exp. Session	✓					✓					✓					
BP, Pulse, Temperature [F]	✓					✓					✓					✓ [G]
Overnight Stay	✓					✓					✓					
90-min Integrative Session		✓	✓		✓		✓	✓		✓		✓	✓	✓		
Phone Call Follow-up [H]		✓					✓					✓				
Weight																✓

[C] All Adverse Events (AEs) includes collecting Serious Adverse Events, AEs of Special Interest, AEs of Psychiatric Status, AEs requiring medical advice or attention, AEs that indicate withdrawal of a participant, and all other AEs [E] Randomize 24 to 48 hours prior to first Experimental Session; obtain container assignment 24 to 48 hours prior to each Experimental Session [F] During Experimental Sessions, vitals are measured before Investigational Product administration, immediately before the supplemental dose is administered (or would be, if supplemental dose not given), and approximately 8 hours after initial dose, and as needed [G] At Study Termination, only blood pressure needs to be measured [H] Four days of phone call follow-up: Day 2 and 7 after the Experimental Session, with two additional calls in between. [I] All Visits must be scheduled to ensure that the CAPS-5 T4 assessment is within the window provided

		Screening		Baseline & Enrollment Confirmation		Treatment 1				Treatment 2				Treatment 3			Follow-up & Study Termination	
	Visit #	Site [A]	IR Screening	IR V3	V4	V5	V6&7	IR V8	V9	V10	V11&12	IR V13	V14	V15	V16&17	V18	IR V19	V20
Visit Description	**~Time to Complete Measure (minutes)**	Site Visit	Tele-medicine	Tele-medicine	Site Visit	Exp. Session 1	Int. Sessions 1.1 & 1.2	CAPS-5 T2: Tele-medicine	Int. Session 1.3	Exp. Session 2	Int. Sessions 2.1 & 2.2	CAPS-5 T3: Tele-medicine	Int. Session 2.3	Exp. Session 3	Int. Sessions 3.1 & 3.2	Int. Session 3.3	CAPS-5 T4: Outcome: Tele-medicine	Study Termi-nation
CAPS-5	90 (Baseline) 60 (all others)			✓				✓				✓					✓	
SDS	2			✓				✓				✓					✓	
C-SSRS [B]	10	✓	✓		✓	✓[C]	✓		✓	✓[C]	✓		✓	✓[C]	✓	✓		✓
MINI	15		✓															
SCID-5-SPQ	20	✓																
SCID-5-PD	60		✓															
PCL-5	8	✓																
LEC-5	5	✓							✓				✓			✓		✓
DDIS [D]	5		✓															
ACE	4				✓													
BDI-II	10 (Baseline) 5 (all others)				✓													✓
CPGS	5				✓													✓
DSP-I	15			✓				✓				✓					✓	
EQ-5D-5L	3				✓													✓
IASC	15				✓													✓
IPF	10				✓				✓				✓			✓		✓
PSQI	10				✓													✓
SCS	6				✓													✓
TAS-20	5				✓													✓
AUDIT	3	✓																✓
DUDIT	3	✓																✓
SRNU	3				✓													✓
EAT-26	6				✓													✓
HPQSF	5				✓													✓
UFEC	3				✓													
~Total Time of Completing Measures (minutes)		**49**	**90**	**107**	**95**	**20**	**10**	**77**	**15**	**10**	**10**	**77**	**15**	**10**	**10**	**15**	**77**	**97**

[A] Ensure that LEC-5 and SCID-5-SPQ results are sent to the Independent Rater who will be conducting the SCID-5-PD
[B] First C-SSRS is a Lifetime assessment, other assessments are Since Last Visit
[C] Conducted pre- and post-Investigational Product administration, and at phone calls on Days 2 and 7 after Experimental Session
[D] The relevant questions (117-130) from the DDIS will be asked by the Independent Rater during the SCID-5-PD assessment. The entire measure will never be administered.

6.0 Randomization and Blinding

Eligible participants will be enrolled in the study and sequentially assigned an identification number. Participants will be assigned to MDMA or placebo treatment groups via an Interactive Web Randomization System (IWRS) based on a randomization schedule developed by an independent third-party vendor to maintain blinding. Randomization will be stratified by clinical site.

To minimize bias, protect the study's double-blind and to ensure data quality, the sponsor plans to use a second Electronic Data Capture (EDC) database that is dedicated to the collection of critical primary and key secondary outcome measures, including the Total Severity Score on the CAPS-5 and item scores on the Sheehan Disability Scale (SDS), administered by the centralized blinded Independent Rater (IR) Pool through live video. This second database is termed the Independent Rater Database (IRDB) and it will be separate from the blinded, clinical EDC database in order to ensure that site and sponsor staff engaged in study conduct are masked from study outcomes. The IRDB will only be accessible by: (1) qualified, observer-blind individuals who are in the established IR Pool, (2) the Senior IR responsible for oversight and data quality of the IR Pool, and (3) the IR Coordinator responsible for data entry based on paper Source Records completed by the IR Pool. All CAPS-5 and SDS scores and supporting documentation will be reviewed by the IR Coordinator prior to data entry.

Once data is entered into the IRDB, the Senior IR will oversee the data quality by centrally by reviewing the critical data. The monitoring conducted by the CRA will utilize the following strategies: 1) To perform remote, source data review of the IRDB to check logic and to confirm reliability between the IR's, and 2) To perform remote source data verification to check the IRDB for administrative, data entry, or transcription errors made by the Independent Rater Coordinator. The IR Program Lead will review data listings for quality and completeness prior to database lock.

7.0 Sample Size and Power Considerations

The statistical power calculations were made by fitting an MMRM model to CAPS-4 data from the Phase 2 study MP1 data to obtain covariance parameter estimates. CAPS-4 data was also converted to the CAPS-5 scale, by dividing the CAPS-4 score by 34 and multiplying by 20 to obtain the CAPS-5 score, and both methods came up with the same level of statistical power. The following inputs created the 90% statistical power estimate using the PASS 14 Mixed Models Repeated Measures (MMRM) module:

a. N=47 per group (allowing for 3 dropouts per treatment group)
b. Alpha = 0.0499
c. 1000 simulations
d. Change from Baseline means of 5.5, 12.8, and 20.5 for placebo, and
e. Change from Baseline means of 17.2, 31.1, and 37.4, for MDMA at Experimental Session 1, 2, and 3, respectively
f. Two levels of treatment factor (MDMA, Placebo)
g. Three within subject levels (3 treatment visits)
h. Subject level variance = 335
i. Variance of R (diagonal elements) = 280

j. Rho = 0.5
k. Unstructured covariance

Additional participants may be added to the sample size at the recommendation of the Data Monitoring Committee (DMC) from the results of an administrative interim analysis (Section 8.9) occurring after 60% of the mITT participants have completed a final assessment of the Primary Outcome (Visit 19) and terminated treatment.

8.0 Analyses

Every effort will be made to ensure complete, accurate and timely data collection and to avoid missing data, to ensure the completeness of the data which can impact the integrity and accuracy of the final study analysis. The statistical analyses will be reported using summary tables, figures, and data listings. All analyses and tabulations will be performed using SAS® Version 9.4 or higher, and R. In general, nominal variables will be described in terms of frequencies and percentages. Ordinal and non-normal continuous variables will be described using sample median and range. Approximately normal variables will be described using sample mean and standard deviations. A subset of demographic and baseline characteristics (age, sex, duration of PTSD, CAPS-5 at Baseline, SDS at Baseline, comorbid major depression, Adverse Childhood Experience Questionnaire [ACE] at Baseline, C-SSRS Lifetime Ideation Score, and prior MDMA use) will be tested formally using chi-squared tests for categorical variables and either Wilcoxon rank sum tests for non-normal continuous variables, t-tests for normal continuous variables with two categories. Except for the primary efficacy analysis (specified below), all statistical tests will be two-sided and a difference resulting in a p-value of less than or equal to 0.05 will be considered statistically significant. All p-values will be rounded to and displayed in four decimals. If a p-value of less than 0.0001 occurs it will be shown in tables as<0.0001. Data not subject to analysis according to this plan will not appear in any tables or graphs, but will be included in the data listings. Selected results may be presented graphically using standard graphical software.

Post-hoc exploratory analyses not identified in this SAP may be performed to further examine the study data. These analyses will be clearly identified as such in the final clinical study report.

8.1 Analysis Sets

- *Modified Intent-To-Treat* (*m*ITT): all randomized participants who receive IP in at least one blinded Experimental Session (Visit 5) and have at least one follow-up CAPS-5 assessment post-treatment
- *Per Protocol* (PP): all randomized participants who meet eligibility criteria, who receive IP in three Experimental Sessions, and have three follow-up CAPS-5 assessments post-treatment
- *Not Per Protocol* (NPP): all participants who are included in the mITT set but not the PP set
- *Safety*: all participants who receive any IP
- *All Enrolled*: all participants who sign informed consent and are initially enrolled
- *All Screened*: all participants who sign informed consent and are screened

Summaries of demographics, disposition, exposure and safety parameters will only be generated for the mITT Set if different from the Safety Set.

8.2 Missing Data Handling

All possible procedures within Good Clinical Practice (GCP) will be used to minimize Post-randomization Early Terminators. Based on MAPP1 data, it is expected that up to 5 - 10% of enrolled and randomized participants will discontinue study treatment early. Post-randomization Early Terminators will be compared to the Completers using baseline demographics and CAPS-5 Total Severity Score at Baseline. For the Post-randomization Early Terminators, data collection by IRs will continue on the same schedule as planned through Study Termination visit procedures in order to limit missing data. All observed CAPS-5 data up to the point of discontinuation of treatment will be included in the MMRM model of the *de jure* estimand from Post-randomization Early Terminators. Participants will not be replaced, and enrollment and treatments will continue until N=100 participants are obtained in the mITT set. All observed CAPS-5 data from Post-randomization Early Terminators collected prior to and after treatment discontinuation will be included in the supportive effectiveness analysis of the *de facto* estimand.

8.2.1 Partial or Missing Dates

The following conventions will be used to impute missing portions of dates for AEs and concomitant medications. Note that the imputed values outlined here may not always provide the most conservative date. In those circumstances, the imputed value may be replaced by a date that will lead to a more conservative analysis.

Start Dates:

1. If the year is unknown, then the date will not be imputed and will be assigned a missing value.
2. If the month is unknown, then:
    a. The month and day of the first dose date will be imputed if the year matches the first dose date year.
    b. Otherwise, ‘January’ will be assigned.
3. If the day is unknown, then:
    a. The day of the first dose date will be imputed if the month and year match the first dose date month and year.
    b. Otherwise, the first day of the month will be assigned.

Stop Dates:

1. If the year is unknown, then the date will not be imputed and will be assigned a missing value.
2. If the month is unknown, then ‘December’ will be assigned or the month of study termination if December is later than the subject’s study termination date.
3. If the day is unknown, then the last day of the month will be assigned or the day of study termination if the last day of the month is later than the subject’s study termination date.

8.3 Protocol Deviations

The number of participants in each protocol deviation category will be summarized by analysis set, treatment group, and protocol deviation severity. Individual participants will appear in a listing. Protocol deviation severity is defined as:

- *Major*: A type of deviation from the protocol that has the potential to significantly affect participant rights, safety, or well-being, or impact the primary or key secondary efficacy endpoint(s) for that subject.
- *Minor*: A type of deviation from the protocol that does not have the potential to significantly affect subject rights, safety, or well-being, or does not have potentially significant impact on the primary or key secondary efficacy endpoint(s) for that subject.

8.4 Baseline Values

Baseline values are from Baseline Visits (Visit 3 or Visit 4) for all measures, except for vital signs. For height, weight and BMI, results collected at Screening will be used as the Baseline value. For blood pressure (BP), pulse, and temperature, baseline is assigned as the pre-dose timepoint of the first experimental session (Visit 5).

8.5 Participant Disposition

The All Screened Set will be included in the summary of participant disposition and accountability. The tabulation of number of participants in each treatment group and overall will be displayed for each analysis set and for the Overnight Stay and No Overnight Stay subgroups. The number and percent of participants who completed or discontinued the study will be displayed for each treatment group and overall together with reasons for Study Termination, where the percent is with respect to the total number of screened participants in that treatment group.

8.6 Demographics and Baseline Characteristics

Participant demographic data and Baseline characteristics will be summarized for the mITT and Safety Sets descriptively by treatment group and overall. Trauma characteristics from the PCL-5 at screening will also be summarized by treatment group and overall.

8.7 Medical History and Substance Use

Medical history terms will be coded using MedDRA Version 24.0 and summarized by system organ class (SOC), preferred term (PT), and treatment group. Subjects with any history of prior therapy will also be summarized by treatment group and overall by the categories listing on the CRF. Substance use, ecstasy use history, and qualitative drug screening during the study will also be summarized by treatment group.

8.8 Efficacy Analyses

For all primary and secondary endpoints, descriptive statistics (n, mean, standard deviation, median, range, or counts and percentages where appropriate) will be provided

by treatment group. Longitudinal CAPS-5 Total Severity Score and mean SDS item scores will be plotted across visits to characterize the onset of treatment effect.

8.8.1 Primary Efficacy Analyses

The *de jure* estimand of treatment efficacy will be used to estimate the causal effect of MDMA-assisted therapy on PTSD symptom severity in the intended population of patients with PTSD from any cause (Estimand 1 in Table 4). All efficacy analyses will be based on the *m*ITT analysis set. The primary treatment comparison will be made at a 2-sided, 0.0499 level of alpha. The primary estimator of effects of initially randomized treatments will be the difference between groups in mean change in CAPS-5 Total Severity Scores from Baseline to 18 weeks after randomization (Visit 19). Least squares (LS) means from a restricted maximum likelihood (REML)-based mixed model for repeated measures (MMRM) will be used to compare treatment groups at Visit 19. LS means, 95% confidence intervals (CI), LS mean difference, and 95% CI for the difference will be displayed for each treatment group at Visits 8, 13, and 19. The p-value for the LS mean difference in treatment groups will be displayed at Visit 19 only. CAPS-5 data collected after treatment discontinuation will be considered to be missing at random (MAR) and will not be included in the *de jure* estimand. Missing data will not be imputed.

The model is specified as follows: Let Yij be the change from Baseline in CAPS-5 score from the ith participant at the jth visit, i = 1, …, n and j = 1, 2, 3, corresponding to post-Baseline Visits 8, 13, and 19, respectively. The treatment variable Xi is the randomized treatment of the ith participant. The variables T_2 and T_3 are variables indicating if the data from that record are from Visit 13 or 19, respectively. Covariates included in the MMRM model are treatment, visit by treatment interaction, baseline CAPS-5 Total Severity Score, a binary variable indicating dissociative subtype at Baseline, and investigative site. The regression model is specified as follows:

$$Y_{ij} = \alpha + \beta_1 X_i + \beta_{22} T_2 I(j=2) + \beta_{23} T_3 I(j=3) + \beta_{32} T_2 X_i I(j=2) + \beta_{33} T_3 X_i\, I(j=3)$$
$$+ \beta_4 \text{Baseline\_CAPS}_i + \beta_5 DS_i + \beta_6 \text{Site}_i + \varepsilon_{ij},$$

where I(j=J) is the indicator function that equals 1 if j=J and 0 otherwise. We assume that the Yij follow a Normal distribution with var(Yij) = σj2, and cov(Yij, Yij′) = σjj′=σj′j (unstructured correlation). The model above has been parameterized with reference cell coding, where the coefficient estimate of β_1, corresponding to the treatment variable, X, captures the mean difference in the change from Baseline in CAPS-5 for MDMA vs. placebo at Visit 8. The coefficient estimates of β_{22} and β_{23} capture the difference in mean change in the CAPS-5 at Visits 13 and 19 from the mean scores at Visit 8, respectively. The coefficient estimates of β_{32} and β_{33}, corresponding to the interaction of the treatment variable and the visit variable, capture the incremental mean difference in the change from Baseline in CAPS-5 for the MDMA group at Visits 13 and 19 from Visit 8 versus the placebo group. The coefficient estimate of β_4 captures the effect on the change from Baseline in CAPS-5 corresponding to the baseline CAPS-5 score. The coefficient estimate of β_5 captures the effect on the change from Baseline in CAPS-5, corresponding to dissociative subtype at Baseline. The coefficient estimate of β_6 captures the difference

in mean change from Baseline in CAPS-5 due to site differences. The following SAS code will implement this model and the statistical test for the MDMA vs. placebo comparison at Visit 19.

```
proc mixed;
    class ID X T Site DS;
    model CAPS = X  T  X*T Baseline_CAPS DS Site / s  DDFM=SATTERTH ;
    repeated T / subject=ID type=un;
    lsmeans trtp trtp*avisitn / diff cl om alpha=.0499 ;
    ods output tests3=tests3;
    ods output Diffs=Diffs;
run;
```

where, the CAPS variable is the change from Baseline in the CAPS-5 Total Severity Score at Visit 8, 13, or 19, X is the randomized treatment assignment (MDMA or placebo), T is Visit (Visit 8, Visit 13, or Visit 19), Baseline_CAPS is the baseline CAPS-5 Total Severity Score for each participant, DS is the dummy variable indicating dissociative subtype at Baseline, Site indicates the clinical site where the participant was treated, and ID uniquely identifies the participant. If the unstructured variance/covariance model assumption doesn't allow the REML algorithm to converge, a heterogeneous Toeplitz structure will be used. If the model doesn't converge under heterogeneous Toeplitz, the homogeneous Toeplitz structure will be used. If that fails to converge, the next variance/covariance attempted will be AR(1). In the event AR(1) structure fails to converge, the compound symmetry assumption will be used. Cohen's d and associated 95% confidence intervals will also be presented to describe the difference in the mean change from Baseline between treatment groups.

Table 4: Proposed Estimands and Key Attributes

Estimand	Endpoint	Hypothesis	Inference	Population	Endpoint	Use of data after early treatment discontinuation in analysis
1	*CAPS-5*	*de jure*	Initially randomized treatment	*m*ITT	Planned endpoint	Not included
2	*CAPS-5*	*de facto*	Treatment Policy	*m*ITT	Planned endpoint	Included
3	*SDS*	*de jure*	Initially randomized treatment	*mITT*	Planned endpoint	Not Included

8.8.2 Supportive Analyses:

To provide information on treatment effectiveness in all randomized participants that could be expected in clinical practice, regardless of adherence to the 3-session course of MDMA-assisted therapy treatment, a *de facto* estimand is included as a supportive analysis (Estimand 2 in Table 4). The *de facto* estimand will include all available CAPS-5 outcome data from all participants in the mITT, including CAPS-5 data collected after treatment discontinuation for Post-randomization Early Terminators. This will be estimated in the mITT set by a statistical comparison between treatment groups using the

MMRM model with a least squares estimate of the change from Baseline in CAPS-5 to 18 weeks after randomization. No *de facto* sensitivity analyses will be performed for the key secondary efficacy endpoint.

The effect of departures from choices and assumptions made for the primary analysis will also be tested as sensitivity analyses for the primary estimand. These are listed below.

a. Distributional assumptions for the residuals of the model will be tested using Shapiro Wilks test and a Q-Q plot; if the distribution is found to be non-normal, we will test a transformation of the data, such as the log transformation, to achieve normality.
b. Effect of covariates will be tested by conducting the primary analysis with and without: (1) dissociative subtype of PTSD based on CAPS-5 and (2) Baseline CAPS-5 Total Severity Score.
c. To examine the robustness of the inferences to the MAR assumption, a worst/best-case imputation technique will be employed where each missing CAPS-5 measure from MDMA treated subjects from each visit will be replaced by the worst observed result among the MDMA subjects at that visit. Each missing CAPS-5 measure from Placebo treated subjects at each visit will be replaced by the best observed measure among Placebo treated subjects at that visit. 1000 complete datasets will be generated by sampling from the data with missing CAPS-5 measures replaced by worst/best case imputation using the MI procedure in SAS; each of which will be analyzed using the primary efficacy model with effects for treatment arm, visit, and covariates. The final results will be obtained by combining the LS means and LS mean differences from these 1000 analyses by applying Rubin's rules [3].
d. If the worst/best case imputation analysis results in the treatment effect estimate to have a p-value greater than 0.0499, a tipping point analysis will be executed. The tipping point analysis will be conducted to estimate the treatment difference in the CAPS-5 scores where model results become statistically non-significant. This sensitivity analysis assumes that all missing CAPS-5 data are Missing Not at Random (MNAR). This analysis will only be performed if the primary analysis results differ from the worst/best case sensitivity analysis. Multiple imputation will be used to impute missing CAPS-5 Visit 13 and/or Visit 19 data. If the data has monotone missingness pattern, then monotone regression procedure be used in SAS PROC MI [4] to impute missing Visit 13 CAPS-5, using Visit 8 CAPS-5, dissociative subtype, site, and treatment as covariates. Then, missing Visit 19 CAPS-5 will be imputed in a similar manner, replacing Visit 8 CAPS-5 with Visit 13 CAPS-5 as a predictor. If data does not have monotone missingness pattern, then fully conditionally specified (FCS) methods will be used for the imputation.

   The imputed MDMA-assisted therapy results will then be penalized by Δ and estimates will be generated from these penalized datasets. Final results for Δ penalized datasets will be combined from 1000 imputation datasets by applying Rubin's Rules. Delta will be increased in increments of x until the final model result p-value crosses the threshold of 0.0499.
e. An additional sensitivity analysis will address the effect of intercurrent events [1, 2, 5], which are identified in this study as:
   - Discontinuation of study because subject withdrew consent
   - Post-Randomization Early Termination and does not take medication for PTSD after termination

- Post-Randomization Early Termination and does take medication for PTSD after termination
- Death, not suicide
- Suicide

This analysis assesses the robustness of inferences from the primary model (which assumes missing data are MAR) by assuming each category of intercurrent events are MNAR. This is done by assigning missing CAPS-5 data to the estimated mean of the MDMA or placebo treatment groups (Table 5:).

Table 5: Inference of Treatment Effect for Missing Data from mITT Set Based on Reasons for Post-Randomization Withdrawal

Post-randomization Events	Available Data Post-Withdrawal	Treatment Assignment	Missing CAPS-5 Imputation
Discontinue Study, Withdraw Consent	No data	MDMA	Assume treatment failure • Assign missing value to Placebo estimated mean at corresponding visit
		Placebo	Assume treatment failure • Assign missing value to Placebo estimated mean at corresponding visit
Post-Randomization Early Termination (or >8 week delay between experimental sessions) and does not take medication for PTSD after discontinuing intervention	Data collection continues whenever possible via remote visits	MDMA	Assume treatment success if data supports treatment response Assign missing value to MDMA estimated mean at corresponding Visit Assume treatment failure if data does not support treatment response • Assign missing value to Placebo estimated mean at corresponding Visit
		Placebo	Assume treatment failure • Assign missing value to Placebo estimated mean at corresponding Visit
Post-Randomization Early Termination (or >8 week delay between experimental sessions) and does take medication for PTSD after discontinuing intervention	Data collection continues whenever possible via remote visits	MDMA	Assume treatment success if data supports treatment response • Assign missing value to MDMA estimated mean at corresponding Visit Assume treatment failure if data does not support treatment response • Assign missing value to Placebo estimated mean at corresponding Visit
		Placebo	Assume treatment failure • Assign missing value to Placebo estimated mean at corresponding Visit
Death, not suicide	No data	MDMA	Assume treatment success if last available data support treatment response • Assign missing value to MDMA estimated mean at corresponding Visit Assume treatment failure last available data does not support treatment response • Assign missing value to Placebo estimated mean at corresponding Visit
		Placebo	Assume treatment failure • Assign missing value to Placebo estimated mean at corresponding Visit
Suicide	No data	MDMA	Assume treatment failure • Assign missing value to Placebo estimated mean at corresponding Visit

Post-randomization Events	Available Data Post-Withdrawal	Treatment Assignment	Missing CAPS-5 Imputation
		Placebo	Assume treatment failure • Assign missing value to Placebo estimated mean at corresponding Visit

Additional baseline covariates (including but not limited to: age, gender, ethnicity, index trauma, complexity and severity of trauma, medication tapering, diagnosis of comorbid depression, diagnosis of comorbid Axis 2 diagnosis, adverse childhood experiences) may be assessed for inclusion in the primary efficacy model at a $p<0.05$ threshold. Each covariate and the corresponding treatment by covariate interaction term will be included in the primary efficacy analysis model one-at-a-time and the p-value of the term and its interaction with treatment with be reported. Covariates with either the estimate of the term or the covariate by interaction term will be included in a full model that includes all covariate/interaction terms that were significant in the primary efficacy model (i.e. the model with only covariate, covariate by treatment, treatment, and visit). Backward elimination will be used to compile a full model.

8.8.3 Key Secondary Efficacy Analyses

For the key secondary analysis of efficacy, the *de jure* estimand will be used to estimate the causal effect of MDMA-assisted therapy on PTSD on the SDS in the intended population of patients with PTSD from any cause (Estimand 3 in Table 4). The SDS is a 3-item scale measuring the severity of disability in the domains of work, family life/home responsibilities, and social/leisure activities. Each of these three domains is scored on a ten-point Likert scale, where a score of 0 is 'not at all impaired', 5 is 'moderately impaired', and 10 is 'very severely impaired'. The summary measure used to analyze the treatment effect on SDS will be the mean of the 3 item responses at each visit. Any participant who did not work during the reporting period due to reasons related to PTSD will be scored as a 10 on Item 1. In cases where participants did not work due to reasons unrelated to PTSD, the score for Item 1 (work/school) will be imputed by averaging the scores of items 2 and 3. If both items are checked for Item 1, it will be scored as a 10. It is expected that no more than 5% of participants will have Item 1 (work/school) imputed, based on prior studies of paroxetine with PTSD. To limit missing data and ensure standardized administration, the SDS will be administered in a clinician-rated format during the same visits as the CAPS-5 by the centralized IR Pool. The clinician-rated administration will be important in limiting misinterpretation of the work/school impairment item that could contribute to elevated rates of skipping out, which can also be completed in reference to school, unpaid or volunteer work, and may not immediately come to mind in a participant-reported format.

If more than 5% of item 1 data in a treatment group is missing at Visit 19, a sensitivity analysis will be conducted by using the sum of the items without imputing missing Item 1 scores and compared to the results from the average SDS with imputation.

A hierarchical testing strategy will be employed to control for Type-I error. That is, the hypothesis for the Key Secondary Endpoint (SDS) will only be tested if the statistical test for the Primary Efficacy comparison rejects the null hypothesis. If the statistical test for

the Primary Efficacy comparison does not reject the null hypothesis, the analysis of the Key Secondary Endpoint (SDS) will be exploratory. The key secondary estimator of effects of initially randomized treatments will be the difference between groups in mean change from Baseline in SDS summary measure at 18 weeks after randomization (Visit 19). The key secondary efficacy analysis will be conducted using the same methodology as the primary efficacy measure, but replacing Baseline CAPS-5 with Baseline SDS as a covariate in the model.

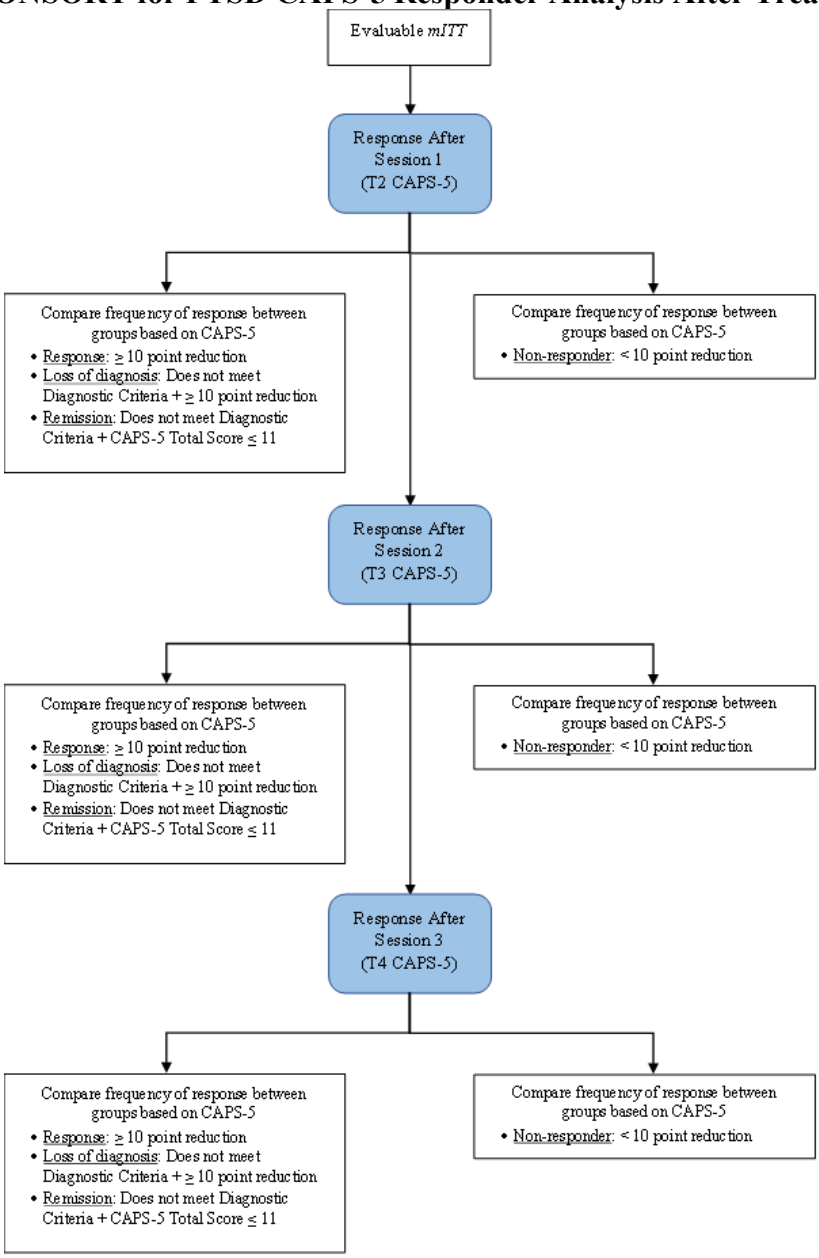
8.8.4 Responder Analyses

The CAPS-5 produces a Total Severity Score based on severity of PTSD symptom domains described in the DSM-5, as well as a categorical rating indicating whether a participant meets PTSD diagnostic criteria. A psychometric validation study found the following severity score ranges for the CAPS-5: Asymptomatic (0-10), Mild (11-22), Moderate (23-34), Severe (35-46), Extreme (47+). Based on these data, a 10-point reduction in CAPS-5 Total Severity Score is clinically meaningful. Four different responder categories will be derived and compared descriptively at each post Experimental Session visit using frequencies and percentages. The categories (Figure 1) are:

- Non-responder: <10 point reduction in CAPS-5 Total Severity Score since Baseline
- Response: 10 point or greater reduction in CAPS-5 Total Severity Score since Baseline
- Loss of Diagnosis: 10 point or greater reduction in CAPS-5 Total Severity Score since Baseline and no longer meeting PTSD diagnostic criteria on CAPS-5
- Remission: CAPS-5 Total Severity Score of 11 or less and no longer meeting PTSD diagnostic criteria on CAPS-5

Note that these groups are not independent. An exploratory analysis will also be conducted using a 12 point reduction in CAPS-5 scores.

Figure 1: CONSORT for PTSD CAPS-5 Responder Analysis After Treatment



[1] Response at T2, T3, and T4 will be in comparison to Baseline at T1.

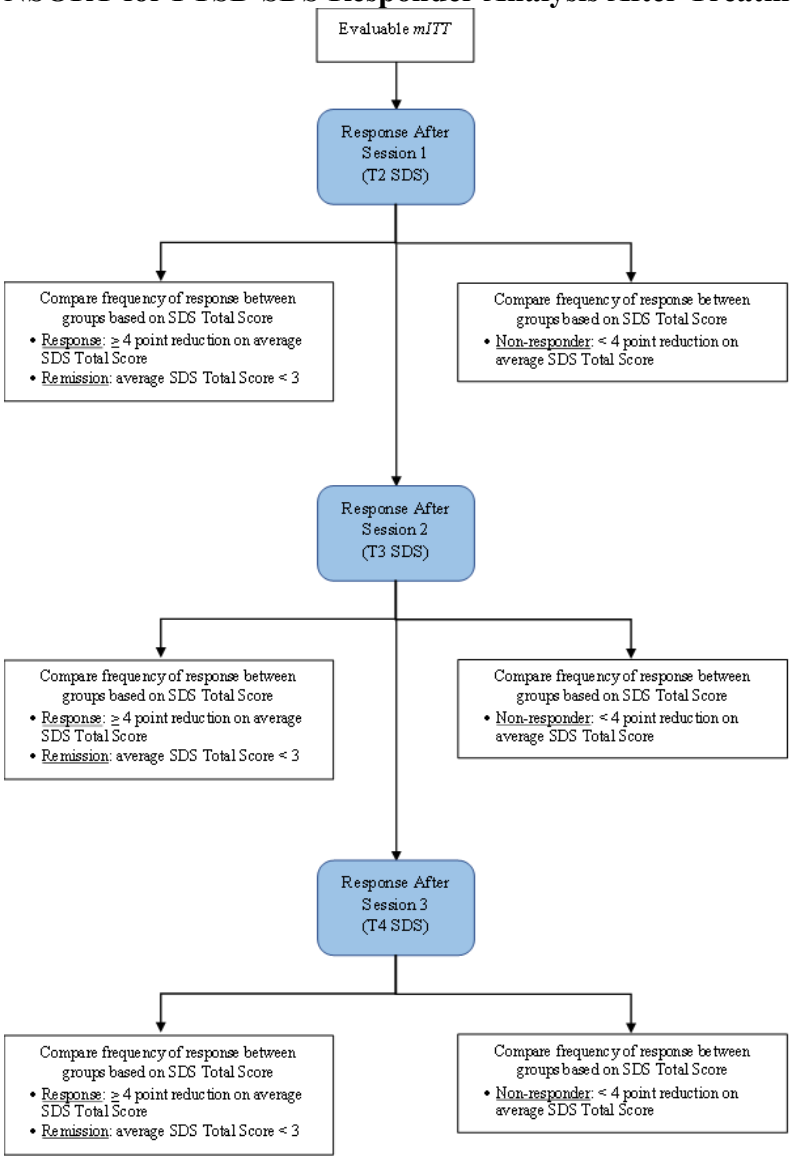
The sponsor will measure changes in functional impairment with the SDS throughout treatment at the same visits as the CAPS-5 in this study. The authors of the SDS have provided the following score ranges to describe functional impairment severity for each of three items describing the functional domains of work, social life, and family life: No impairment (0), Mild impairment (1-3), Moderate impairment (4-6), Marked impairment (7-9), Extreme impairment (10). Based on the psychometric development publication [4] and prior paroxetine studies for treatment of PTSD, a 4-point mean reduction across SDS items would represent a clinically meaningful treatment response, as this would correspond to downward change from Extreme to Moderate, Marked to Moderate, or Moderate to Mild in severity of any domain of functional impairment. If a participant drops to a mean score of less than 3 across SDS items, this would be defined as a

functional remission indicating Mild or No impairment. Three different SDS responder categories will be derived and compared descriptively at each post Experimental Session visit using frequencies and percentages. The categories are:

- Non-responder: <4 point reduction in the average SDS Total Score since Baseline
- Response: 4 point or greater reduction in the average SDS Total Score since Baseline
- Remission: average SDS Total Score is less than 3 at the visit

Note that these groups are not independent. The exploratory responder analysis that was used for CAPS-5 will be conducted using the SDS measure and the thresholds described below (Figure 2). If more than 5% of item 1 data in a treatment group is missing at a visit, then SDS responder categories will also be derived using the sum of the items without imputation (see Section 8.8.3).

Figure 2: CONSORT for PTSD SDS Responder Analysis After Treatment



[1] Response at T2, T3, and T4 will be in comparison to Baseline at T1.

8.9 Safety Analyses

Safety assessment analyses for this study will include summaries of drug exposure, unsolicited adverse events, concomitant medications, suicidal ideation, intensity, and behavior from the C-SSRS, and vital signs.

8.9.1 Analysis of Exposure

The frequencies and percentages of participants with exposure will be summarized overall and by treatment group. The total dose administered (either initial with no supplemental or initial plus supplemental) and elapsed time (minutes) from primary to supplemental dose (if applicable) will be produced along with summary statistics for each Experimental Session. The total dose administered over all Experimental Sessions and the average time from primary to supplemental dose for each subject across the Experimental Sessions will also be summarized. Data will be tabulated for the Safety Set.

8.9.2 Analysis of Adverse Events

The primary measure of safety will be the reporting of unsolicited AEs. All AEs collected from Enrollment to Study Termination will be categorized as follows:

- Pretreatment AEs – AEs that occur during Preparatory Period prior to first dose in first Experimental Session
- Treatment Emergent AEs – AEs that occur during Treatment Period from first Experimental Session to last Integrative Session.
- AEs that occur on Days 0, 1 and 2 after MDMA or placebo administration
- AESIs – AEs specified in the protocol related to cardiac function and abuse liability.
- Follow-up Period AEs – AEs that occur during Follow-up Period after last Integrative Session through Study Termination.
- AEs leading to discontinuation of IP
- AEs resulting in death or hospitalization
- SAEs
- AEs continuing at Study Termination

Verbatim terms on case report forms will be mapped to preferred terms (PT) and system organ class (SOC) using the Medical Dictionary for Regulatory Activities (MedDRA) Version 24.0. Frequency and incidence of AEs will be displayed and sorted by SOC then PT and summarized by treatment group, category (as defined above), severity, and seriousness. AEs will be analyzed and presented as follows:

- If a participant has more than one AE mapped to the same PT, that AE will be reported once using the highest severity.
- Relationship will be determined based on relative incidence of TEAEs with at least two-fold difference between MDMA vs. placebo.
- Compare relative incidence of AEs during Experimental Sessions such as clinical signs and symptoms, such as chest pain, shortness of breath, or neurological symptoms or any other signs or symptoms that may be indicative of a medical complication of the Investigational Product.

- AEs that occur on day of Experimental Sessions up to 2 days after IP administration will be presented separately.

TEAEs with a preferred term of "Suicidal Ideation" will be summarized by history of suicidal ideation from the Lifetime C-SSRS at Screening (Yes or No). The time to onset (days post experimental session), duration of the events (days) will be summarized by treatment group. The number of events not resolved with also be presented.

8.9.3 Concomitant Medications

A secondary measure of safety will be the reporting of concomitant medications. All concomitant medications collected from Screening to Study Termination will be categorized as follows:

- Pretreatment medications are defined as medications taken prior to and after signing informed consent and those taken during the Preparatory Period prior to the first Experimental Session. A stop date is expected prior to the first Experimental Session for any medications requiring a change in dose, a skipped dose, or tapering.
- Treatment Period concomitant medications are defined as those taken or continued during the Treatment Period from the first Experimental Session to the last Integrative Session. A stop date is expected prior to each Experimental Session for any medications requiring a change in dose, a skipped dose, or tapering.
- Concomitant medications include those with a start date of the day of and up to 2 days after IP administration.
- Follow-up Period concomitant medications are defined as those taken or continued during the Follow-up Period after the last Integrative Session through Termination.
- Any concomitant medications that are tapered.
- Any concomitant medications that are taken to treat an AE.
- Any concomitant medications that are taken to treat an SAE.
- Any excluded concomitant medications taken as a deviation from the protocol.

Concomitant medications on case report forms will be classified using the WHO Drug Dictionary Global (WHODrug Global). Frequency and incidence of concomitant medications will be summarized by treatment group, Anatomical Therapeutic Chemical (ATC) Class Level 1, ATC Class Level 3, preferred term, and indication. Concomitant medications will be analyzed and presented as follows:

- Pretreatment medications
- Prior and concomitant medications
- Concomitant medications taken on the day of and up to 2 days after IP administration will be presented separately
- Any central nervous system concomitant medications by period

Other categories defined in this section will be presented in listings.

8.9.4 Analysis of C-SSRS

Suicidal ideation, ideation intensity, and behavior will be summarized according to suggestions made in the C-SSRS Scoring and Data Analysis Guide [8]. A positive

response for suicidal ideation is counted when a participant answers "yes" to any one of the five suicidal ideation questions (Categories 1-5) on the C-SSRS (i.e., a score>0 for suicidal ideation). Serious suicidal ideation is a suicidal ideation score of 4 or 5. A positive response for suicidal behavior occurs when a participant answers "yes" to any one of the five suicidal behavior questions (Categories 6-10) on the C-SSRS (i.e., a score>0 for suicidal behavior). The number and percent of positive responses of Positive Ideation, Serious Ideation, and Positive Behavior will be tabulated by treatment group and time period [lifetime, screening, Baseline, each Experimental Session (pre- and post-IP), Integrative Sessions, and endpoints]. Frequency and incidence of suicidal ideation, ideation intensity, and suicidal behavior will be presented by time period (listed above), treatment group and overall using descriptive statistics in tabular format. A shift table of suicidal ideation at Baseline (Visit 4) compared to the maximum post-treatment suicidal ideation score will be produced by treatment group.

Additionally, the number of subjects with suicide-related treatment emergent (TE) suicidal ideation and behavior based on the C-SSRS during treatment will be compared by treatment group for categories below. Treatment-emergent is defined as any C-SSRS measurement taken after the first dose of IP. The difference the proportions between treatment groups will be compared using 2-sided Fisher's exact tests and p-values will be presented for each category.

- TE suicidal ideation (1-5) compared to recent history = number of subjects with post-treatment C-SSRS score of 1-5 / number of subjects with at least one post-baseline suicidal ideation score where the baseline suicidal ideation score is < 5
- TE serious suicidal ideation (0-3 to 4-5) compared to recent history = number of subjects with post-treatment C-SSRS score of 4-5 / number of subjects with at least one post-baseline suicidal ideation score where the baseline suicidal ideation score is 0-3
- Emergence of serious suicidal ideation (0 to 4-5) compared to recent history = number of subjects with post-treatment C-SSRS score of 4-5 / number of subjects with at least one post-baseline suicidal ideation score where the baseline suicidal ideation score is 0
- Improvement in suicidal ideation at endpoint compared to baseline = number of subjects with post-treatment C-SSRS score is less than the baseline C-SSRS score / number of subjects with at least one post-baseline suicidal ideation score where the baseline suicidal ideation score is >0
- Emergence of suicidal ideation (1-5) compared to all prior history = number of subjects with post-treatment C-SSRS score is greater than the Lifetime suicidal ideation score / number of subjects with at least one post-baseline suicidal ideation score where the Lifetime suicidal ideation score <5
- Emergence of suicidal behavior (6-10) compared to all prior history = number of subjects with post-treatment suicidal behavior score of 6-10 / number of subjects with at least one post-baseline suicidal behavior score where the Lifetime suicidal behavior is less than 6

8.9.5 Analysis of Vital Signs

Vital signs (heart rate, BP, and body temperature) for Experimental Sessions will be summarized using descriptive statistics in tabular format listing values at pre-IP, interim (prior to the supplemental dose), and endpoint (at the end of each Experimental Session) by treatment group. Change from Baseline in vital signs will be described for each visit

and time point (predose, interim, and endpoint) by treatment group and overall. Summaries will also be produced by age group (< 65 years old and >= 65 years old). Shift tables will be generated by initial treatment dose (0 mg [placebo], 80 mg, and 120 mg), and age group for pre-dose values compared to the maximum post-dose value using 180 mmHg as a threshold for systolic blood pressure, 110 mmHg for diastolic blood pressure, and an increase of 1 degree C for temperature. The frequency and percent of subjects who had vital signs greater than or equal to the thresholds in the Observed Value and Change categories in Table 6 will also be summarized by age group and initial treatment dose.

Table 6: Shift Table Vital Sign Categories

Vital Sign	Observed Value	Change
Systolic Blood Pressure	>=180 mmHg <180 mmHg	Increase of >=40 mmHg
Diastolic Blood Pressure	>=110 mmHg <110 mmHg	Increase of >=25 mmHg
Temperature	>=38 degree C <38 degree C	Increase >= 1 degree C

8.10 Administrative Interim Analysis

An unblinded administrative interim analysis is planned to be performed at a single time point after the first 60% of participants in the mITT set have completed the final CAPS-5 follow-up assessment and terminated treatment, including early treatment discontinuation participants who have completed their final CAPS-5, and N=100 participants have been enrolled. The objective of the administrative interim analysis is to conduct a sample size re-estimation. The Key Secondary Endpoint will not be evaluated at the interim analysis. The detailed results of the interim analysis will not be revealed to the site staff, participants, IRs, or sponsor staff/designees except for a single designated sponsor representative from MAPS. The independent DMC statistician will perform the unblinded administrative interim analysis as described in the DMC Charter.

Following the DMC review of the MAPP1 results, the DMC may recommend changes to the purpose and criteria of the interim analysis. For instance, the DMC may recommend adding an efficacy interim analysis executed at a specific number of subjects with Primary Endpoint data. Regardless of the DMCs suggested changes, overall Type-I error will be preserved.

Sample size re-estimation: Based on the result of calculating conditional power using the estimated effect size from the first 60% of participants in the mITT set, the sample size may be increased. The theoretical table below shows the range of possible effect sizes that could be observed at the interim analysis and what the guidance from the DMC could be for the adjusted sample size. Note that this table is an example of possible effect sizes and sample size increases based on a simpler model than the actual model that will be used in the interim analysis.

Table 7: Observed Effect Sizes and Corresponding Sample Size for 90% Conditional Power Guidance at the Interim Analysis (N=60)

Effect Size	≤ 0.32	0.32	0.34	0.36	0.38	0.40	0.42	0.44
Mean Change	-	5.76	6.12	6.48	6.84	7.2	7.56	7.92
Conditional Power[1]	< 50%	52%	60%	67%	74%	80%	85%	89%
Increase in Sample Size to Restore 90% Conditional Power	No increase	240	180	140	110	70	40	10

[1] Conditional Power based on the alternative hypothesis effect size as observed in the first 60 subjects.

Note: this table is theoretical and not based on the true efficacy model. It is an example based on a simplier modeling strategy based on theoretical effect sizes.

For Table 6, the conditional power estimates and the numbers to increase the sample size to restore 90% conditional power were calculated in PASS version 14 using the module for two-sample t-tests, which is based on the conditional power definitions from Jennison and Turnbull [9]. Since the minimum effect size for all possible scenarios where the sample size would be increased achieves greater than 50% conditional power for N=60, the type-I error rate is preserved with this administrative sample size re-estimation method [10]. However, the sponsor has chosen to allocate 0.2% of the alpha (0.0001) for this unblinded administrative interim analysis to account for any possible downward bias in the variance estimate. At the time of the interim analysis, the independent unblinded statistician will fit the cleaned data on the first 60 participants to the primary analysis model to estimate the parameters. These parameter estimates will be input to the PASS 14 Mixed Model Repeated Measures module to re-estimate the power. If the power is less than 90% and more than 50%, the independent statistician will increase the sample size up to the point where the power reaches 90%. These results will be provided by the unblinded independent statistician to the DMC.

The DMC will review the group-unblinded interim analysis results provided by the independent statistician and will give recommendations to the designated sponsor representative from MAPS. The designated sponsor representative from MAPS will provide written information to the trial organizer delegate of MPBC indicating how many participants, if any, to add to the IWRS but will not provide any information on the corresponding conditional power the additional participants are being added to achieve. Only the re-estimation of sample size will be completed at the interim analysis, this will not be used to drop or add treatment arms or doses, change entry criteria, change randomization ratio or change Primary Endpoint. The sponsor representative from MAPS will not be part of any trial decisions or administration after the sample size re-estimation has been communicated to MPBC who will continue to be the trial organizer. If the DMC reports a sample size increase to the independent statistician DMC member, they will increase the randomization cap from 100 to the new sample size within the IWRS accordingly. Thereafter, the IWRS will cap randomization at the increased sample size

resulting from the interim analysis. If the interim analysis results in a sample size increase, it will not be revealed to the investigators, participants, or sponsor.

8.11 Subgroup Analysis

The primary efficacy analysis will be run on subgroups Overnight Stay/No Overnight Stay and Dissociative Subtype/No Dissociative Subtype (removing this as a covariate).If additional exploratory covariates are significant (Section 8.8.2), these will also be considered for subgroup analysis of the primary efficacy endpoint. Additionally, demographic and baseline characteristics, CAPS-5 responder analysis, and select adverse event tables will be produced by Overnight Stay/No Overnight Stay subgroups.

9.0 Data Monitoring Committee

A DMC with appropriate expertise in the conduct of PTSD clinical trials will act in an advisory capacity to monitor participant safety on a routine basis throughout the trial by reviewing safety and study data provided by an Independent Statistical Group. The DMC will also monitor individual participant tolerability. The composition of the DMC will include two clinician experts in PTSD clinical trials and a biostatistician.

The DMC may:

- Review the study protocol and informed consent documents and plans for data monitoring.
- Evaluate the progress of the trial, study data quality, timeliness, participant recruitment, accrual and retention, participants' risk versus potential benefit, and other factors that could affect the study outcome.
- Review the results of MAPP1 and make recommendations on updating the adaptive design elements to the study protocol, interim analysis, and/or data analysis of the MAPP2 study.
- Implement the adaptive design aspects of the study at the interim analysis by reviewing the interim analysis results and making the appropriate decision to the independent statistician on sample size re-estimation.
- Consider relevant information that may have an impact on the safety of the participants or the ethics of the study.
- Make other recommendations to MAPS concerning continuation, Study Termination or other modifications of the study based on their observations of the study.

A full description of the DMC duties will be detailed in the DMC Charter.

10.0 Timing of Analyses

The primary efficacy analysis will be conducted after all participants complete Visit 20 and the database is locked.

11.0 References

1. Tang, Y., *An efficient monotone data augmentation algorithm for multiple imputation in a class of pattern mixture models.* J Biopharm Stat, 2017. **27**(4): p. 620-638.
2. Molenberghs, G., et al., Analyzing incomplete longitudinal clinical trial data. Biostatistics, 2004. 5(3): p. 445-64.
3. Rubin, D.B., Inference and missing data. Biometrika, 1976. 63(3): p. 581-592.
4. SAS Help Center: The MI Procedure. SAS/STAT ® 15.2 User's Guide Last updated October 28, 2020.
5. Iddrisu, AK., Gumedze, F. *An application of a pattern-mixture model with multiple imputation for the analysis of longitudinal trials with protocol deviations.* BMC Med Res Methodol 19, 10 (2019). https://doi.org/10.1186/s12874-018-0639-y
6. *An application of a pattern-mixture model with multiple imputation for the analysis of longitudinal trials with protocol deviations* | BMC Medical Research Methodology | Full Text (biomedcentral.com)
7. Sheehan, K.H. and D.V. Sheehan, *Assessing treatment effects in clinical trials with the discan metric of the Sheehan Disability Scale.* Int Clin Psychopharmacol, 2008. **23**(2): p. 70-83.
8. Nilsson, M.E., et al., *Columbia Suicide Severity Rating Scale Scoring and Data Analysis Guide*, in *CSSRS Scoring Version 2.0*. 2013: http://www.cssrs.columbia.edu/documents/ScoringandDataAnalysisGuide_Feb2013.pdf.
9. Jennison, C. and B.W. Turnbull, *Group sequential methods with applications to clinical trials*. 2000, Boca Raton, FL: Chapman & Hall/CRC Press.
10. Chen, Y.H., D.L. DeMets, and K.K. Lan, *Increasing the sample size when the unblinded interim result is promising.* Stat Med, 2004. **23**(7): p. 1023-38.