STATISTICAL ANALYSIS PLAN (SAP) FOR COMP 001

THE SAFETY AND EFFICACY OF COMP360 IN PARTICIPANTS WITH TREATMENT-RESISTANT DEPRESSION (P-TRD)

DRUG:	COMP360
TITLE:	The Safety and Efficacy of COMP360 in Participants with Treatment-Resistant Depression (P-TRD)
CLINICAL PHASE	IIb
SPONSOR:	COMPASS Pathfinder Limited 3rd Floor, 1 Ashley Road Altrincham, Cheshire, WA14 2DT, UK
VERSION NUMBER:	Final Version 2.0
VERSION DATE:	26 October 2021

CONFIDENTIALITY STATEMENT

SAP APPROVAL FORM

Protocol Number: COMP 001

Title: The Safety and Efficacy of COMP360 in Participants with Treatment-Resistant Depression (P-TRD)

SAP Version and Date: Version 2.0 26 October 2021

This SAP was reviewed and approved by the sponsor. The information contained in this SAP is consistent with:

- The current risk benefit evaluation of the investigational product.
- The moral, ethical and scientific principles governing clinical research as set out in the Declaration of Helsinki, and principles of Good Clinical Practices (GCP) as described in the Code of Federal Regulations (CFR) 21 CFR parts 50, 54, 56 and 312 and according to applicable local requirements.

Author Details and Signature:

Signature: _____ Date: _____

Name (print) Niccolò Bassani

Title: Principal Statistician, Worldwide Clinical Trials, Inc.

Approver Details and Signature:

Signature: _____ Date: _____

Name (print) Sunil Mistry

Title: Director of Statistics, COMPASS Pathfinder Limited

SUMMARY OF CHANGES

Updates made from the protocol (Version 4.0, dated 19 July 2019) and changes made to the prior version of the SAP (Version 1.0, dated 06 December 2019) are summarised below.

Section	Revision / Addition	Rationale
Protocol Specific Changes (Version 4.0, dated 19 July 2019)		
All sections	Minor changes were made to the objectives and endpoints wording from the protocol. These are all listed in Appendix I 'COMPASS Changes to COMP 001 Protocol'.	The rationale for each change is detailed in Appendix II (Section 19.1).
7	Analysis Sets were modified as follows: • modified Intent-to-Treat was removed • Full Analysis Set definition updated to match the modified Intent-to-Treat definition	The definition of the Safety Analysis Set and Full Analysis Set in version 1 of the SAP and in the Protocol were the same, so this change eliminates any redundancy.
10.3	Changed 'prior psychedelic experience' wording to 'prior psilocybin experience'.	To correct for the discrepancy in the Statistics section of the protocol with the rest of the protocol. The correct term to use is 'prior psilocybin experience' not 'prior psychedelic experience', as described in the protocol Section 4.1 (Study Design).
11	Removed 'prior psychedelic experience' from all analysis models.	Since only 10% of participants will have had prior psilocybin experience, adding this covariate to the analysis was anticipated not to have a large impact and thus removed from the analysis model.
11.2	The analysis of the key secondary endpoints has been changed from the originally planned Cochran-Mantel-Haenszel (CMH) procedure to either a GLMM (response and remission analysis at week 3) or a logistic regression (sustained response at week 12).	To increase statistical power by adjusting for more covariates than a stratified CMH would allow and to account for the assumed missing data mechanism.

Section	Revision / Addition	Rationale
11.4.3	Added EQ-VAS and PANAS as endpoints to be statistically analysed as well as providing summaries	This ensures a better assessment of these two endpoints.
11.4.4	The caregiver EQ-5D-3L scores will only be listed and not analysed (this was previously in Section 6.12.7)	As the caregiver EQ-5D-3L was not a mandatory assessment and is provisional of optional caregiver consent.
11.5	Removed the analysis to investigate the site/country effect	Due to the sparseness of data, it has been agreed to create a simpler by-country descriptive summary of MADRS total scores.
SAP Specific Changes (Version 1.0, dated 06 December 2019)		
	SAP is documented in a new format.	The new format follows COMPASS Pathfinder Limited's SAP template, with more detail being provided in certain sections for added clarity. See Appendix II (Section 19.2) for details of where to find sections from Version 1.0 in the new format). Also includes a list of abbreviations and parameters.
4, 11	The efficacy analysis methodology text has been restructured to align with the estimands framework and all analyses have been classified using the following hierarchy: Primary or Secondary Endpoint → Primary or Secondary Estimand → Main, Sensitivity or Supplementary Analysis (this was previously in Section 6.12)	Version 1.0 had an old structure that did not account for estimands. This has been rectified and detailed explanations of estimands, methods and rationale for assumptions are provided.
5	Added details regarding the study design, study schematic and schedule of assessments.	For completeness and to align with the protocol.
6.2	Added details confirming that the interim analysis was not to be performed (this was Section 7).	To provide notification of a decision that was made.

Section	Revision / Addition	Rationale
6.3	Added details regarding the Data Safety Monitoring Board (DSMB).	For completeness and to provide information.
8	Added general considerations including the versions of SAS, MedDRA and WHO-DDE to be used. Reporting of unscheduled visits has been generalised to both pre- and post-baseline occurrences across domains (this was previously in Section 6.2.10).	For completeness and to provide information. The original description only included post-baseline assessments and for a limited set of variables, the definition is now broader.
9.1.1	Clearer definition for change from baseline to include both 'absolute' and 'percentage' changes (this was previously in Section 6.2.9).	The original section only defined an absolute change; however, percentage change is part of the derivation of one key secondary endpoint, hence specifying it here ensures alignment of derivations.
9.2.1	Added details of the categorisation of protocol deviations. Also reworded to 'important' rather than 'major'.	To reflect the protocol deviation handling plan.
9.2.2, 9.2.3, 9.2.4, 9.2.5, 9.2.6	Added details concerning the derivation of baseline variables	For completeness and to provide information.
9.2.7	The definition of prior medications has been changed to reflect the study protocol (this was previously in Section 6.9).	The prior medications definition from the protocol includes a cut-off of 30 days prior to screening to classify a drug as prior.
9.3.1	Added sections to explicitly define how key secondary endpoints are to be derived.	Previously definitions were scattered in different sections where the analysis was detailed, this change ensures they are all centralised in one location.
9.3.1.4	Added a definition for MADRS relaxed sustained response.	To better understand the impact on MADRS total scores with a more flexible definition of sustained response.
9.3.1.5	Added a definition for MADRS partial responder status.	To better understand the impact on MADRS total scores with a less strict definition of response.
9.3.15, 9.3.16	Added sections to include all study instruments collected and to be reported.	EBI and PANAS were not included in Version 1.0.

Section	Revision / Addition	Rationale
10.1, 10.2, 11.4.1, 12.1	Added outputs related to the COVID-19 pandemic: missed visits or visits that were conducted remotely, PDs, primary endpoint, and adverse events.	To better evaluate the impact of COVID-19 on study conduct, safety and efficacy,
10.3	Added 'Prior psilocybin experience', height and BMI to baseline summaries (this was previously in Section 6.7).	This is to ensure all relevant variables are summarised.
11.3	Cox regression has been added in the analysis of time to event endpoints to provide a direct comparison between treatment groups.	The hazard ratio obtained from Cox regression model can provide a useful metric to better evaluate the impact of treatment on participants' depression.
11.4.2	Additional responders analysis added.	To better assess the response to treatment.
11.5	Added to the list of subgroup variables: country, number of failed pharmacological treatments for the current episode of depression, and depression severity at baseline (this was previously in Section 6.12.6).	To better understand the impact of baseline variables on the primary and key secondary endpoints.
12.1	Added addition adverse event summary tables (this was previously in Section 6.13.1).	For a more comprehensive analysis of safety.
13.1	Added section regarding the summary and analysis of the biomarker C-reactive protein (CRP)	This was omitted from Version 1.0.
15.1	Added section to clarify that Mindstrong app data is not in the scope of the current SAP/CSR.	Since this data will be analysed separately, this has been explicitly mentioned in this SAP.
18	Updated the list of tables, figures and listings. Added column headings indicating if they are required for topline and the source shell number (if repeat) (this was previously in Section 11).	To reflect the latest shells and topline requirements.

SAP SYNOPSIS

Protocol Number	COMP 001
Title:	The Safety and Efficacy of COMP360 in Participants with Treatment-Resistant Depression (P-TRD)
EudraCT Number:	2017-003288-36
Investigational Medicinal Product:	COMP360
ClinicalTrials.gov Identifier:	NCT03775200
Clinical Phase:	IIb
Rationale:	A recent open-label study of the effects of psilocybin in participants with treatment-resistant depression (TRD) showed rapid significant decrease of depressive symptoms after treatment with psilocybin coupled with psychological support. Over 40% of participants sustained response at three months. In this study, the aim is to assess effectiveness of 3 different doses of COMP360 (COMPASS' proprietary pharmaceutical-grade synthetic psilocybin formulation that has been optimised for stability and purity) - 1 mg, 10 mg, and 25 mg - in TRD.
Number of Participants:	216

Objectives:	**Primary Objective**
	
	The primary objective of this study is to evaluate the efficacy of COMP360 (25 mg or 10 mg) compared to 1 mg, administered under supportive conditions to adult participants with TRD, in improving depressive symptoms, as assessed by the change in the Montgomery-Asberg Depression Rating Scale (MADRS) total score from baseline. Baseline is defined as the assessment score obtained on day -1. The primary timepoint is week 3; this variable will be analysed for the change from baseline to day 2, and weeks 1, 3, 6, 9, and 12.
	
	Secondary Objectives
	
	The secondary objectives are:
	
	• To assess the efficacy of COMP360 compared to 1 mg COMP360 on:
	  • Proportion of participants with response defined as a ≥ 50% decrease in MADRS total score from baseline to week 3. This will also be assessed at day 2 and at weeks 1, 6, 9, and 12.
	  • The proportion of participants who have a sustained response at week 12. Sustained response is defined as the proportion of participants fulfilling response criteria at any visit up to and including week 3, that also fulfils response criteria at all subsequent visits up to and including week 12. Response is defined as ≥ 50% decrease in MADRS total score from baseline.
	• To evaluate the safety and tolerability of COMP360 in participants with TRD based on adverse events (AEs), changes in vital signs, and suicidal ideation/behaviour (measured using the Columbia-Suicide Severity Rating Scale [C-SSRS]) score at all visits.

Exploratory Objectives

The exploratory objectives are:

- To evaluate the effects of COMP360 on quality of life and wellbeing, functioning and associated disability, cognitive function, and anxiety compared to 1 mg COMP360 on:
  - Quality of life in participant EuroQoL (EQ) 5-dimension 3 level scale (EQ-5D-3L) score change from baseline to week 3. This will also be assessed at week 12.
  - Quality of life in caregiver EQ-5D-3L score change from baseline to week 3. This will also be assessed at week 12. This assessment is not mandatory.
  - Functioning and associated disability in the Sheehan Disability Scale (SDS) score change from baseline to week 3. This will be also assessed at week 12.
  - Cognitive function as measured by the Digit Symbol Substitution Test (DSST) score change from baseline to week 3. This will also be assessed at day 2 and week 12.
  - Level of anxiety as measured using the change in Generalised Anxiety Disorder 7 item scale (GAD-7) total score change from baseline to week 3. This will also be assessed at week 12.
  - Participant determined level of depression as measured using the change in Quick Inventory of Depressive Symptomatology Self-Rated (QIDS-SR-16) total score from baseline to week 3. This will also be assessed at screening, day 2, and weeks 1, 2, 6, 9, and 12.
  - Psychosocial functioning as measured using the change in Work and Social Adjustment Scale (WSAS) from baseline to week 3. This will also be assessed at week 12.
- To evaluate the impact of different COMP360 doses on real life functional activity estimated from passive data streams collected on a mobile app on participants' mobile phones. The data collected from the participant's phone will include:
  - Number of and time of phone calls/e-mails/texts (content will not be collected)
  - Gestures used (taps, swipes, other)
  - Gyroscope (orientation) of the phone (the way the phone is pointing)
  - Acceleration of the phone (sudden movements of the phone)
  - Keystroke patterns with characters redacted
  - Location information from the GPS
  - The app also maintains a histogram of daily words that the participant types on their phone. These words will be stripped from their context and syntax, thus preventing the content of any particular message from being deciphered.

- Positive and Negative Affect Schedule (PANAS), Five Dimension Altered States of Consciousness Questionnaire (5D-ASC) and Scale to Assess Therapeutic Relationship (Clinician and Patient version, STAR-C and STAR-P, respectively) will be assessed for correlation with the primary outcome as possible predictors of response.

Study Design and Procedures:

This is a phase II, international multicentre, randomised, fixed-dose, double-blind trial. The study population will include adult men and women, 18 years of age and older, with TRD. Participants with TRD are defined as those who meet Diagnostic and Statistical Manual of Mental Disorders (5th Edition; DSM-5) diagnostic criteria for single or recurrent episode of major depressive disorder (MDD) without psychotic features which have failed to respond to an adequate dose and duration of 2, 3, or 4 pharmacological treatments for the current episode; if single-episode MDD, the duration of the current episode must be at least 3 months but not more than 2 years.

The majority of participants will have no prior exposure to psilocybin or so-called magic mushrooms; however, to reflect the prevalence of experience in general population, we will allow up to 10% of participants with prior recreational experience with psilocybin or magic mushrooms. Past exposure to psilocybin has to be more than 12 months prior to screening and not during the current depressive episode. This will be constrained by the centralised randomisation process, Interactive Web-based Response System (IWRS).

After signing the informed consent form (ICF), participants will be assessed for their eligibility with the Mini International Neuropsychiatric Interview, version 7.0.2 (MINI 7.0.2), the 17 item Hamilton Depression Rating Scale (HAM-D-17), the Massachusetts General Hospital Antidepressant Treatment History Response Questionnaire (MGH-ATRQ), the C-SSRS, and McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). Those who meet the eligibility criteria will enter the screening period, which will last between 3 and 6 weeks. At the initial screening visit (V1), the participant will also be evaluated with the QIDS-SR-16, and the Adult Self-Report Scale (ASRS). Additionally, a medical history, an electrocardiogram (ECG), blood tests, and vital signs will be obtained.

During the screening period, participants who are on antidepressant medications will be expected to complete the taper at least 2 weeks prior to baseline (V2). Participants will be given a choice of the tapering rate. During the tapering period all participants will be supported by the study clinician.

Once a participant completes all V1 assessments and all screening data is entered into the Electronic Data Capture (EDC), the Medical Monitor (MM) and Clinical Assessment Technologies Team (CAT) will review data entered and issue approval, if the participant is eligible. Once approval is issued, the participant should then be invited for a screening V1a visit. The V1a visit is the point at which the participant begins tapering off their antidepressant and/or antipsychotic medications, if appropriate. The participant must complete the taper within the first 4 weeks of this period, prior to 2 weeks completely off antidepressant and/or antipsychotic medications, before baseline. The tapering period used in the study is set at the industry standard for depression trials.

The participant will meet with a therapist for a minimum of 3 visits during screening. These are safety sessions and will cover what to expect during the

COMP360 administration session. The therapist and the participant will review psychoeducational materials provided at the time of enrolment.

All participants will be evaluated for safety at the clinic weekly for a minimum of 3 weeks prior to COMP360 administration to ensure safe discontinuation of current antidepressant therapy required by the protocol. Participants' companions (friends or family members) will be educated about the signs of worsening of depression and suicidality and instructed on ways to contact the study team in case of significant worsening of depression. Any safety assessment visits during the screening period will be called V1a, V1b, etc. During these visits, the C-SSRS and any changes in medications since the previous visit will be obtained in addition to other assessments at the study clinician's discretion.

The day before COMP360 administration session, the participants will undergo a baseline assessment (3 to 6 weeks after initial screening [V1]) that will consist of the HAM-D-17, MADRS, QIDS-SR-16, C-SSRS, SDS, GAD-7, DSST, EQ-5D-3L (administered to both participant and caregiver [the latter is not mandatory]), WSAS, vital signs, urinalysis, urine drug screen, and urine pregnancy test (only for women of childbearing potential). Both the therapist and the participant will be asked to fill out a therapeutic alliance evaluation questionnaire, STAR-C and STAR-P, respectively. After baseline data is entered into EDC, the CAT team will complete a final review to ensure the participant's continued eligibility. Participants cannot be progressed to V3 until this approval is received.

The COMP360 administration session (V3, day 1) will last approximately 6 hours (h) and will be supported by a trained therapist. After the acute effects of COMP360 pass, participants will be evaluated for safety and accompanied home. On day 2 (V4), the day following COMP360 administration, participants will be seen in person for a safety check, assessment of suicidality, and to discuss their experience during the COMP360 administration session.

All participants will be asked to remain off their antidepressant medications for at least 3 weeks following the psilocybin session until the primary endpoint assessment, or longer. Rescue medications are allowed as noted in the protocol. Participants who restart their antidepressant medications during the first 3 weeks after the psilocybin treatment administration will be assessed for reasons of resuming their medications and followed until 12 weeks post psilocybin administration.

The treatment period will determine the optimal therapeutic dose; 216 participants will be randomised in a 1:1:1 ratio to receive 1 mg psilocybin, 10 mg psilocybin, or 25 mg psilocybin.

Participants will be seen at the clinic for screening (plus a minimum of three safety visits), baseline (day -1), study drug administration day (day 1), day 2, week 1, week 2, week 3, and week 12. Participants will also be contacted for follow-up, which can be conducted remotely, at week 6 and week 9. The MADRS will be done by telephone and the other assessments will be done electronically. Participants are seen at the clinic for safety visits between the initial screening and the baseline visit.

Primary Endpoint: The primary endpoint is the change in MADRS total score from baseline (day -1) to 3 weeks post COMP360.

Secondary Endpoints:

The secondary endpoints are:

- The proportion of participants with a response (defined as a ≥ 50% improvement in MADRS total score from baseline) at week 3 after the COMP360 session.
- The proportion of participants with remission (defined as a MADRS total score ≤ 10) at week 3 post COMP360.
- The proportion of participants who have a sustained response at week 12. Sustained response is defined as the proportion of participants fulfilling response criteria at any visit up to and including week 3, that also fulfils response criteria at all subsequent visits up to and including week 12. Response is defined as ≥ 50% decrease in MADRS total score from baseline.
- Time to event measures: restart antidepressant medication for any reason, restart medication for continuing depressive symptoms, and relapse from a previously recovered state. Participants who withdraw from the study will be censored from the time to event analysis.

Exploratory Endpoints:

The exploratory endpoints are:

- Change from baseline in the following:
  - Participant EQ-5D-3L at week 3
  - Caregiver EQ-5D-3L at week 3 (this assessment is not mandatory)
  - SDS at week 3
  - DSST at week 3
  - GAD-7 at week 3
  - QIDS-SR-16 at week 3
  - WSAS at week 3
- Measures derived from smart phone usage via the Mindstrong app

Safety Assessments

- ECG
- Vital signs
- Blood test including liver function tests
- Suicide risk as assessed by the C-SSRS
- AEs and Serious AEs

Estimand Framework

The Hypothetical Strategy will be the primary estimand to best understand the clinical effects of COMP360 without the effects of participants taking new treatments for depression. Specifically, the primary estimand will be as follows:

The mean difference (population-level summary) between COMP360 (25 mg or 10 mg) and COMP360 1 mg (treatment) in participants with TRD, as defined by the inclusion/exclusion criteria (population) assessing the change from baseline in MADRS total score at week 3 (variable) if new treatment for depression is not available post-dose (intercurrent event).

The Treatment Policy will be the secondary estimand to understand the clinical effects of COMP360 versus inactive control including participants taking new treatment for depression if needed in both arms. Specifically, the secondary estimand will be as follows:

The mean difference (population-level summary) between COMP360 (25 mg or 10 mg) and COMP360 1 mg (treatment) in participants with TRD, as defined by the inclusion/exclusion criteria (population) assessing the change from baseline in MADRS total score at week 3 (variable) regardless of whether participants take new treatment for depression (intercurrent event).

Sample Size Determination

The intent of the primary efficacy analysis is to demonstrate superiority of at least one therapeutic dose of COMP360 (10 mg or 25 mg) versus the 1 mg COMP360 dose based on the change from baseline in MADRS total score at week 3. The three treatment groups will be randomised in a 1:1:1 ratio.

For this primary analysis, a sample size of 216 randomised participants (72:72:72) will provide 90% power at the alpha = 0.05 level to detect a 6-point difference in average MADRS total score between the optimal therapeutic dose of COMP360 and COMP360 1 mg, assuming the common standard deviation (SD) is 11.0.

It is assumed that up to 90% of randomised participants may not have prior psilocybin experience. The power for this post-hoc subgroup is approximately 86%, if the maximum number of participants to have prior psilocybin experience was 10% of randomised participants.

Analysis Sets

The Randomised Analysis Set: All randomised participants, whether or not they receive study drug.

The Safety Analysis Set: All randomised participants who receive study drug.

The Full Analysis Set (FAS): All participants randomised who receive study drug and have at least 1 post-dose efficacy assessment.

The Per-Protocol (PP) Analysis Set: All participants in the FAS who do not have a protocol deviation that is thought to significantly affect the integrity of the participant's efficacy data.

Hypothesis

The primary efficacy endpoint is the change from baseline in MADRS total score at week 3. The primary analysis will be the comparison between COMP360 (25 mg or 10 mg) versus COMP360 1 mg and will be based on a two-sided hypothesis testing approach. The null and alternative hypotheses associated for this comparison are:

H_0: $\mu_{\text{COMP360 (25 mg or 10 mg)}} - \mu_{\text{COMP360 1 mg}} = 0$

H_1: $\mu_{\text{COMP360 (25 mg or 10 mg)}} - \mu_{\text{COMP360 1 mg}} \neq 0$

Where $\mu_{\text{COMP360 (25 mg or 10 mg)}}$ and $\mu_{\text{COMP360 1 mg}}$ represent the mean in change from baseline in MADRS total score at week 3 for COMP360 (25 mg or 10 mg) and COMP360 1 mg, respectively.

The null hypothesis is that there is no difference between treatment groups for the primary efficacy endpoint and the alternative hypothesis is that there is a difference between treatment groups for the primary efficacy endpoint.

Primary Endpoint Analyses	The primary efficacy endpoint (change from Baseline in MADRS total score at Week 3) will be evaluated with a mixed-effects model for repeated measures (MMRM) analysis following multiple imputation of MADRS data. The MMRM model will include treatment, visit, study site and treatment-by-visit interaction as fixed effects and baseline MADRS total score as a continuous covariate. Correlation between repeated observations within a participant will be accounted for by specifying an unstructured correlation. Comparison of the COMP360 optimal dose versus 1 mg COMP360 will be performed at the 0.05 testing level and results for each imputation will be pooled to obtain a single treatment effect estimate. A sensitivity analysis will be performed on the primary MMRM model using a tipping-point approach. A supplementary analysis using Analysis of Covariance (ANCOVA) will also be performed following the same multiple imputation as for the primary MMRM analysis.
Key Secondary Endpoint Analyses	The three key secondary efficacy endpoints (proportion of participants who are responders, remitters, and sustained responders) will be analysed using either a Generalised Linear Mixed Model (responders and remitters) with the same covariates and within-participant correlation structure as the MMRM model for the primary endpoint or a logistic regression model (sustained responders) including treatment and baseline MADRS total score as covariates, to compare the COMP360 optimal therapeutic dose versus 1 mg COMP360. A hierarchical procedure to maintain control over type I error will be employed.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Definition
ANCOVA	Analysis of Covariance
bpm	beats per minute
CAT	Clinical Assessment Technologies
CDISC	Clinical Data Interchange Standards Consortium
CFR	Code of Federal Regulations
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
COMPASS	COMPASS Pathfinder Limited
CRP	C-Reactive Protein
CSR	Clinical Study Report
dp	decimal place
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th edition
DSMB	Data and Safety Monitoring Board
eCRF	electronic Case Report Form
eCTD	electronic Common Technical Document
EDC	Electronic Data Capture
EOS	End of Study
ET	Early Termination
FAS	Full Analysis Set
GCP	Good Clinical Practice
GLMM	Generalised Linear Mixed Model
HR	Hazard Ratio
HRQoL	Health-Related Quality of Life
ICF	Informed Consent Form
ICH	The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IP	Independent programming
IWRS	Interactive Web-based Response System
kg	Kilogram
l	Litres

Abbreviation	Definition
LS	Least Squares
MAR	Missing At Random
MCMC	Markov Chain MonteCarlo
MDD	Major Depressive Disorder
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
min	minute(s)
mg	Milligram
mITT	modified Intent-to-Treat
MM	Medical Monitor
mmHg	millimetres of mercury
MMRM	Mixed Model for Repeated Measures
MNAR	Missing Not At Random
msec	millisecond(s)
OR	Odds Ratio
PD	Protocol Deviation
PP	Per Protocol
PT	Preferred Term
P-TRD	Psilocybin for Treatment-Resistant Depression
QC	Quality Control
SAP	Statistical Analysis Plan
SD	Standard Deviation
SE	Standard Error
SOC	System Organ Class
TFLs	Tables, Figures and Listings
TRD	Treatment-resistant Depression
UK	United Kingdom
WHO	World Health Organisation
Worldwide	Worldwide Clinical Trials, Inc.

LIST OF PARAMETERS

Parameter	Definition
Efficacy	
5D-ASC	Five Dimensional - Altered States of Consciousness questionnaire
ASRS	Adult Self-Report Scale
DSST	Digits Symbol Substitution Test
EBI	Emotional Breakthrough Inventory
EQ-5D-3L	EuroQoL-5 Dimensions-3 Levels
EQ VAS	EuroQoL Visual Analogue Scale
GAD-7	Generalised Anxiety Disorder Scale – 7 item
HAM-D-17	Hamilton Depression Rating Scale – 17 item
MADRS	Montgomery-Asberg Depression Rating Scale
MGH-ATRQ	Massachusetts General Hospital-Antidepressant Treatment History Questionnaire
MINI 7.0.2	Mini International Neuropsychiatric Interview, Version 7.0.2
MSI-BPD	McLean Screening Instrument for Borderline Personality Disorder
PANAS	Positive and Negative Affect Schedule
QIDS-SR-16	Quick Inventory of Depressive Symptomatology – Self-Rated 16 item
SDS	Sheehan Disability Scale
STAR-C	Scale to Assess Therapeutic Relationship – Clinician version
STAR-P	Scale to Assess Therapeutic Relationship – Patient version
VAS	Visual Analogue Scale
WSAS	Work and Social Adjustment Scale
Study Disposition	
BMI	Body Mass Index
Safety	
ADR	Adverse Drug Reaction
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALP	alkaline phosphatase

Parameter	Definition
ALT	alanine aminotransferase
AST	aspartate aminotransferase
C-SSRS	Columbia-Suicide Severity Rating Scale
ECG	electrocardiogram
SAE	Serious Adverse Event
SAR	Serious Adverse Reaction
S-STS	Sheehan Suicidality Tracking Scale
TEAE	Treatment-Emergent Adverse Event
TESAE	Treatment-Emergent Serious Adverse Event

1 INTRODUCTION

This statistical analysis plan (SAP) is based on the final protocol (V4.0) dated 19 July 2019. The plan covers statistical analysis, tables, figures and listings (TFLs) of the study data to investigate the safety and efficacy of COMP360 in participants with treatment-resistant depression (TRD). COMP360 is COMPASS Pathfinder Limited's (COMPASS) proprietary pharmaceutical-grade synthetic psilocybin formulation that has been optimised for stability and purity.

The SAP is prepared by Worldwide Clinical Trials, Inc. (Worldwide). The statistical analyses and production of the outputs described in the SAP, as well as the quality control (QC), will be conducted by Worldwide using SAS® Version 9.4 or later.[1] The final analyses and outputs will be approved by COMPASS.

Any statistical analysis details described in this document supersede any description of statistical analysis in the protocol.

2 STUDY OBJECTIVES

The main purpose of this study is to allow COMPASS to determine the optimal dose of COMP360, either 10 mg or 25 mg. The intent of the primary efficacy analysis is to demonstrate superiority of at least one optimal therapeutic dose of COMP360 (10 mg or 25 mg) versus the 1 mg COMP360 via the following objectives.

2.1 Primary

The primary objective of this study is to evaluate the efficacy of COMP360 (25 mg or 10 mg) compared to 1 mg, administered under supportive conditions to adult participants with TRD, in improving depressive symptoms, as assessed by the change in the Montgomery-Asberg Depression Rating Scale (MADRS) total score from baseline. Baseline is defined as the assessment score obtained on day -1. The primary timepoint is week 3; this variable will be analysed for the change from baseline to day 2, and weeks 1, 3, 6, 9, and 12.

2.2 Secondary

The secondary objectives are:

- To assess the efficacy of COMP360 compared to 1 mg COMP360 on:
    - Proportion of participants with response defined as a ≥ 50% decrease in MADRS total score from baseline to week 3. This will also be assessed at day 2 and at weeks 1, 6, 9, and 12.
    - The proportion of participants who have a sustained response at week 12. Sustained response is defined as the proportion of participants fulfilling response criteria at any visit up to and including week 3, that also fulfils response criteria at all subsequent visits up to and including week 12. Response is defined as ≥ 50% decrease in MADRS total score from baseline.
- To evaluate the safety and tolerability of COMP360 in participants with TRD based on adverse events (AEs), changes in vital signs, and suicidal ideation/behaviour (measured using the Columbia-Suicide Severity Rating Scale [C-SSRS]) score at all visits.

2.3 Exploratory

The exploratory objectives are:

- To evaluate the effects of COMP360 on quality of life and wellbeing, functioning and associated disability, cognitive function, and anxiety compared to 1 mg COMP360 on:

- Quality of life in participant EuroQoL (EQ) 5 dimensions 3 levels scale (EQ-5D-3L) score change from baseline to week 3. This will also be assessed at week 12.

- Quality of life in caregiver EQ-5D-3L score change from baseline to week 3. This will also be assessed at week 12. This assessment is not mandatory.

- Functioning and associated disability in the Sheehan Disability Scale (SDS) score change from baseline to week 3. This will be also assessed at week 12.

- Cognitive function as measured by the Digit Symbol Substitution Test (DSST) score change from baseline to week 3. This will also be assessed at day 2 and week 12.

- Level of anxiety as measured using the change in Generalised Anxiety Disorder 7 item scale (GAD-7) total score change from baseline to week 3. This will also be assessed at week 12.

- Participant determined level of depression as measured using the change in Quick Inventory of Depressive Symptomatology Self-Rated (QIDS-SR-16) total score from baseline to week 3. This will also be assessed at screening, day 2, and weeks 1, 2, 6, 9, and 12.

- Psychosocial functioning as measured using the change in Work and Social Adjustment Scale (WSAS) from baseline to week 3. This will also be assessed at week 12.

- To evaluate the impact of different COMP360 doses on real life functional activity estimated from passive data streams collected on a mobile app on participants' mobile phones. The data collected from the participant's phone will include:

  - Number of and time of phone calls/e-mails/texts (content will not be collected)

  - Gestures used (taps, swipes, other)

  - Gyroscope (orientation) of the phone (the way the phone is pointing)

  - Acceleration of the phone (sudden movements of the phone)

  - Keystroke patterns with characters redacted

  - Location information from the GPS

  - The app also maintains a histogram of daily words that the participant types on their phone. These words will be stripped from their context and syntax, thus preventing the content of any particular message from being deciphered.

- Positive and Negative Affect Schedule (PANAS), Five Dimension Altered States of Consciousness Questionnaire (5D-ASC) and Scale to Assess Therapeutic Relationship (Clinician and Patient version, STAR-C and STAR-P, respectively) will be assessed for correlation with the primary outcome as possible predictors of response.

3 STUDY ENDPOINTS

3.1 Primary

The primary endpoint is the change in MADRS total score from baseline (day -1) to 3 weeks post COMP360.

3.2 Secondary

The secondary endpoints are:

- The proportion of participants with a response (defined as a ≥ 50% improvement in MADRS total score from baseline) at week 3 after the COMP360 session.

- The proportion of participants with remission (defined as a MADRS total score ≤ 10) at week 3 post COMP360.

- The proportion of participants who have a sustained response at week 12. Sustained response is defined as the proportion of participants fulfilling response criteria at any visit up to and including week 3, that also fulfils response criteria at all subsequent visits up to and including week 12. Response is defined as ≥ 50% decrease in MADRS total score from baseline.

- Time to event measures: restart antidepressant medication for any reason, restart medication for continuing depressive symptoms, and relapse from a previously recovered state. Participants who withdraw from the study will be censored from the time to event analysis.

3.3 Exploratory

The exploratory endpoints are:

- Change from baseline in the following:

    - Participant EQ-5D-3L at week 3

    - Caregiver EQ-5D-3L at week 3 (this assessment is not mandatory)

    - SDS at week 3

    - DSST at week 3

    - GAD-7 at week 3

- QIDS-SR-16 at week 3
- WSAS at week 3

- Measures derived from smart phone usage via the Mindstrong app

3.4 Safety Assessments

- Electrocardiogram (ECG)
- Vital signs
- Blood test including liver function tests
- Suicide risk as assessed by the C-SSRS
- AEs and serious adverse events (SAEs)

4 ESTIMAND FRAMEWORK

4.1 Primary Endpoint

The primary objective is to understand the clinical effects of COMP360 without the effects of participants taking new treatment for depression. This is reflected in the trial design, where participants are encouraged not to take new treatments for depression for at least 3 weeks post-dose. However, it would be unethical to preclude the use of any rescue medication at any stage during the study. Participants who start antidepressant treatment during the first three weeks of the study will be followed up in the study until 12 weeks post study drug administration regardless, as will those who initiate new treatments after three weeks. To answer this objective, the estimand framework and statistical analyses will need to account for whether participants begin new treatments for depression. The Hypothetical Strategy best answers this.

The estimand attributes for the primary endpoint are:

- Treatment: COMP360 (25 mg or 10 mg) versus COMP360 1 mg
- Population: Participants with TRD, as defined by the inclusion/exclusion criteria
- Variable: Change from baseline in MADRS total score at week 3
- Intercurrent event: Initiation of new treatment for depression post-dose
- Population-level summary: Mean difference

The Hypothetical Strategy will be the primary estimand to best understand the clinical effects of COMP360 without the effects of participants taking new treatments for depression, as outlined above. Specifically, the primary estimand will be as follows:

The mean difference (population-level summary) between COMP360 (25 mg or 10 mg) and COMP360 1 mg (treatment) in participants with TRD, as defined by the inclusion/exclusion criteria (population) assessing the change from baseline in MADRS total score at week 3 (variable) if new treatment for depression is not available post-dose (intercurrent event).

The Treatment Policy will be the secondary estimand to understand the clinical effects of COMP360 including participants taking new treatment for depression if needed. Specifically, the secondary estimand will be as follows:

The mean difference (population-level summary) between COMP360 (25 mg or 10 mg) and COMP360 1 mg (treatment) in participants with TRD, as defined by the inclusion/exclusion criteria (population) assessing the change from baseline in MADRS total score at week 3 (variable) regardless of whether participants take new treatment for depression post-dose (intercurrent event).

Table 1 below provides further details for each of the above estimands including analysis methods, imputation approaches and any planned supplementary/sensitivity analyses.

Table 1 Summary of Estimand Framework – Primary Endpoint

Study Objective	Study Endpoint	Estimand	Analysis Type	Analysis Details
To evaluate the efficacy of COMP360 (25 mg or 10 mg) compared to 1 mg, administered under supportive conditions to adult participants with TRD, in improving depressive symptoms.	Change in MADRS total score from baseline (day -1) to 3 weeks post COMP360.	Hypothetical Strategy: the mean difference between COMP360 (25 mg or 10 mg) and COMP360 1 mg in participants with TRD, as defined by the inclusion/exclusion criteria, assessing the change from baseline in MADRS total score at week 3 if new treatment for depression is not available post-dose.	Primary	Missing not at random (MNAR) imputation for data after start of new treatment for depression and a missing at random (MAR) and MNAR imputation for data-points after study discontinuation and MAR for non-monotone missing data + Mixed Model for Repeated Measures (MMRM)
			Sensitivity	Tipping-point analysis: delta-imputation approach (MNAR) for all data imputed in the primary analysis + MMRM
			Supplementary	Primary analysis imputation methods + Analysis of Covariance (ANCOVA)
		Treatment Policy: the mean difference between COMP360 (25 mg or 10 mg) and COMP360 1 mg in participants with TRD, as defined by the inclusion/exclusion criteria, assessing the change from baseline in MADRS total score at week 3 regardless of whether participants take new treatment for depression post-dose.	Main	MAR and MNAR imputation for data-points after study discontinuation and MAR for non-monotone missing data + MMRM
			Sensitivity	1) Tipping-point analysis: delta-imputation approach (MNAR) for all data imputed in the main analysis of this estimand + MMRM 2) MMRM with no imputation for missing data
			Supplementary	Main analysis imputation methods + ANCOVA

4.2 Key Secondary Endpoint

A composite strategy will be adopted to define the primary estimand for the following key secondary endpoints:

- The proportion of participants with a response (defined as a ≥ 50% improvement in MADRS total score from baseline) at week 3 after the COMP360 session.

- The proportion of participants with remission (defined as a MADRS total score ≤ 10) at week 3 post COMP360.

- The proportion of participants who have a sustained response at week 12. Sustained response is defined as the proportion of participants fulfilling response criteria at any visit up to and including week 3, that also fulfils response criteria at all subsequent visits up to and including week 12. Response is defined as ≥ 50% decrease in MADRS total score from baseline.

The formal estimand definition for the first key secondary endpoint (MADRS response at week 3) using the Composite Strategy is as follows:

The odds ratio (population-level summary) between COMP360 (25 mg or 10 mg) and COMP360 1 mg (treatment) in participants with TRD, as defined by the inclusion/exclusion criteria (population), assessing the MADRS response at week 3 where a responder is defined as a participant satisfying the MADRS response criteria who does not take new treatment for depression (variable and intercurrent event).

A similar definition applies for the other two key secondary endpoints.

A secondary estimand following the Treatment Policy will also be considered, using a similar definition as for the primary endpoint. The formal estimand definition for the first key secondary endpoint (MADRS response at week 3) using the Treatment Policy strategy is as follows:

The odds ratio (population-level summary) between COMP360 (25 mg or 10 mg) and COMP360 1 mg (treatment) in participants with TRD, as defined by the inclusion/exclusion criteria (population) assessing the MADRS response at week 3 where a responder is defined as a participant satisfying the MADRS response criteria (variable) regardless of whether participants take new treatment for depression (intercurrent event).

All estimands and analyses for the primary and key secondary endpoints are specified in Table 2.

Table 2 Summary of Estimand Framework – Key Secondary Endpoints

Study Objective	Study Endpoint	Estimand	Analysis Type	Analysis Details
To evaluate the efficacy of COMP360 (25 mg or 10 mg) compared to 1 mg, administered under supportive conditions to adult participants with TRD, in improving depressive symptoms	The proportion of participants with a MADRS response at week 3 after the COMP360 session	Composite Strategy: the odds ratio between COMP360 (25 mg or 10 mg) and COMP360 1 mg in participants with TRD, as defined by the inclusion/exclusion criteria, assessing MADRS response at week 3 where a responder is defined as a participant satisfying the MADRS response criteria who does not take new treatment for depression	Main	Non-responder imputation for all data-points after (i) the start of new treatment for depression or (ii) study discontinuation (lack of efficacy or adverse event [AE]), and dichotomised MADRS scores from missing at random (MAR) imputation (main analysis) of Hypothetical Strategy estimand for the primary endpoint + Generalised Linear Mixed Model (GLMM)
			Sensitivity	-
			Supplementary	Non-responder/MAR imputation (see main analysis) + logistic regression by visit
		Treatment Policy: the odds ratio between COMP360 (25 mg or 10 mg) and COMP360 1 mg in participants with TRD, as defined by the inclusion/exclusion criteria assessing MADRS response at week 3 regardless of whether participants take new treatment for depression	Main	Non-responder imputation for all data-points after study discontinuation (lack of efficacy or AE), and dichotomised MADRS scores from MAR imputation (main analysis) of the Treatment Policy estimand for the primary endpoint + GLMM
			Sensitivity	GLMM with no imputation
			Supplementary	Non-responder imputation for all data-points after study discontinuation (lack of efficacy or AE), and dichotomised MADRS scores from MAR imputation (main analysis) of the Treatment Policy estimand for the primary endpoint + logistic regression by visit
	The proportion of participants with MADRS	Composite Strategy: the odds ratio between COMP360 (25 mg or 10 mg) and COMP360 1 mg in participants with TRD, as defined	Main	See first secondary endpoint
			Sensitivity	
			Supplementary	

Study Objective	Study Endpoint	Estimand	Analysis Type	Analysis Details
	remission at week 3 after the COMP360 session	by the inclusion/exclusion criteria, assessing MADRS remission at week 3 where a remitter is defined as a participant satisfying the MADRS remission criteria who does not take new treatment for depression		
		Treatment Policy: the odds ratio between COMP360 (25 mg or 10 mg) and COMP360 1 mg in participants with TRD, as defined by the inclusion/exclusion criteria assessing MADRS remission at week 3 regardless of whether participants take new treatment for depression	Main	See first secondary endpoint
			Sensitivity	
			Supplementary	
	The proportion of participants with MADRS sustained response at week 12 after the COMP360 session	Composite Strategy: the odds ratio between COMP360 (25 mg or 10 mg) and COMP360 1 mg in participants with TRD, as defined by the inclusion/exclusion criteria, assessing MADRS sustained response at week 12 where a sustained responder is defined as a participant satisfying the MADRS sustained response criteria who does not take new treatment for depression	Main	Non-responder imputation for all participants that (i) start new treatment for depression or (ii) discontinue study (lack of efficacy or AE), and dichotomised MADRS scores from MAR imputation (main analysis) of Hypothetical Strategy estimand for the primary endpoint + logistic regression
			Sensitivity	-
			Supplementary	-
		Treatment Policy: the odds ratio between COMP360 (25 mg or 10 mg) and COMP360 1 mg in participants with TRD, as defined by the inclusion/exclusion criteria assessing MADRS sustained response at week 12 regardless of whether participants take new treatment for depression	Main	Non-responder imputation for all participants that discontinue study (lack of efficacy or AE), and dichotomised MADRS scores from MAR imputation (main analysis) of the Treatment Policy estimand for the primary endpoint + logistic regression
			Sensitivity	-
			Supplementary	Observed case analysis – logistic regression

5 STUDY DESIGN

5.1 Study Design

This is a phase IIb, international multicentre, randomised, fixed-dose, double-blind trial. The study population will include adult men and women, 18 years of age and older, with TRD. Participants with TRD are defined as those who meet Diagnostic and Statistical Manual of Mental Disorders (5th Edition; DSM-5) diagnostic criteria for single or recurrent episode of major depressive disorder (MDD) without psychotic features which have failed to respond to an adequate dose and duration of 2, 3, or 4 pharmacological treatments for the current episode; if single-episode MDD, the duration of the current episode must be at least 3 months but not more than 2 years.

The majority of participants will have no prior exposure to psilocybin or so-called magic mushrooms; however, to reflect the prevalence of experience in general population, we will allow up to 10% of participants with prior recreational experience with psilocybin or magic mushrooms. Past exposure to psilocybin has to be more than 12 months prior to screening and not during the current depressive episode. This will be constrained by the centralised randomisation process, Interactive Web-based Response System (IWRS).

After signing the informed consent form (ICF), participants will be assessed for their eligibility with the Mini International Neuropsychiatric Interview, version 7.0.2 (MINI 7.0.2), the 17 item Hamilton Depression Rating Scale (HAM-D-17), the Massachusetts General Hospital Antidepressant Treatment History Response Questionnaire (MGH-ATRQ), the C-SSRS, and McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). Those who meet the eligibility criteria will enter the screening period, which will last between 3 and 6 weeks. At the initial screening visit (V1), the participant will also be evaluated with the QIDS-SR-16, and the Adult Self-Report Scale (ASRS). Additionally, a medical history, an electrocardiogram (ECG), blood tests, and vital signs will be obtained.

During the screening period, participants who are on antidepressant medications will be expected to complete the taper at least 2 weeks prior to baseline (V2). Participants will be given a choice of the tapering rate. During the tapering period all participants will be supported by the study clinician.

Once a participant completes all V1 assessments and all screening data is entered into the Electronic Data Capture (EDC), the Medical Monitor (MM) and Clinical Assessment Technologies Team (CAT) will review data entered and issue approval, if the participant is eligible. Once approval is issued, the participant should then be invited for a screening V1a visit. The V1a visit is the point at which the participant begins tapering off their antidepressant and/or antipsychotic medications, if appropriate. The participant must complete the taper within the first 4 weeks of this period, prior to 2 weeks completely off antidepressant and/or antipsychotic medications, before baseline. The tapering period used in the study is set at the industry standard for depression trials.

The participant will meet with a therapist for a minimum of 3 visits during screening. These are safety sessions and will cover what to expect during the COMP360 administration session. The therapist and the participant will review psychoeducational materials provided at the time of enrolment.

All participants will be evaluated for safety at the clinic weekly for a minimum of 3 weeks prior to COMP360 administration to ensure safe discontinuation of current antidepressant therapy required by the protocol. Participants' companions (friends or family members) will be educated about the signs of worsening of depression and suicidality and instructed on ways to contact the study team in case of significant worsening of depression. Any safety assessment visits during the screening period will be called V1a, V1b, etc. During these visits, the C-SSRS and any changes in medications since the previous visit will be obtained in addition to other assessments at the study clinician's discretion.

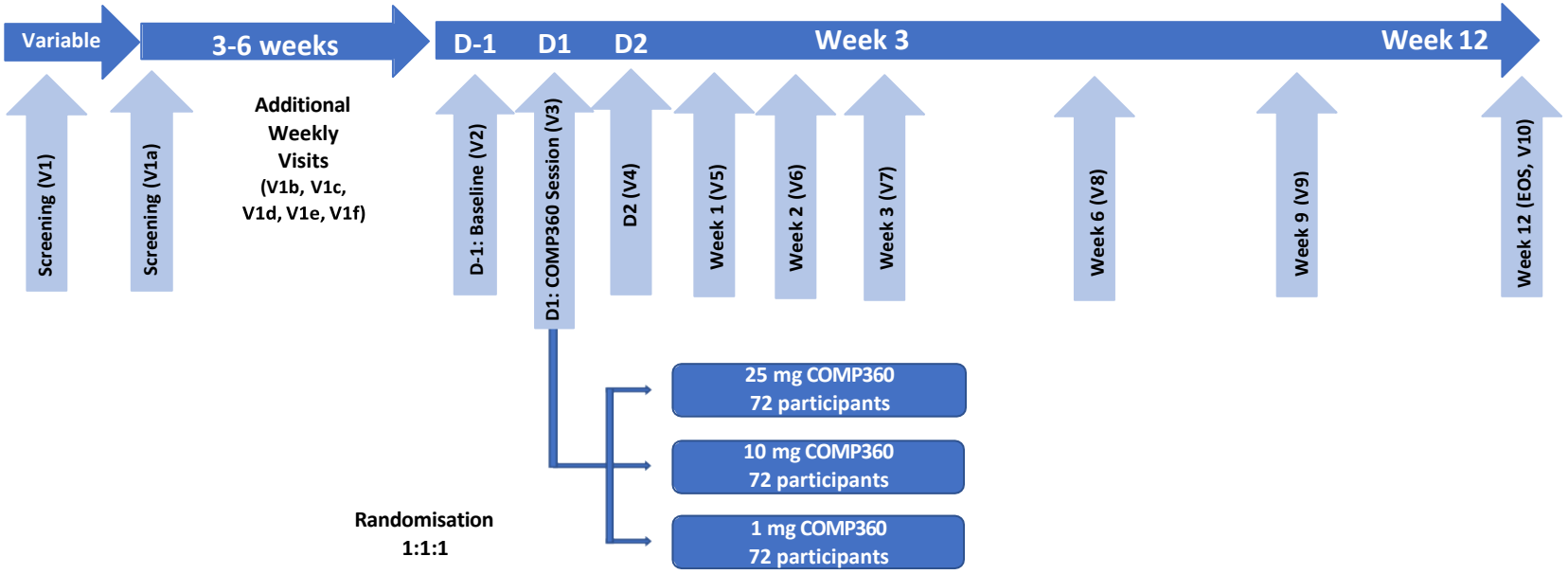
The day before COMP360 administration session, the participants will undergo a baseline assessment (3 to 6 weeks after initial screening [V1]) that will consist of the HAM-D-17, MADRS, QIDS-SR-16, C-SSRS, SDS, GAD-7, DSST, EQ-5D-3L (administered to both participant and caregiver [the latter is not mandatory]), WSAS, vital signs, urinalysis, urine drug screen, and urine pregnancy test (only for women of childbearing potential). Both the therapist and the participant will be asked to fill out a therapeutic alliance evaluation questionnaire, STAR-C and STAR-P, respectively. After baseline data is entered into EDC, the CAT team will complete a final review to ensure the participant's continued eligibility. Participants cannot be progressed to V3 until this approval is received.

The COMP360 administration session (V3, day 1) will last approximately 6 hours (h) and will be supported by a trained therapist. After the acute effects of COMP360 pass, participants will be evaluated for safety and accompanied home. On day 2 (V4), the day following COMP360 administration, participants will be seen in person for a safety check, assessment of suicidality, and to discuss their experience during the COMP360 administration session.

Participants will be seen at the clinic for screening (plus a minimum of three safety visits), baseline (day -1), study drug administration day (day 1), day 2, week 1, week 2, week 3, and week 12. Participants will also be contacted for follow-up, which can be conducted remotely, at week 6 and week 9. The MADRS will be done by telephone and the other assessments will be done electronically. Participants are seen at the clinic for safety visits between the initial screening and the baseline visit.

The study schematic is presented in Section 5.2 and the schedule of assessments is presented in Section 5.3.

5.2 Study Schematic



Abbreviations: D = day; EOS = End of Study; V = Visit

5.3 Schedule of Assessments

		3-6 weeks prior to Baseline			Time Post COMP360 Session						
	Screen Visit[2]	Screening Period	Baseline (Day -1)	COMP360 session (Day 1)	Day 2	Week 1 Day 8	Week 2 Day 15	Week 3 Day 22	Week 6 Day 43	Week 9 Day 64	Week 12 Day 85 (ET)
Visit	1	1a, 1b, etc	2	3	4	5	6	7	8	9	10
Location of Visit[1]	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Remote	Remote	Clinic
Allowable Window		Weekly		≤ 7 days	none	±1 day	±1 day	± 1 day	± 3 days	± 3 days	± 7 days
Clinic Assessments and Procedures											
Informed Consent	✓										
Medical History	✓		✓								
Inclusion/exclusion Criteria	✓		✓								
MINI	✓										
HAM-D-17	✓		✓								
MGH-ATRQ	✓										
STAR-C			✓								
C-SSRS[3]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Vital signs	✓		✓	✓	✓						
Weight	✓										
Height	✓										
ECG	✓				✓						
Clinical laboratory tests and biomarker[4]	✓				✓			✓			
Urinalysis[4]	✓				✓						

		3-6 weeks prior to Baseline			Time Post COMP360 Session						
	Screen Visit[2]	Screening Period	Baseline (Day -1)	COMP360 Session (Day 1)	Day 2	Week 1 Day 8	Week 2 Day 15	Week 3 Day 22	Week 6 Day 43	Week 9 Day 64	Week 12 Day 85 (ET)
Visit	1	1a, 1b, etc	2	3	4	5	6	7	8	9	10
Location of Visit[1]	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Remote	Remote	Clinic
Allowable Window		Weekly		≤ 7 days	none	±1 day	±1 day	± 1 day	± 3 days	± 3 days	± 7 days
Urine Drug Screen[4]	✓		✓								
Urine Pregnancy Test[5]	✓		✓								
Documentation of Birth Control to be used[6]	✓										
2a polymorphism (optional)	✓										
Activate/deactivate Mindstrong (optional)	✓										✓
Provide access to psychoeducational material (Longboat)	✓	✓[9]									
COMP360 dose				✓[10]							
Prior/Concomitant Medications	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
AE/SAE				✓	✓	✓	✓	✓	✓	✓	✓
Randomisation			✓								
Participant Completed Assessments											
5D-ASC				✓[7]							
ASRS	✓										

		3-6 weeks prior to Baseline			Time Post COMP360 Session						
	Screen Visit[2]	Screening Period	Baseline (Day -1)	COMP360 Session (Day 1)	Day 2	Week 1 Day 8	Week 2 Day 15	Week 3 Day 22	Week 6 Day 43	Week 9 Day 64	Week 12 Day 85 (ET)
Visit	1	1a, 1b, etc	2	3	4	5	6	7	8	9	10
Location of Visit[1]	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Remote	Remote	Clinic
Allowable Window		Weekly		≤ 7 days	none	±1 day	±1 day	± 1 day	± 3 days	± 3 days	± 7 days
DSST			✓		✓			✓			✓
EBI					✓						
EQ-5D-3L[8]			✓					✓			✓
GAD-7			✓					✓			✓
MSI-BPD	✓										
PANAS			✓		✓			✓			
QIDS-SR-16	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
SDS			✓					✓			✓
STAR-P			✓								
WSAS			✓					✓			✓
Remote-rater Assessment											
MADRS			✓		✓	✓		✓	✓	✓	✓

Abbreviations: 5D-ASC, Five Dimension Altered States of Consciousness Questionnaire; AE, adverse event; ASRS, Adult Self-Report Scale; C-SSRS, Columbia-Suicide Severity Rating Scale; DSM-5 Diagnostic and Statistical Manual of Mental Disorders (5th edition); DSST, Digit Symbol Substitution Test; EBI, Emotional Breakthrough Inventory;
ECG, electrocardiogram; EQ-5D-3L, EuroQoL 5-Dimension 3-Level; ET, early termination; GAD-7, Generalised Anxiety Disorder scale - 7 item; h, hour(s); HAM-D-17, Hamilton Depression Rating scale - 17 item; MADRS, Montgomery-Asberg Depression Scale; MGH-ATRQ, Massachusetts General Hospital Antidepressant Treatment History Questionnaire; MINI, Mini International NeuropsychiatricInterview; MSI-BPD, McLean Screening Instrument for Borderline Personality Disorder; PANAS, Positive and Negative Affect Schedule; QIDS-SR-16, Quick Inventory of Depressive

Symptomatology-Self-rated; SAE, serious adverse event; SDS, Sheehan Disability Scale; STAR-C, Scale to Assess the Therapeutic Relationship-Clinician; STAR-P, Scale to Assess the Therapeutic Relationship-Patient; WSAS, Work and Social Adjustment Scale

[1] On site clinic visits; visits allowed remotely will have the MADRS performed by telephone and other assessments will be done electronically.
[2] If additional visits are needed to ensure adequate time for discontinuation of prior antidepressant therapy, visits should occur weekly prior to the COMP360 administration session (V3). At subsequent screening period visits (V1a, V1b, etc), medications taken and any changes in medications since the previous visit and C-SSRS will be obtained, in addition, to other assessments at the study clinician's discretion. Assessments may be performed over several days, but all scales should be completed on the same day.
[3] The "Last 12 Months" version will be administered at screening and the "Since Last Visit" version will be administered at all other visits.
[4] See Protocol Section 7.2.4 for complete list of required tests to be performed.
[5] For women of child-bearing potential only.
[6] Site is to document method of contraception agreed to be used by each participant.
[7] To be administered immediately after the COMP360 administration session.
[8] The EQ-5D-3L will be administered to both the participant and their caregiver (latter is not mandatory).
[9] Longboat access can be granted at V1 or V1a at the study team's discretion.
[10] After baseline data is entered into EDC the CAT team will complete a final review to ensure the participant's continued eligibility. Participants cannot complete COMP360 administration at V3 until this approval is received.
[11] Screening period visits should not be initiated until initial MM and CAT approval is received. Screening visits are to occur weekly 3-6 weeks prior to baseline, number of visits is determined by the length of the participant's taper off antidepressant medication.

5.4 Hypotheses and Treatment Comparisons

The primary efficacy endpoint is the change from baseline in MADRS total score at week 3. The primary analysis will be the comparison between COMP360 (25 mg or 10 mg) versus COMP360 1 mg and will be based on a two-sided hypothesis testing approach. The null and alternative hypotheses associated for this comparison are:

H_0: $\mu_{\text{COMP360 (25 mg or 10 mg)}} - \mu_{\text{COMP360 1 mg}} = 0$

H_1: $\mu_{\text{COMP360 (25 mg or 10 mg)}} - \mu_{\text{COMP360 1 mg}} \neq 0$

Where $\mu_{\text{COMP360 (25 mg or 10 mg)}}$ and $\mu_{\text{COMP360 1 mg}}$ represent the mean in change from baseline in MADRS total score at week 3 for COMP360 (25 mg or 10 mg) and COMP360 1 mg, respectively.

The null hypothesis is that there is no difference between treatment groups for the primary efficacy endpoint and the alternative hypothesis is that there is a difference between treatment groups for the primary efficacy endpoint.

5.5 Multiplicity

To control the overall Type I error rate a sequential test procedure will be applied across the primary and key secondary efficacy endpoints for both COMP360 doses (25 mg and 10 mg) following the hierarchy specified below, with all testing done at the 0.05 alpha level:

- Primary endpoint, change in MADRS total score from baseline to week 3 (25 mg vs 1 mg)
- Primary endpoint, change in MADRS total score from baseline to week 3 (10 mg vs 1 mg)
- Secondary endpoint, proportion of participants with a response (≥50% improvement in MADRS total score from baseline) at week 3 (25 mg vs 1 mg)
- Secondary endpoint, proportion of participants with a response (≥50% improvement in MADRS total score from baseline) at week 3 (10 mg vs 1 mg)
- Secondary endpoint, proportion of participants with remission (MADRS total score ≤10) at week 3 (25 mg vs 1 mg)
- Secondary endpoint, proportion of participants with remission (MADRS total score ≤10) at week 3 (10 mg vs 1 mg)
- Secondary endpoint, proportion of participants who have a sustained response at week 12 (25 mg vs 1 mg)
- Secondary endpoint, proportion of participants who have a sustained response at week 12 (10 mg vs 1 mg)

The above hierarchical procedure will apply only to results from the main analysis of the primary study estimand ("Hypothetical Strategy") and will continue until the first non-significant test is observed. If a non-significant p-value is observed, the remaining endpoints will still have nominal p-values presented but will only be assessed descriptively.

5.6 Sample Size Considerations

The intent of the primary efficacy analysis is to demonstrate superiority of at least one therapeutic dose of COMP360 (10 mg or 25 mg) versus the 1 mg COMP360 dose based on the change from baseline in MADRS total score at week 3. The three treatment groups will be randomised in a 1:1:1 ratio.

For this primary analysis, a sample size of 216 randomised participants (72:72:72) will provide 90% power at the alpha = 0.05 level to detect a 6-point difference in average MADRS total score between the optimal therapeutic dose of COMP360 and COMP360 1 mg, assuming the common standard deviation (SD) is 11.0.

The 6-point difference in average MADRS total score on a group level is supported by the pilot study with psilocybin (non COMP360 formulation)[2] and is within the range of potential differences that have been used to power other studies in major depression and treatment-resistant depression (TRD). The assumed SD of 11.0 is based on a review of other studies for this condition.

It is assumed that up to 90% of randomised participants may not have prior psilocybin experience. The power for this post-hoc subgroup is approximately 86%, if the maximum number of participants to have prior psilocybin experience was 10% of randomised participants.

5.7 Randomisation

Study participants will be randomised at a 1:1:1 ratio to receive either:

- COMP360 1 mg (72 participants)
- COMP360 10 mg (72 participants)
- COMP360 25 mg (72 participants)

6 PLANNED ANALYSES

6.1 Final Analysis

The SAP will be finalised before database lock. Final data analysis will be conducted after database lock. The tables, figures and listings planned in this document will be included in the final analysis.

6.2 Interim Analysis

The protocol stated there may be a blinded interim analysis after approximately 50% of the planned 216 participants had been randomised and had an opportunity to complete three weeks of the study after receiving the study drug. The purpose of this blinded interim would have been to perform an assessment of the overall combined (across all treated participants) SD of the change from baseline in the MADRS collected at week 3. However, upon review, it was decided not to perform a blinded interim analysis and to await the final results.

6.3 Data Safety Monitoring Board (DSMB)

An independent DSMB, composed of experts in the management of participants with the disease under study and a biostatistician, reviewed selected safety data at predefined intervals during the study. The primary purpose of this committee was to review safety data for the protection of participant safety. A DSMB Charter defines the primary responsibilities of the DSMB, COMPASS, and Worldwide, the purpose and timing of meetings, quorum, and voting details. The Charter also provides the procedures for confidentiality, communication, and a description of the deliverables that were provided to and reviewed by the DSMB.

The DSMB advised COMPASS and made recommendations regarding continuation of the study and modification to the protocol or study procedures in order to protect the participants enrolled in the study.

7 ANALYSIS SETS

Analysis Set	Definition
Randomised Analysis Set	All randomised participants, whether or not they receive study drug.
Safety Analysis Set	All randomised participants who receive study drug. Outputs based on this analysis set will use the actual treatment received by the study participant.
Full Analysis Set (FAS)	All participants randomised who receive study drug and have at least 1 post-dose efficacy assessment. Outputs based on this analysis set will use the treatment the study participant was randomised to.
Per Protocol (PP) Analysis Set	All participants in the FAS who do not have a protocol deviation (PD) that is thought to significantly affect the integrity of the participant's efficacy data.

The Safety Analysis Set will be used for all safety-related evaluations. The Randomised and Safety Analysis Sets will be used for study participant-related evaluations. The FAS will be used for all efficacy-related evaluations, with the PP Analysis Set being used for sensitivity analyses for certain efficacy-related evaluations. Outputs based on the FAS, PP and Randomised Analysis Set will use the treatment the study participant was randomised to.

PDs thought to significantly affect the integrity of the particpant's efficacy data will be identified and documented during the blinded data review meeting prior to database lock and confirmed at the time of database lock, prior to routine study unblinding.

8 GENERAL CONSIDERATIONS

All TFLs will be created using SAS® version 9.4 or later.[1]

Listings will be sorted in the following order: treatment group, participant, parameter, and visit unless otherwise stated. All data will be listed.

Participant disposition, baseline/demographics and safety summaries will be presented jointly for all treatment arms whereas efficacy analysis and summaries will be split by study.

Unless otherwise specified, "baseline" is defined as the last observed value of the parameter of interest prior to dosing for each period. Unless otherwise stated, if baseline data is missing no derivation will be performed and baseline will be set to missing. For numerical variables, change from baseline will be calculated as the difference between the value of interest and the corresponding baseline value.

Continuous data will be summarised descriptively using N (number of participants in the analysis set), n (number of observations), mean, SD, median, minimum and maximum. Categorical data will be summarised using frequency counts and percentages.

The denominator for a percentage will be the total number of participants in the relevant treatment group or analysis set, unless otherwise specified (on some occasions, percentages may be calculated using the total number of participants with available data at a particular visit or time point as the denominator).

Summaries will be presented by treatment (and overall, where relevant), unless otherwise specified.

Unless otherwise specified, all formal statistical tests will be performed at the 5% two-sided significance level.

Unscheduled visits and retests (same visit number assigned) will not be displayed in by-visit summary tables but will be included in the data listings.

All individual participant data will be listed. Listings will include scheduled, unscheduled, retest and early study discontinuation data.

All medical histories and AEs will be coded from the actual term using the Medical Dictionary for Regulatory Activities (MedDRA) version 23.0 or later.

All prior and concomitant medications will be coded using the World Health Organization Drug Dictionary Enhanced (WHO-DDE) (March 2020 Version or later).

9 DATA DERIVATIONS

This section describes the derivations required for statistical analysis. Unless otherwise stated, variables derived in the source data will not be re-calculated.

9.1 General

9.1.1 Change from Baseline

Definition	Derivation
Change from baseline	$= post\text{-}dose\ visit\ value - baseline\ value$
Change from baseline (%)	$=\left[\frac{post\text{-}dose\ visit\ value - baseline\ value}{baseline\ value}\right] \times 100$

If there is no baseline or post-dose visit value, then both the absolute and percentage change from baseline will be set to missing.

9.1.2 Missing and Partial Dates

All rules explained below for partial / missing dates will be followed unless contradicted by any other data recorded on the electronic Case Report Form (eCRF).

All dates presented in the individual participant listings will be as recorded on the eCRF. In case of partial eCRF entries, the rules described in Sections 9.2.5 and 9.4.2 will be used for date imputation for the purposes of deriving ancillary quantities (e.g. study day, duration of an event, etc) but only partial dates will be presented in the listings.

9.1.3 Treatment Exposure and Compliance

Since administration of COMP360 occurs only once, calculations of treatment exposure and compliance are not applicable.

9.1.4 Inexact Values

In the case where a variable is recorded as "> x", "≥ x", "< x" or "≤ x", a value of x will be taken for analysis purposes. This rule will be applied to continuous laboratory parameter values only for summary tables, whereas listings will show the actual value recorded (inclusive of the non-numeric symbol).

9.2 Study Participants

9.2.1 Protocol Deviations

A PD is any change, divergence or departure from the protocol or investigational plan. PDs will be categorised as follows (detailed PD definitions can be found in the protocol deviation handling plan:

1. 'Important' from a reporting (TFL/CSR) perspective

2. PDs that would exclude the participant from the PP analysis set

9.2.2 Age Categories

The following age categories will be derived:

- 18-34 years
- 35-64 years
- 65-84 years
- > 84 years

9.2.3 Body Mass Index (BMI)

The BMI will be recalculated at screening using the formula:

$$weight\ (kg)/height\ (m)^2$$

BMI will not be calculated for participants enrolled under protocol versions prior to version 4.0 because height was not planned to be collected at screening for earlier protocol versions.

9.2.4 Length of Current Depressive Episode

The length in months of the current depressive episode will be calculated as follows:

$$12 \times \left[\frac{screening\ date - start\ date\ of\ the\ depressive\ episode}{365.25}\right]$$

The length of the current depressive episode will also be categorised as follows:

- < 1 year
- ≥ 1 year to < 2 years
- ≥ 2 years

9.2.5 Missing Major Depressive Disorder (MDD) Episode Start Dates

If the month and year for the start date of the current MDD episode are present but the day is missing, the diagnosis date will be set to the first day of the relevant month. If only the year is recorded the diagnosis date will be set as "01-Jan" for that year.

This imputation rule will only be implemented to derive the length of the current depressive episode (see Section 9.2.4), and the imputed date will not be used elsewhere (ie partial dates will be presented in the listings, as recorded).

9.2.6 Prior Psilocybin Experience

A participant will be considered to have a prior psilocybin experience if they have had a prior psychedelic experience that was greater than 12 months prior to screening and that this experience was psilocybin (if the experience is missing then this will be considered as having a prior psilocybin experience). In addition, the date of the last psilocybin experience must have been earlier than the current MDD start date.

9.2.7 Prior and Concomitant Medications

Prior medication refers to any medication that was stopped prior to the start of the study/taken during the 30-day period before signing the ICF. Concomitant medication refers to the use of any ongoing medication use at the time of the start of the study/all medications taken after the ICF has been signed.

9.3 Efficacy Assessments

9.3.1 Montgomery-Asberg Depression Rating Scale Score (MADRS)

The MADRS is a clinician rated scale measuring depression severity, consisting of the following 10 individual items:

1. reported sadness
2. apparent sadness
3. inner tension
4. reduced sleep
5. reduced appetite
6. concentration difficulties
7. lassitude
8. inability to feel
9. pessimistic thoughts
10. suicidal thoughts

Each item has a score ranging from 0 (normal) to 6 (severe), with higher scores denoting greater severity.

MADRS data will be obtained at baseline, day 2 (day 1 according to the study protocol, see Section 17.2 for mapping), and weeks 1, 3, 6, 9, and 12.

9.3.1.1 Total Score

The MADRS total score (ranging from 0 – 60) is derived by summing up the scores from the 10 individual items.

9.3.1.2 MADRS Responder Status

A participant is defined as a MADRS responder at a post-baseline visit if they meet the following:

$$-50\% \leq change\ from\ baseline\ (\%)$$

The derivation of MADRS change from baseline (%) is defined in Section 9.1.1.

9.3.1.3 MADRS Remitter Status

A MADRS remitter is defined as a participant whose MADRS total score value is ≤10 at a post-baseline visit.

9.3.1.4 MADRS Sustained Response

A participant is defined as a MADRS sustained responder to treatment if they fulfil the MADRS response criteria defined in Section 9.3.1.2 at any visit up to and including week 3, and also fulfil MADRS response criteria at all subsequent visits up to and including week 12. In practice, this means that to be a sustained responder at week 12 a participant needs to satisfy the MADRS response criteria at all timepoints from week 3 to week 12. The table below summarises the scenarios that would lead to assigning a sustained response status at week 12:

Day 2	**Week 1**	**Week 3**	**Week 6**	**Week 9**	**Week 12**
O	O	O	O	O	O
O	X	O	O	O	O
X	O	O	O	O	O
X	X	O	O	O	O

O: Response; X: Non-Response.

A further relaxed definition of sustained responder is if a participant fulfils the MADRS response criteria at any visit up to and including week 3 and also fulfil the response criteria at week 12 and at least at one of week 6 or week 9 (where the previous main definition requires the participant to qualify as a responder at week 6, 9 and 12 to be classified as sustained responder). A similar table to the one presented above for the main sustained response definition is displayed below. The key difference between the two tables is that the 4 combinations for day 2 to week 3 that make up the previous table are now tripled to include the following options for week 6 to 12:

- Participant is a MADRS responder at all assessments from week 6 to week 12 (first 4 lines, same as previous table)
- Participant is a MADRS responder at week 6 and 12 but not at week 9 (lines 5 to 8)
- Participant is a MADRS responder at week 9 and 12 but not at week 6 (lines 9 to 12)

Day 2	Week 1	Week 3	Week 6	Week 9	Week 12
O	O	O	O	O	O
O	X	O	O	O	O
X	O	O	O	O	O
X	X	O	O	O	O
O	O	O	O	X	O
O	X	O	O	X	O
X	O	O	O	X	O
X	X	O	O	X	O
O	O	O	X	O	O
O	X	O	X	O	O
X	O	O	X	O	O
X	X	O	X	O	O

O: Response; X: Non-Response.

9.3.1.5 MADRS Partial Responder Status

A participant is defined as a MADRS partial responder at a post-baseline visit if they meet the following:

$$-25\% \leq change\ from\ baseline\ (\%) < -50\%$$

The derivation of change from baseline (%) is defined in Section 9.1.1.

9.3.1.6 Severity Categories

Baseline MADRS scores will be categorised as indicators of depression severity using the following categories[3]:

- ≤ 10 = Subthreshold
- 11 to 19 = Mild
- 20 to 30 = Moderate
- ≥ 31 = Severe

9.3.2 Quick Inventory of Depressive Symptomatology (QIDS-SR-16)

The 16-item self-rated scale is designed to assess the severity of depressive symptoms. The following 9 symptom domains are used to obtain the total score:

1. The highest score on any 1 of the 4 sleep items (1-4)
2. The score on item 5
3. The highest score on any 1 of the appetite/weight items (6-9)
4. The score on item 10
5. The score on item 11
6. The score on item 12
7. The score on item 13

8. The score on item 14
9. The highest score on either of the 2 psychomotor items (15 and 16)

The total score is obtained by summing the 9 symptom scores. The total score ranges from 0 to 27 with 0 representing no depression and 27 representing severe depression.

QIDS-SR-16 will be obtained at screening, baseline and all post-baseline timepoints up to week 12.

9.3.3 European Quality of Life 5-dimension 3-level Scale (EQ-5D-3L)

The EQ-5D-3L is a multi-attribute instrument used in assessing the health-related quality of life (HRQoL) and includes two parts:

- The EQ-5D descriptive system contains five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has three levels: 1 = no problems, 2 = some problems, 3 = severe problems. The responses to the five EQ-5D dimensions will be converted into a 5-digit number (eg 12212) called a 'health state'. The health state reflects how good or bad the health state is according to the preferences of the general population of a country/region.
- The EQ visual analogue scale (VAS) records the participant's self-rated health on a vertical VAS, where participants are asked to mark their self-rated health on a scale from 0 ('worst imaginable health state') to 100 ('best imaginable health state').

In order to convert the health state to a continuous index value, value sets (ie weights ranging from 1 for a health state of 11111 to negative values for health states reflective of a poor condition such as 33333) will be retrieved from the EuroQoL website[4]. As suggested by EuroQoL, one collection of value sets will be used for all countries in the study, namely those for the United Kingdom (UK). The derivation is described in a paper from Dolan et al.[5] and briefly described below.

For each dimension of the EQ-5D-3L, two variables are defined: the first one takes values of 1 (if the dimension has a value of 2), 2 (if the dimension has a value of 3) and 0 otherwise, whereas the second takes a value of 1 if the dimension has a value of 3 and 0 otherwise. In addition, an 'intercept' is defined, which can either be 0 (if all dimensions have a value of 1) or 1 otherwise, and a so-called interaction term, which can either be 0 (if no dimension has a value of 3) or 1 otherwise (at least one dimension has a value of 3). For each of these 12 variables a regression coefficient is used to derive the 'weight' associated with a given 5-digit health state. The model coefficients (taken from Table 1 of Dolan paper[5]) are reported in the table below:

Model term	**Regression Coefficient**
Intercept (any dimension > 1) – D1	0.081
Mobility (Level 2) - MO	0.069

Model term	Regression Coefficient
Self-care (Level 2) - SC	0.104
Usual Activities (Level 2) - UA	0.036
Pain/Discomfort (Level 2) - PD	0.123
Anxiety/Depression (Level 2) - AD	0.071
Mobility (Level 3) – MO2	0.176
Self-care (Level 3) – SC2	0.006
Usual Activities (Level 3) – UA2	0.022
Pain/Discomfort (Level 3) – PD2	0.140
Anxiety/Depression (Level 3) – AD2	0.094
Interaction (any dimension = 3) – N3	0.269

The continuous values are then derived using the following formula:

$$EQ\text{-}5D\text{-}3L_i = 1 - D1*\beta_0 - MO*\beta_1 - SC*\beta_2 - UA*\beta_3 - PD*\beta_4 - AD*\beta_5 - MO2*\beta_6 - SC2*\beta_7 - UA2*\beta_8 - PD2*\beta_9 - AD2*\beta_{10} - N3*\beta_{11}$$

Where β_j are the coefficients regression from the above table. As an example, a health state of 12212 would have an associated continuous value derived as follows:

- The following variables have a value of 0: MO, PD, MO2, SC2, UA2, PD2, AD2, N3. Thus, the respective coefficients are irrelevant in the above formula
- The prediction formula simplifies as follows:

$$EQ\text{-}5D\text{-}3L_i = 1 - 0.081 - 1*0.104 - 1*0.036 - 1*0.071 = 0.708$$

If a component of the EQ-5D-3L was not collected, the 5-digits health state will be recorded with '-' to identify these gaps (e.g. 132-1, implying that the pain/discomfort item was not marked) and the continuous score will not be derived and set to missing for analysis purposes.

EQ-5D-3L data will be obtained at baseline, weeks 3 and 12.

9.3.4 Sheehan Disability Scale (SDS)

This brief, 5-item self-report inventory assesses functional impairment in work/school (score ranges from 0 to 10), social life (score ranges from 0 to 10), and family life (score ranges from 0 to 10). The sum of the above 3 domains scores lead to a total score ranging from 0 to 30 with 0 representing no impairment and 30 representing severe impairment. If a participant ticks the 'I have not worked/studied at all during the past week for reasons unrelated to the disorder' box, then no SDS total score can be derived (irrespective of whether the work/school item was answered or not) and only the individual answers for that participant will only be included in the listings.

SDS data will be obtained at baseline, weeks 3 and 12.

9.3.5 Digit Symbol Substitution Test (DSST)

The DSST consists of a look-up table showing pairs of digits and hieroglyphic-like symbols and rows of boxes with a digit in the top section and an empty space in the bottom section of each box. A participant's score is the number of correct symbols within the allowed time (90 sec).

The number of correct symbols will be categorised for reporting purposes as follows: ≤29, > 29 and ≤ 39, > 39 and ≤ 48 and > 48.

DSST data will be obtained at baseline, day 2 weeks 3 and 12.

9.3.6 Generalised Anxiety Disorder Scale (GAD-7)

The Generalised Anxiety Disorder (GAD)-7 is a 7 item self-report anxiety questionnaire designed to assess the participant's health status during the previous 2 weeks. Scores of 0, 1, 2 or 3 are given for experiencing symptoms 'not at all', for 'several days', for 'more than half the days' and for 'nearly every day', respectively. The scores are then totalled and presented from 0 to 21.

GAD-7 data will be obtained at baseline, weeks 3 and 12.

9.3.7 Work and Social Adjustment Scale (WSAS)

The Work and Social Adjustment Scale (WSAS) is a 5-item self-report scale used to assess psychosocial functioning. Each of the 5 questions is rated on a scale from 0 to 8, where 0 is no impairment and 8 is very severe impairment. The total WSAS score is obtained by adding up responses for each individual question, with a range from 0 to 40. A WSAS score above 20 appears to suggest moderately severe or worse psychopathology. Scores between 10 and 20 are associated with significant functional impairment but less severe clinical symptomatology. Scores below 10 appear to be associated with subclinical populations.

WSAS data will be obtained at baseline, weeks 3 and 12.

9.3.8 Mini International Neuropsychiatric Interview

The Mini International Neuropsychiatric Interview (MINI) is the structured psychiatric interview of choice for psychiatric evaluation and outcome tracking in clinical psychopharmacology trials and epidemiological studies. The MINI is divided into modules corresponding to a diagnostic category. The MINI has similar reliability and validity properties compared to Structured Clinical Interview for DSM Disorders using Version 5 of the DSM (SCID-5) and the Composite International Diagnostic Interview (CIDI), a structured interview developed by the World Health Organization. In this study, version 7.0.2 of the MINI will be used.

MINI data will be obtained at screening only.

9.3.9 Hamilton Depression Rating Scale – 17-Item (HAM-D-17)

The Hamilton Depression Rating scale (HAM-D) has been used in determining a participant's level of depression before, during, and after treatment. The questionnaire is available in both a 17 and 21 item version. In this study, the 17-item version was utilised. Nine items are scored on a 5-point scale ranging from 0 to 4, 7 items are scored on a 3-point scale ranging from 0 to 2, and the remaining item is scored on a 4-point scale ranging from 0 to 3. The total score is obtained as the sum of the individual items scores from this assessment (values ranging from 0 to 53) and will be used as an eligibility criterion prior to treatment (minimal total symptom score ≥ 18).

HAM-D-17 total scores at baseline will be categorised as moderate (18-23) or severe (≥24) depression.

HAM-D-17 data will be obtained at screening and baseline only.

9.3.10 The Massachusetts General Hospital Antidepressant Treatment Response Questionnaire (MGH-ATRQ)

The MGH-ATRQ is a self-rated scale used to determine treatment-resistance in major depressive disorder (MDD). The MGH-ATRQ defines 6 weeks on an adequate dose of antidepressant medication as an adequate duration of treatment. It also provides specific operational criteria for adequate dosage for each of the most commonly used antidepressants. The ATRQ examines the efficacy using the following categorised ratings: <25% improved; 25%-49% improved; 50%-75% improved and >75% improved. A rating of 100% is completely improved and a rating of 0% is not improved at all.

MGH-ATRQ data will be obtained at screening only.

9.3.11 McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD)

The MSI-BPD is a commonly used measure to assess for BPD. The scale consists of 10 items based on the DSM-5 BPD criteria; the first 8 items represent the first eight criteria in the DSM-5 for BPD diagnosis, while the last two questions assess the paranoia and dissociation criteria for BPD. Scores for the MSI-BPD range from 0 to 10, with each item rated as "1" if present and "0" if absent. A score of 7 or higher (obtained as the sum of the individual items scores) indicates a likelihood for the participant to meet criteria for BPD.

The MSI-BPD will be obtained at screening only.

9.3.12 Adult Self-Report Scale (ASRS)

The ASRS is a self-reported questionnaire used to determine the presence of attention deficit hyperactivity disorder (ADHD) in adults. The screener (6-item) version of the scale based on the DSM-5 will be used in this study. The first question concerns inattention and the other 5 questions assess hyperactivity-impulsivity. Each question is answered on a 5-

point Likert-type scale with the following possible outcomes: 0 = "Never", 1 = "Rarely", 2 = "Sometimes", 3 = "Often" and 4 = "Very Often". The scoring system follows what described in literature[6], where the total score is obtained by summing the scores for each item, with a total ranging from 0 to 24.

ASRS data will be obtained at screening only.

9.3.13 Scale to Assess Therapeutic Relationships – Patient (STAR-P) and Clinician (STAR-C) Versions

The STAR-P and STAR-C are 12-item measures assessing the therapeutic relationship between the participant and therapist on three components: Collaboration, Positive Clinician Input, and Emotional Difficulties (clinician version)/Non-Supportive Clinician (patient version) input. The Collaboration subtest reflects a good rapport and a shared understanding of goals, mutual understanding, openness, and trust. Positive Clinician Input reflects the perception (by the participant) of the clinician to encourage, support, and listen to the participant. Emotional Difficulties/Non-Supportive Clinician Input reflect problems in the relationship. Each of the 12 items are rated on a scale from 0 (never) to 4 (always). The total scores are obtained by adding up answers to each item and range between 0 and 48 for both versions, with a higher score suggesting better therapeutic relationships.

The STAR-P and STAR-C will be obtained at baseline only.

9.3.14 Five Dimensional Altered States of Consciousness Questionnaire (5D-ASC)

The 5D-ASC measures the acute drug effects using 5 primary dimensions and respective subdimensions to assess alterations in mood, perception, and experience of self in relation to environment and thought disorder. The 5 dimensions include oceanic boundlessness, anxious ego dissolution, visionary restructuralization, auditory alterations, and reduction of vigilance. In addition to these 5 dimensions, the 5D-ASC can be split into 11 subscales, as described by Studerus et al.[7]: experience of unity, spiritual experience, blissful state, insightfulness, disembodiment, impaired control and cognition, anxiety, complex imagery, elementary imagery, audio-visual synaesthesia and changed meaning of percepts. Participants were instructed to respond to the described experiences by placing vertical marks on a horizontal VAS 100 millimetres long. The VAS of the altered states of consciousness rating scales (OAV and 5D-ASC) are anchored as "No, not more than usual" on the left and as "Yes, much more than usual" on the right. The items are scored by measuring the millimetres from the low end of the scale to the participant's mark (integers from 0–100). Because the low end of the scale indicates a neutral response, the response format of these items can be considered as strictly unipolar according to the response format typology of Russell and Carroll.[7-8]

The mapping of individual items to the 5 dimensions and to the 11 subscales is displayed in the table below (the former accounts for 93 out of 94 individual items [item 66 is excluded], whereas the latter only accounts for 42 items, as described in the paper from Studerus et al.[7]). Subscales have been ordered and displayed in accordance with the

relevant dimension they are included in (apart for the last 2 dimensions for which no subscale is defined)[9].

Dimension	Items	Subscale	Items
Oceanic Boundlessness	1, 3, 9, 12, 16, 18, 26, 34, 35, 36, 40, 41, 42, 45, 50, 52, 57, 62, 63, 69, 71, 73, 81, 86, 87, 91, 94	Experience of Unity	18, 34, 41, 42, 52
		Spiritual Experience	9, 81, 94
		Blissful State	12, 86, 91
		Insightfulness	50, 69, 77
Anxious Ego Dissociation	6, 8, 21, 27, 32, 38, 43, 44, 46, 47, 53, 56, 60, 64, 67, 78, 79, 80, 85, 88, 89	Disembodiment	26, 62, 63
		Impaired Control and Cognition	8, 27, 38, 47, 64, 67, 78
		Anxiety	32, 43, 44, 46, 56, 89
Visual Restructuralization	7, 14, 20, 22, 23, 28, 31, 33, 39, 54, 58, 70, 72, 75, 77, 82, 83, 90	Complex Imagery	39, 72, 82
		Elementary Imagery	14, 22, 33
		Audio-Visual Synaesthesia	20, 23, 75
		Changed Meaning of Percepts	28, 31, 54
Auditory Alterations	4, 5, 11, 13, 19, 25, 30, 48, 49, 55, 65, 74, 76, 92, 93	-	-
Reduction of Vigilance	2, 10, 15, 17, 24, 29, 37, 51, 59, 61, 68, 84	-	-

The overall score for each subscale and dimension is obtained by averaging the score for each individual item within that subscale/dimension.

5D-ASC data will be obtained at day 1 immediately after the COMP360 session.

9.3.15 Emotional Breakthrough Inventory (EBI)

The EBI is an 8-item brief measure intended to index the degree to which an individual experiences their emotion during the COMP360 session. It is a VAS, with units from 0 to 100, typically rated within 24 h of a psychedelic experience and ideally within 5 hours of the 'end' of the psychedelic experience or once acute drug effects have significantly subsided. An overall EBI score is obtained by averaging individual item responses.

EBI data will be obtained at day 2 only.

9.3.16 The Positive and Negative Affect Schedule (PANAS)

The PANAS measures the acute emotional drug effects and comprises 2 mood scales that measure positive and negative affect. Participants respond to 20 items using a 5-point scale that ranges from “very slightly or not at all (1)” to “extremely (5)”. The Positive Affect score is obtained by summing up scores for items 1, 3, 5, 9, 10, 12, 14, 16, 17, and 19, whereas the Negative Affect score is obtained by summing up the scores for items 2, 4, 6,

7, 8, 11, 13, 15, 18, and 20. A total higher score on the positive affect questions (ranging from 10 to 50) indicates higher levels of a Positive Affect while a lower score on the negative affect questions indicates lower levels of a Negative Affect.

PANAS data will be obtained at baseline, day 2 and week 3.

9.4 Safety

9.4.1 Adverse Events

A treatment-emergent adverse event (TEAE) is defined as any AE that has an onset on or after the dose of study drug, or any pre-existing AE condition that has worsened on or after the dose of study drug.

A treatment-related TEAE is defined as an AE reported by the investigator to be possibly related or related to the study drug. If an AE has missing relationship, it will be assumed to be related to the study drug for analysis purposes. For reporting purposes, a treatment-related TEAE will be referred to as an adverse drug reaction (ADR) and a serious ADR will be referred to as a serious adverse reaction (SAR).

Maximum severity will be assumed for an AE with missing severity.

AEs will be categorised by time of onset (day 1, day 2, day 3-8, > day 8) and duration (≤ 1 day, > 1 and ≤ 2 days, > 2 and ≤ 7 days, >7 days or ongoing).

Adverse Events of Special Interest (AESI) will be identified as follows:

AESI term	**MedDRA Preferred Terms**
Euphoric mood	Euphoric mood
Dissociative disorder	Dissociative disorder; Dissociative identity disorder; Dissociative amnesia; Dissociation; Depersonalisation/derealisation disorder
Hallucination	Hallucination; Hallucination, auditory; Hallucination, synaesthetic; Hallucination, tactile; Hallucination, visual; Hallucination, olfactory; Hallucinations, mixed; Somatic hallucination
Psychotic disorder	Psychotic disorder; Psychotic behaviour; Acute psychosis; Hysterical psychosis; Reactive psychosis; Substance-induced psychotic disorder; Brief psychotic disorder with marked stressors; Brief psychotic disorder without marked stressors; Transient psychosis
Cognitive disorder	Cognitive disorder
Disturbance in attention	Disturbance in attention
Mood altered	Mood altered; Depressed mood; Affect lability; Fluctuating mood symptoms
Psychomotor skills impaired	Psychomotor skills impaired
Inappropriate affect	Inappropriate affect
Overdose	Overdose

AESI term	MedDRA Preferred Terms
Intentional product misuse	Intentional product misuse

In addition AESIs will include those AEs identified as AESIs by the investigator in the eCRF. In the event that the associated preferred term does not feature in the list above then they will be assigned an AESI term of "Other".

9.4.2 Missing / Partial Start / Stop Dates of AEs and Concomitant Medications

Missing and partial start and stop date will be imputed for analysis purposes as detailed below.

Partial or missing stop date will be imputed as follows:

- If the stop date is completely missing and the event has resolved, or the participant has stopped taking the concomitant medication, the stop date will be imputed as the date of the participant's last clinic visit in the study.
- If only the year is known, and it is equal to the year of the participant's last clinic visit in the study, the stop date will be imputed as the last clinic visit date, otherwise if the years differ it will be imputed as "31-Dec".
- If the month and year are known, the stop date will be imputed as the last day of that month unless the stop date corresponds to the same month as the participant's last clinic visit in which case the date of participant's last clinic visit in the study will be used instead.

Missing start date will be imputed as follows:

- If the stop date occurs on or after the dose of study drug or the event / concomitant medication is ongoing, the start date will be imputed as the date of the dose of study drug.
- If the stop date occurs before the start of study drug, the start date of the event / concomitant medication will be imputed as the participant's screening date or the stop date of the event / concomitant medication whichever the earlier.

Partial start date (year present, but month and day missing)

- If the stop date occurs on or after the dose of study drug or the event / concomitant medication is ongoing, and the year is the same as the year of dosing the start date will be imputed as the date of the dose of study drug. If the year is different from the year of dosing "01-Jan" will be used.
- If the stop date occurs before the dose of study drug, the start date of the event / concomitant medication will be imputed as the "01-Jan" of the same year.

Partial start date (month and year present, but day missing)

- If the stop date occurs on or after the dose of study drug or the event / concomitant medication is ongoing, the start date will be imputed as the first day of the same month and year unless this partial start date is in same month as the dose of study drug in which case the date of dose of study drug will be used.
- If the stop date occurs before the dose of study drug, the start date will be imputed as the first day of the month and year of the partial start date.

If the start time of an AE is missing it will be imputed only in the case where the start date corresponds to the date of the dose of study drug. The time will be imputed as the same time as the dose of study drug. In all other cases the time will not be imputed.

9.4.3 Vital Signs

Starting with protocol version 4.0, blood pressure measurements were collected as triplicate measurements. For analysis purposes replicate measurements (where available) have been averaged and the resulting average used in all summaries and listings.

9.4.3.1 Vital Signs Ranges of Clinical Importance

Values outside the following ranges for the vital signs parameters will be considered clinically important:

Vital sign	**Lower Limit**	**Upper Limit**
Systolic blood pressure (mmHg)	90 mmHg	160 mmHg
Diastolic blood pressure (mmHg)	50 mmHg	100 mmHg
Pulse rate (bpm)	50 bpm	100 bpm
Respiration rate (breaths / min)	11 breaths/min	20 breaths/min
Body temperature (°C)	-	37.5 °C

9.4.4 ECG Categorical Intervals

The ECG categorical intervals of interest are:

- QTcF ≤ 450 msec
- 450 < QTcF ≤ 480 msec
- 480 < QTcF ≤ 500 msec
- QTcF > 500 msec
- Change from baseline in QTcF interval
  - QTcF interval increases from baseline > 30 and ≤ 60 msec
  - QTcF interval increases from baseline > 60 msec

9.4.5 Columbia-Suicide Severity Rating Scale (C-SSRS)

The following outcomes are C-SSRS categories and have binary responses (yes / no). The categories have been re-ordered from the actual scale to facilitate the definition of composite endpoints:

Category 1	Wish to be Dead
Category 2	Non-specific Active Suicidal Thoughts
Category 3	Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
Category 4	Active Suicidal Ideation with Some Intent to Act, without Specific Plan
Category 5	Active Suicidal Ideation with Specific Plan and Intent
Category 6	Preparatory Acts or Behaviour
Category 7	Aborted Attempt
Category 8	Interrupted Attempt
Category 9	Actual Attempt (non-fatal)
Category 10	Completed Suicide

Suicidal Ideation since baseline – A "yes" answer at any time during double blind treatment to any one of the 5 suicidal ideation questions (categories 1-5) on the C-SSRS.

Suicidal Behaviour since baseline – A "yes" answer at any time during double blind treatment to any one of the 5 suicidal behaviour questions (categories 6-10) on the C-SSRS.

There will be no imputation of missing data for C-SSRS.

9.4.6 Sheehan Suicidality Tracking Scale (S-STS)

The S-STS is a 16-item scale that assesses the seriousness of suicidality phenomena on a Likert-type scale (0–4) ranging from "not at all" (0) to "extremely" (4). In addition, 6 extra questions are to be completed from the clinician in case a participant does not return for a scheduled follow-up visit to identify the reason for the missed appointment.

At screening, a significant risk of suicide is defined as (a) a score of "3" or "4" on Questions 2 or 13; or (b) a score of "2" or higher on Questions 1a, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 14 over the past 13 months.

A significant risk of suicide during the study is defined as (a) a score of "3" or "4" on Questions 2, 3, 4, 5, 6, 7, 8, or 13; or (b) a score of "2" or higher on any of Questions 1a, 9, 10, 11, 12, 14, 15, 20; or (c) if Question 17 is a "Yes" (suicide results in death).

Additionally, if a participant responds with a score of "2" or higher on any of Questions 1a, 3, 4, 5, 6,7, 8, 9, 10, 11, 12, 14, or 20, or if any responses to items of concern based on clinical judgment, the medical monitor should be contacted to determine need for reporting as an AE or SAE.

9.4.7 Algorithm to Map Sheehan Suicidality Tracking Scale and Columbia-Suicide Severity Rating Scale

Version 4 of the protocol (dated 22 July 2019) replaced the S-STS with the C-SSRS and was finalised when a number of study participants had been already enrolled into the study. For analysis purposes these two scales will be mapped to universal categories. Sheehan et al. (2014)[10] provided detailed instructions on the algorithms they used to map item responses from the S-STS and C-SSRS to the FDA Classification Algorithm of Suicide Assessment (FDA-CASA) 2012 categories. The table below presents these categories.

FDA-CASA 2012 CODE NUMBER	FDA-CASA 2012 CLASSIFICATION	S-STS MAP TO FDA-CASA 2012	C-SSRS MAP TO FDA-CASA 2012
1	Passive suicidal ideation	A positive response to 2	A "yes" response to item 1: "Wish to be Dead"
2	Active suicidal ideation: nonspecific (no method, intent, or plan)	A positive response to 3 AND a negative response to 4, 5, 6, and 7	A "yes" response to item 2: 'Nonspecific Active Suicidal Thoughts"
3	Active suicidal ideation: method, but no intent or plan	A positive response to 3 and 4 AND a negative response to 5, 6, and 7	A "yes" response to item 3: "Active Suicidal Ideation With Any Methods (no plan) Without Intent to Act"
4	Active suicidal ideation: method and intent, but no plan	A positive response to 3 and 4 and (6 or 7) and a negative response to 5	A "yes" response to Item 4: "Active Suicidal Ideation With Some Intent to Act, Without Specific Plan"
5	Active suicidal ideation: method, intent, and plan	A positive response to 3 and 4 and (6 or 7) and a negative response to 5	A "yes" response to item 5: "Active Suicidal Ideation With Specific Plan and Intent"
6	Completed suicide	A "yes" response to 13	NA
7	Suicide attempt	A positive response to 10 or 16 OR a "yes" response to 1b	A "yes" response to "Actual Attempt"
8	Interrupted suicide attempt	A positive response to 8 with at least one Level 3 on 12	A "yes" response to "Interrupted Attempt"
9	Aborted suicide attempt	A positive response to 8 with at least one Level 2 on 12	A "yes" response to "Aborted Attempt"
10	Preparatory acts toward imminent suicidal behavior"-"not counting aborted or interrupted attempts	A positive response to 8 with at least one Level 1 on 12	A "yes" response to "Preparatory Acts or Behavior"

FDA-CASA 2012 CODE NUMBER	FDA-CASA 2012 CLASSIFICATION	S-STS MAP TO FDA-CASA 2012	C-SSRS MAP TO FDA-CASA 2012
11	Self-Injurious Behavior Without Suicidal Intent	A positive response to 9 OR a positive response to 1a and a negative response to 1b	A "yes" response to "Has subject engaged in Non-Suicidal Self-Injurious Behavior?"
12	Self-injurious behavior, intent unknown	A positive response to 1a with 1b and 6 and 7 and 10 unanswered	NA
13	Not enough information (fatal)	A "yes" response to 14	NA
14	Not enough information (nonfatal)	Yes, if question 17 is "yes" and FDA-CASA 2012 code numbers 1-13 are ALL "no"	NA
15	Other (accidental, psychiatric medical), no deliberate self-harm	(A positive or blank response to 1 and a negative response to questions 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, and 14) OR a positive response to 15	NA

Abbreviations: C-SSRS, Columbia-Suicide Severity Rating Scale; NA, not applicable; S-STS, Sheehan Suicidality Tracking Scale.

10 STUDY PARTICIPANTS

10.1 Disposition of Participants

The summaries will be presented for each treatment group and overall.

The number of participants screened, screen failures, and the number and percentage of participants in each analysis set will be summarised by country (including overall category) for all screened participants (*Table 14.1.1.1*). The number and percent of participants who completed the study and number and percentage of participants who discontinued from the study, with reasons for discontinuation, will be summarised using the Randomised Analysis Set (*Table 14.1.1.2*). The reasons for screen failures will also be tabulated (*Table 14.1.1.3*). Reasons for exclusion from the PP Analysis Set will be summarised using the FAS (*Table 14.1.1.4*).

Participant disposition will be listed by study discontinuations (*Listing 16.2.1.1*) and completers (*Listing 16.2.1.2*) using the Safety Analysis Set. Participation in each defined analysis set, including exclusion from the PP Analysis Set will be listed using the Randomised Analysis Set (*Listing 16.2.3*). Inclusion criteria not met/exclusion criteria met at screening and at baseline will be listed for all screened participants (*Listing 16.2.4.3 and Listing 16.2.4.4*).

A summary of missed visits or visits that were conducted remotely rather than at the clinic because of COVID-19 will be presented for the Safety Analysis Set by treatment group and overall (*Table 14.1.1.5*), and a listing created which will also include missed assessments, if any (*Listing 16.2.1.3*).

10.2 Protocol Deviations

The frequency and percentage of participants in each important PD category will be summarised and presented for each treatment group and overall using the Safety Analysis Set (*Table 14.1.1.6*). All PDs will be listed using the Safety Analysis Set *(Listing 16.2.2.1)*.

In addition, a summary table of COVID19-related PDs as well as a separate listing will also be presented (*Table 14.1.1.7*, *Listing 16.2.2.2*).

10.3 Demographic and Baseline Characteristics

The Safety Analysis Set will be used for all demographic and baseline characteristics-related evaluations, unless otherwise specified. The summaries will be presented for each treatment group and overall.

Demographic and baseline characteristics (age [years; continuous and by age categories (see Section 9.2.2)], gender [male, female], race [White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander,

Other], prior psilocybin experience [yes, no; see Section 9.2.6], height [cm], weight [kg], recalculated BMI [kg/m^2] (see Section 9.2.3)) will be summarised (*Table 14.1.2*).

Other baseline characteristics (number of failed treatments in the current depressive episode [2, 3, 4], MADRS severity at baseline [continuous and by severity categories (see Section 9.3.1.5)], HAM-D-17 total score, HAM-D-17 severity at baseline (see Section 9.3.9), MGH-ATRQ [<25% improved, 25-49% improved, 50-75% improved, >75% improved], STAR-P total score, STAR-C total score, ASRS total score, MSI-BPD total score, length of current depressive episode [years; continuous and by length categories (see Section 9.2.4)], 5-HT2A polymorphism and number of lifetime depressive episodes as collected in question A6 on the MINI questionnaire) will be summarised (*Table 14.1.4.2*). The MINI version 7.0.2 will be summarised (*Table 14.1.4.1*) – note this summary will only display whether participants meet the criteria for specific modules as displayed on the questionnaire cover pages.

A listing for all demographic and baseline characteristics, and other baseline characteristics, including a listing for the MINI version 7.0.2 will be presented (*Listing 16.2.4.1 and Listing 16.2.9.1 – Listing 16.2.9.9*).

Other baseline measurements (Urine Drug Screen and Pregnancy Test) will be listed only (*Listing 16.2.8.6 and Listing 16.2.8.7*). The pregnancy test listing will include results from both the urine test and, where this was done, of the serum test.

10.4 Medical History

Medical history collected at screening and baseline will be summarised by primary system organ class (SOC) and preferred term (PT) and presented for each treatment group and overall using the Safety Analysis Set (*Table 14.1.3.1*), including both the number and percentage of participants with a given medical history item as well as the number of occurrences of such item.

All medical history data will be listed using the Safety Analysis Set (*Listing 16.2.4.2*).

10.5 Prior and Concomitant Medications

The Safety Analysis Set will be used for all prior and concomitant medication-related evaluations, unless otherwise specified. The summaries will be presented for each treatment group and overall.

Prior and concomitant medications will be summarised (*Table 14.1.3.2 and Table 14.1.3.3*). In addition, a separate summary will be provided for psychoanaleptic and psycholeptic-related concomitant medications by treatment group and by stop date of the medication with respect to COMP360 dose (pre-dose/post-dose) (*Table 14.1.3.4*).

Prior and concomitant medications will be listed (*Listing 16.2.4.5 and Listing 16.2.4.6*), and a separate listing will also be provided for psychoanaleptic and psycholeptic-related medications (*Listing 16.2.4.7*).

10.6 Treatment Exposure

Study drug administration will be listed (*Listing 16.2.5.1*).

11 EFFICACY

The FAS will be used for all efficacy-related evaluations, unless otherwise specified. The summaries and analyses will be presented for each treatment group.

11.1 Primary Efficacy

The primary endpoint is the change from baseline in total MADRS score at week 3.

11.1.1 Hypothetical Strategy Estimand (Primary)

11.1.1.1 Main Analysis

The main analysis will be conducted, in line with the estimand definition, by imputing data-points after the start of a new treatment for depression post COMP360 dosing under a MNAR mechanism, to hypothesise what would have happened to the MADRS total score had new treatment for depression not been available to use. It will be assumed that MADRS values will have a strictly monotone increasing pattern defined by a factor c as follows:

$$y(t) = y(t-1) + c*y(t-1)$$

where *y* is the MADRS total score data point to be imputed, *t* is the time of the measurement and *c* will depend on the ordinality of the data-point following the start of the antidepressant:

First data-point → c = x

Second data-point → c = x/2

Third data-point → c = x/4

Fourth data-point → c = x/8

Fifth data-point → c = x/16

Sixth data-point → c = x/32

The value of *c* to be applied will be confirmed in the blinded data review meeting. For illustration purposes *c* has been assumed as 0.1 in the example below:

First data-point → c = 0.1

Second data-point → c = 0.05

Third data-point → c = 0.025

Fourth data-point → c = 0.0125

Fifth data-point → c = 0.0625

Sixth data-point → c = 0.003125

In this example, this means that for the first data point to be imputed the last evaluable MADRS total score will be increased by 10%, whereas for the second data point the imputation will take the first imputed value and increase it by 5% and so on until all data points have been imputed. Should the first data point to be imputed be preceded by one or more missing data points, the last available MADRS total score measurement prior to the start of new treatment for depression will be used as a starting point.

The starting value of *c* for the first data point to impute will be generated 100 times (using 123 as seed) using a normal distribution with the given *c* value as the mean and SD = *c*/5 (this value for the SD was chosen arbitrarily to ensure a balance between reasonable uncertainty and the need for a strictly positive *c* value without resorting to truncation of negative values). The subsequent c values will be derived according to the above specified rule and MADRS values will be imputed using the approach described above.

Using the example of a starting value for *c* of 0.1, see the table below, where a participant is assumed to have started taking medication at week 3, such that records from week 6 onwards need imputing:

Visit	**Observed**	**Imputed**			
	MADRS Total Score	**Imputation 1 (c = 0.1017)**	**Imputation 2 (c = 0.0905)**	**…**	**Imputation 100 (c = 0.1032)**
Baseline	35	35	35	…	35
Day 2	32	32	32	…	32
Week 1	35	35	35	…	35
Week 3	36	36	36	…	36
Week 6	34***	39.66	39.26	…	39.72
Week 9	15***	41.68	41.03	…	41.77
Week 12	10***	42.73	41.96	…	42.84

*** Denotes observations obtained after the start of new treatment for depression.

The timepoint where the new treatment for depression is started will be identified during a blinded review of the concomitant medication data. If the date where the new medication starts is the same date of a MADRS assessment, that assessment will still be considered 'evaluable' and the above imputation rule applied from the following timepoint.

Non-monotone missing data and missing data occurring after study discontinuation will be imputed under a MAR mechanism, ie assuming that the missing data depend only on the observed values from participants with similar baseline covariates and MADRS score values profile up to the missing timepoint. However missing data occurring after study discontinuation for reasons including (but not limited to) lack of efficacy or AE will be

imputed under the MNAR approach described above. This assumption will be evaluated during the blinded data review meeting and any additional reasons for study discontinuation for which to apply MNAR confirmed.

SAS PROC MI will be applied to the input dataset containing all the observed MADRS total scores at day 2, weeks 1, 3, 6, 9 and 12. The Markov Chain Monte Carlo (MCMC) method will be used for the multiple imputations. One hundred (100) multiple imputed datasets will be created.

A sample SAS code for the MAR imputation is as follows:

```
proc mi data = <input-dataset> nimpute = 100 seed = 99 out = <output-
dataset>;
  by trt01pn;
  var base d2 wk1 wk3 wk6 wk9 wk12;
  mcmc chain = multiple impute = full initial = em prior = jeffreys;
run;
```

where TRT01PN is the treatment group and *d2*, *wk1*, *wk3*, *wk6*, *wk9* and *wk12* are variables containing the MADRS total score at baseline, day 2, weeks 1, 3, 6, 9 and 12, respectively. The IMPUTE = FULL means that all missing data points will be imputed, whether they belong to a monotone missing data pattern or not, and the PRIOR = JEFFREYS specify a non-informative prior for the imputation process, so that the posterior distribution used for the imputation is largely affected by the observed data themselves. Upon completion of the imputation process, all imputed data will be rounded to the closest integer to reflect the potential values that could have been observed had the data points not been missing (that is, an imputed value of 24.7 is certainly wrong whereas a value of 25 is more plausible), and caps will be applied to ensure that the MADRS range (0 to 60) is not exceeded in any direction. For each of the multiple imputed data sets, change from baseline in MADRS total scores will be calculated at each timepoint. The imputed datasets will be merged with the ones obtained for the MNAR imputation of data following start of new treatment for depression or study discontinuation due to lack of efficacy or AE, using participant, visit and imputation as key variables. In doing so, only imputed values for missing data either for participants that never start any new treatment for depression or for missing data prior to the start of such treatment will be retained from the MAR imputation step.

The analysis will be performed using a MMRM model (*Table 14.2.1.1*). The dependent variable will be the change from baseline in MADRS total score at day 2 and weeks 1, 3, 6, 9 and 12 (visits 4, 5, 7, 8, 9 and 10 respectively). The model will include fixed effects for treatment, visit, study site and treatment by visit interaction and baseline MADRS total score as a covariate. Should a study site randomise fewer than 5 participants, a pooling-based approach may be applied for analysis purposes to avoid extreme imbalance in treatment assignment within each level of the study site factor in the model. Details of specific pooling occurrences and full rationale will be determined during the blinded data review meeting.

Correlation between repeated observations within a participant will be accounted for by specifying an unstructured correlation. If this model does not converge then the following

covariance matrices will be used in this particular order until the model converges: Toeplitz, first-order autoregressive, and compound symmetry. Analysis will be performed for each imputed dataset using PROC MIXED in SAS, and the resulting F-tests will be based on using Kenward-Roger's adjusted degrees of freedom. Model assumptions will be examined graphically via inspection of conditional residuals scatterplots (residuals vs predicted values), quantile-quantile plots and histograms. If appropriate, alternative analyses (eg analysis after data transformation or non-parametric analysis) may be performed; these analyses and a detailed motivation for performing them will be detailed in the Clinical Study Report (CSR).

The treatment groups will be compared at all timepoints using a two-sided test at a 5% significance level. The sample SAS code below implements the model detailed above with an unstructured covariance matrix (TYPE = UN in the REPEATED statement). The code assumes that AVISITN has 6 values (one for each visit), and TRT01PN representing treatment is coded as 1 = 25 mg COMP360 dose, 2 = 10 mg COMP360 dose and 3 = 1 mg COMP360 dose.

```
ods output estimates = est LSMeans = lsm diffs = diff;
proc mixed data = <input-dataset> plots (only) = studentpanel(marginal);
by imputation;
class subjid trt01pn siteid avisitn;
model chg = trt01pn siteid avisitn trt01pn*avisitn base / ddfm = kr;
repeated avisitn / type = un subject = subjid;
lsmeans trt01pn*avisitn / diff cl;
estimate "25 mg COMP360 dose vs 1 mg COMP360 dose at visit 3"
        trt01pn 1 0 -1 trt01pn*avisitn 0 0 1 0 0 0 0 0 0 0 0 0 0 0 -1 0 0 0 / cl;
estimate "10 mg COMP360 dose vs 1 mg COMP360 dose at visit 3"
        trt01pn 0 1 -1 trt01pn*avisitn 0 0 0 0 0 0 0 0 1 0 0 0 0 0 -1 0 0 0 / cl;
run;
```

The point estimates (LS means and their differences between treatment arms) and the standard error (SE) will be pooled using Rubin's combination rules[11] to incorporate the between-imputation with the within-imputation variability and to obtain one single point and interval estimate using the PROC MIANALYZE procedure. The below sample SAS code implements the pooling across imputations:

```
proc mianalyze data = <input-dataset>;
  by trt01pn avisitn;
  modeleffects estim;
  stderr sem;
  ods output parameterestimates = parmest2(keep = parm estimate stderr
lclmean uclmean probt);
run;
```

where *estim* and *sem* are the point estimate and SE from the MMRM analysis for each imputed dataset whereas *estimate*, *stderr*, *lclmean*, *uclmean*, and *probt* are the final pooled point estimate, SE, lower and upper confidence limits, and the p-value for the between group comparison. In the above code, the BY statement ensures that the pooling is done across imputations within each level of TRT01PN and AVISITN (these variable names might change depending on the actual variable used in the analysis).

Least squares (LS) means alongside their SE as well as 95% confidence intervals (CIs) for each treatment group and timepoint will be displayed; in addition, differences in LS means between each dose group and 1 mg COMP360 dose with their SE and 95% CIs as well as p-values will also be provided.

The above analysis will also be repeated for the PP Analysis Set (*Table 14.2.1.2*).

The analysis tables for both the FAS and PP Analysis Set will also report descriptive summaries obtained using only MADRS data up until the start of new treatment for depression after baseline.

All observed MADRS data will also be listed (*Listing 16.2.6.1*).

LS means of change from baseline in total MADRS score at each post-baseline timepoint as estimated from this primary analysis will be graphically displayed with their 95% CI for each treatment group using both the FAS (*Figure 14.2.1.1*) and the PP Analysis Set (*Figure 14.2.1.2*). Differences (and 95% CIs) in LS means between COMP360 25 or 10 mg and the 1 mg dose over time will also be graphically displayed for both the FAS (*Figure 14.2.1.3*) and the PP Analysis Set (*Figure 14.2.1.4*).

Differences between treatment groups in LS means (and 95% CIs) at week 3 and 12 for both treatment groups will also be displayed as a forest plot, including results for both the primary and the secondary estimand and main (using both FAS and PP Analysis Set)/sensitivity/supplementary analyses, organised as follows (*Figure 14.2.2.5*):

- One panel for each treatment (COMP360 25 mg and COMP360 10 mg, both vs COMP360 1mg)
- Primary and secondary estimand analyses results arranged from top to bottom (ordered as main/sensitivity/supplementary, where applicable)
- Point estimates and CIs added as annotation text

A descriptive summary of percentage change from baseline will also be created for the FAS (*Table 14.2.1.3*) by treatment group and graphically displayed as well (*Figure 14.2.1.5*), including only data prior to the start of a new treatment for depression after baseline.

11.1.1.2 Sensitivity Analysis

Should the primary analysis display a significant p-value for either of the treatment comparisons at week 3 or week 12, a further sensitivity analysis will be performed by assuming systematically worse results than what's been imputed in the primary analysis (Section 11.1.1.1). For this purpose, a delta-adjustment approach will be followed, whereby a systematic shift will be added to all values of the total MADRS score from the two experimental arms imputed according to the rules in Section 11.1.1.1, whereas no shift will

be added to the COMP360 1 mg arm, to estimate the treatment effect in a conservative scenario. The choice of these delta-values will be based on clinical/practical judgement to reflect plausible differences that could have occurred: if a tipping-point, that is a shift that leads to a change in the analysis results such as a non-significant p-value, is identified, clinical judgement will be further used to ascertain the likelihood of the relevant shift to have happened in practice. The larger the shift, the more robust the results from the primary analysis will be.

This delta-adjustment will be applied for all data-points that were imputed for the primary analysis apart from non-monotone missing data, ie for the following cases:

Baseline	**Day 2**	**Week 1**	**Week 3**	**Week 6**	**Week 9**	**Week 12**
X	X	X	X	X	X	X
O	X	X	X	X	X	X
	O	X	X	X	X	X
		O	X	X	X	X
			O	X	X	X
				O	X	X
					O	X

X: missing data point/new concomitant treatment for depression;
O: non-missing data point/no new concomitant treatment for depression.
Timepoints prior to the latest non-missing data point or the latest data point with no new concomitant treatment for depression can either be missing or non-missing.

The delta-adjustment will be implemented using the following procedure:

1) Select the shift value
2) All values, apart from non-monotone missing ones, imputed for the primary analysis are 'adjusted' by adding a constant shift, consistent between the 25 mg and 10 mg treatment arms
3) Step 2) is repeated for all relevant shift values

As an example, below we've taken the example data presented in Section 11.1.1.1 to illustrate the MNAR imputation and applied a shift of 2 to all imputed values:

Visit	**Observed**	**Imputation 1 (c = 0.1017)**		
	MADRS Total Score	**Original Imputed Values**	**Rounded Values**	**Shifted Values**
Baseline	35	35	35	35
Day 2	32	32	32	32
Week 1	35	35	35	35

Week 3	36	36	36	36
Week 6	34***	39.66	40	42
Week 9	15***	41.68	42	44
Week 12	10***	42.73	43	45

*** Denotes observations obtained after the start of new treatment for depression.

The shifts (or delta) values will be specified as identical for both treatment arms since it is reasonable to assume that if a worsening was to happen this would not differ between treatment groups. The following positive shifts for the total MADRS score may be considered: 2, 4, 6, 8 and 10. For each of these shift values, the same analytical approach as the primary efficacy (see Section 11.1.1.1) will be followed, ie a MMRM model will be fit by imputation and results pooled using Rubin's rules. Pooled estimates for the difference between COMP360 25 mg/10 mg and COMP360 1 mg in the LS means for the change from baseline to week 3 and 12 in the total MADRS score will be presented alongside p-values for each set of shift values (*Table 14.2.1.4*).

Results for the tipping-point analysis for both week 3 and 12 will also be displayed graphically (*Figure 14.2.1.6*), with the shift displayed along the x-axis and the associated p-values on the y-axis for each treatment group (COMP360 25 mg and 10 mg), with a horizontal reference line at 0.05 (the threshold alpha value), with one panel for each timepoint.

11.1.1.3 Supplementary Analysis

In addition to the MMRM model specified for the primary analysis in Section 11.1.1.1, ANCOVA models using the change in total MADRS score from baseline as response will be fit separately for all visits, using data imputed from the primary analysis. The model will include treatment and study site as factors and baseline total MADRS score as continuous covariate. As for the primary analysis LS means for each treatment group at each timepoint and the difference with the COMP360 1 mg arm will be displayed alongside 95% CIs and p-values. A sample SAS code to fit such a model is displayed below:

```
ods output lsmeans = lsm diff = diffs;
proc mixed data = <input-dataset>;
  by avisitn imputation;
  class trt01pn siteid;
  model chg = trt01pn siteid base;
  lsmeans trt01pn / diff cl;
run;
```

Estimates from the above model will be combined as for the primary analysis using Rubin's rules (*Table 14.2.1.5*).

11.1.2 Treatment Policy Estimand (Secondary)

11.1.2.1 Main Analysis

The main analysis of the Treatment Policy estimand will assume a MNAR mechanism for missing data after study discontinuation due to lack of efficacy or AE and a MAR mechanism for all remaining missing data (both non-monotone and following study discontinuation for other reasons) and thus perform an imputation similar to what described in Section 11.1.1.1 for the Hypothetical Strategy estimand, using the same example SAS code for the MAR part (PROC MI with MCMC statement).

The imputed datasets will be analysed using the same MMRM model as for the primary estimand and the same quantities will be pooled across imputations via Rubin's rules and displayed in a table (*Table 14.2.1.6*) alongside summary statistics including all observed MADRS values. In addition, plots for both the change from baseline in MADRS score within each treatment arm as well as differences between treatment arms will be graphically displayed for each treatment arm (*Figure 14.2.1.7 and Figure 14.2.1.8*).

11.1.2.2 Sensitivity Analysis

The first sensitivity analysis will follow the same approach as the sensitivity analysis for the Hypothetical Strategy estimand (see Section 11.1.1.2) and will implement a tipping-point analysis, where a shift will be added to all MNAR/monotone MAR-imputed data in the COMP360 25 and 10 mg arms used for the primary analysis as described in Section 11.2.2.1. Similar quantities will be displayed (*Table 14.2.1.7*, *Figure 14.2.1.9*).

An additional sensitivity analysis will involve a MMRM model with no explicit imputation for any missing data (intermittent or discontinuations). Under a MAR assumption, these models are known to produce unbiased estimates of the treatment effect (see, for example, Panel on Handling Missing Data in Clinical Trials[12]). The same MMRM model as for the primary estimand will be fit and similar quantities displayed (*Table 14.2.1.8),* and LS means over time plotted for all treatment arms *(Figure 14.2.1.10*).

11.1.2.3 Supplementary Analysis

An ANCOVA model using the same model structure and SAS code as detailed in Section 11.1.1.3 will be estimated using the multiple imputed data obtained in the main analysis for this estimand (Section 11.1.2.1). LS means for each treatment group and the COMP360 25 mg and 10 mg differences with COMP360 1 mg will be displayed for the change from baseline to each post-baseline visit for the total MADRS score alongside 95% CIs and p-values (*Table 14.2.1.9*).

11.1.3 Missing data

To better understand the extent of intercurrent events occurrences as well as missing data, both due to study discontinuation and non-monotone, a summary by visit will be provided displaying the counts and percentages of participants that:

1) Have started a new treatment for depression
2) Have missing MADRS data due to study discontinuation

3) Have missing MADRS data for any other reason other than study discontinuation

The above summary (*Table 14.2.1.10*) will be based on the FAS and displayed both by treatment and overall.

Should a visit be completely missing, and no date be available to compare against the early termination date, the planned date of the missed visit will be derived in accordance with the study schedule. If this derived date is after the date of study discontinuation the data point will be considered as missing due to early termination.

11.2 Key Secondary Efficacy

The key secondary endpoints are:

- The proportion of participants with response (defined as a ≥50% improvement in MADRS total score from baseline) at week 3.
- The proportion of participants with remission (defined a MADRS total score ≤10) at week 3.
- The proportion of participants who have a sustained response at week 12. Sustained response is defined as the proportion of participants fulfilling response criteria at any visit up to and including week 3 that also fulfils response criteria at all subsequent visits up to and including week 12. Response is defined as ≥ 50% decrease in MADRS total score from baseline.

11.2.1 Composite Strategy Estimand (Primary)

11.2.1.1 Main Analysis

The main analysis for the key secondary endpoints will be conducted, in accordance with the Composite Strategy estimand definition, by imputing any data point from the start of a new treatment for depression or after study discontinuation due to lack of efficacy or AE as a non-responder or non-remitter. If the date where the new medication starts is the same date of a MADRS assessment, that assessment will still be considered 'evaluable' and the above imputation applied from the following timepoint. Missing MADRS data due to study discontinuation other than lack of efficacy or AE for participants that do not start any new treatment for depression will be imputed using the same MAR approach as described for the main analysis of the Hypothetical Strategy estimand for the primary endpoint (Section 11.1.1.1) and values will then be dichotomised to derive responder/remitters status.

The analysis model for the first two key secondary endpoints (proportion of participants with response at week 3 and proportion of participants with remission at week 3) will be a Generalised Linear Mixed Model (GLMM), with a binomial distribution and a logit link function, and it will include fixed effects for treatment, visit, study site and treatment by visit interaction, and baseline MADRS total score as covariate. Within-participant correlation will be accounted for by using the _RESIDUAL_ option in the RANDOM statement to maintain consistency with the primary efficacy analysis (Section 11.1.1.1)

and similarly an unstructured covariance matrix will be assumed (exploring further options using the same approach as for the MMRM should the model fail to converge). Should convergence problems arise even with a simpler correlation structure, the model will be simplified by removing the study site effect. Sample SAS code to fit the model is displayed below:

```
ods output estimate = <output-dataset>;
proc glimmix data = data;
  by imputation;
  class subjid siteid trt01pn avisitn;
  model aval = trt01pn avisitn trt01pn*avisitn base siteid / dist = binomial link
= logit;
  lsmeans trt01pn*avisitn / diff cl ilink oddsratio;
  random _residual_ / subject = subjid type = un;
  estimate "25 mg COMP360 dose vs 1 mg COMP360 dose at visit 3"
           trt01pn 1 0 -1 trt01pn*avisitn 0 0 1 0 0 0 0 0 0 0 0 0 0 -1 0 0 0 /
cl exp ilink;
  estimate "10 mg COMP360 dose vs 1 mg COMP360 dose at visit 3"
           trt01pn 0 1 -1 trt01pn*avisitn 0 0 0 0 0 0 0 1 0 0 0 0 0 -1 0 0 0 /
cl exp ilink;
run;
```

ESTIMATE statements for all other post-baseline timepoints can be derived similar to those presented above. The odds ratio (OR) of being a responder (remitter) at all post-baseline visits will be presented for the comparison of COMP360 25 mg/10 mg vs COMP360 1 mg, alongside CIs and p-values. Results will then be pooled across imputations using Rubin's rules to present a single treatment effect estimate. The main difference with respect to the pooling across imputation is that the OR is not normally distributed, and as such the log-ORs and their SE will be pooled rather than the OR themselves. These are stored in the estimate and stderr of the ODS output datasets for PROC GLIMMIX. The pooled estimates and the 95% confidence limits can then be exponentiated to give the summary measures on the OR scale (whereas the p-value can be provided as obtained directly from the combination process).

This analysis will be performed for the FAS (*Table 14.2.2.1.1* for response at all post-baseline visits and *Table 14.2.2.2.1* for remission at all post-baseline visits) and for the PP Analysis Set (*Table 14.2.2.1.2* for response at all post-baseline visits and *Table 14.2.2.2.2* for remission at all post-baseline visits). ORs over time for the comparison of COMP360 25mg/10mg vs 1 mg will also be graphically displayed with their 95% CIs for the FAS (*Figure 14.2.2.1.1* and *Figure 14.2.2.2.1* for response and remission, respectively) and the PP Analysis Set (*Figure 14.2.2.1.2* and *Figure 14.2.2.2.2* for response and remission, respectively). The proportion of responders and remitters will also be descriptively summarised alongside 95% Wald-type asymptotic CIs for all timepoints up to week 12, only including records up to the start of new treatment for depression after baseline, and graphically displayed via line plots including the estimated 95% CI for the FAS (*Figure 14.2.2.1.3* and Figure *14.2.2.2.3*, respectively) and the PP Analysis Set (Figure *14.2.2.1.4*

and Figure *14.2.2.2.4*, respectively). Sample SAS code for the estimation of 95% Wald-type CIs for proportions is displayed below:

```
proc freq data = <input-dataset>;
  by trt01pn avisitn;
  tables aval / binomial (level = 2 cl = wald);
run;
```

The third (and last) key secondary endpoint (sustained response at week 12) main analysis will involve a standard logistic regression model using site and treatment as main effects and baseline MADRS total score as a continuous covariate. Every participant that started a new treatment for depression or discontinued due to lack of efficacy or AE will be regarded as a non-responder under both the standard and relaxed definition and thus analysed as such, whereas for participants that discontinued the study for any other reason without ever starting new treatment for depression the dichotomised MAR imputed data (see Section 11.1.1.1) as obtained for the other two key secondary endpoints will be used to derive sustained responder status.

Similar to the other key secondary endpoints, the OR of having sustained response (vs not having it) will be displayed for the relevant pairwise treatment comparisons alongside 95% CIs and p-values. Sample SAS code for the logistic regression is as follows:

```
ods output lsmeans = <output-dataset>;
proc genmod data = <input-dataset>;
  by imputation;
  class trt01pn (ref = '3');
  model aval = trt01pn base siteid / dist = binomial link = logit;
  lsmeans trt01pn / diff ilink exp cl;
run;
```

Using the REF = '3' option in the CLASS statement has the effect of ensuring that the ORs are then derived using the COMP360 1 mg arm as reference (in the above code it is assumed this corresponds to a TRT01PN value of 3). Results will be pooled across imputations using Rubin's rules and the same method described for the other two key secondary endpoints. In addition, the proportion of sustained responders will also be descriptively summarised for each treatment group alongside Wald-type asymptotic 95% CIs, using similar SAS PROC FREQ code to what was presented earlier in this section, only including participants that do not start a new treatment for depression after baseline. The above analysis will be conducted for the FAS (*Table 14.2.2.3.1.1*) and the PP Analysis Set (*Table 14.2.2.3.2*), jointly displaying the ORs and the proportions (with their 95% CI).

The analysis for sustained responders will also be performed using the relaxed definition of sustained responders for the FAS only (*Table 14.2.2.3.1.2*), using the same methods as for the main definition.

The derived response/remission/sustained response status will be listed as 'Yes'/'No' values in the same listing as the MADRS total score data (*Listing 16.2.6.1*), with sustained response being available only for week 12.

Similar to the primary endpoint (see Section 11.1.1.1) a forest plot including all ORs and their 95% CIs at week 3 and 12 for the response/remission endpoints and week 12 for the sustained response endpoint (only the main definition) for both estimands and all analysis methods will be displayed using the same logic (*Figures 14.2.2.6.1, 14.2.2.6.2 and 14.2.2.6.3*).

11.2.1.2 Sensitivity Analysis

No sensitivity analysis will be performed.

11.2.1.3 Supplementary Analysis

A supplementary analysis will be performed for the first two key secondary endpoints, using a logistic regression model fit separately for each visit using the same imputed data considered for the main analysis of this estimand (ie non-responder/remitter imputed) and covariates and code similar to the one for the sustained response endpoint discussed in Section 11.2.1.1 with the addition of a BY statement to allow for separate model fitting by visit and imputation.

OR and their 95% CI and associated p-values for all post-baseline timepoints will be displayed using the FAS only (*Table 14.2.2.1.3* and *Table 14.2.2.2.3*, respectively for the two endpoints).

11.2.2 Treatment Policy Estimand (Secondary)

11.2.2.1 Main Analysis

The main analysis for the key secondary endpoints will be conducted, in accordance with the Treatment Policy estimand definition, by imputing any data point after study discontinuation due to lack of efficacy or AE as a non-responder or non-remitter or non-sustained responder. Missing MADRS data due to study discontinuation other than lack of efficacy or AE will be imputed using the same MAR approach as described for the main analysis of the Treatment Policy estimand for the primary endpoint (Section 11.1.2.1) and values will then be dichotomised to derive responder/remitter/sustained responder status.

The datasets will be analysed using either the GLMM (responders/remitters) or logistic regression (sustained responder, both definitions) analysis model for each imputation, and results will then be combined using Rubin's rules, using a similar approach as described in Section 11.2.1.1 .

OR and their 95% CIs and associated p-values for all post-baseline timepoints will be displayed using the FAS only (*Table 14.2.2.1.4*, *Table 14.2.2.2.4* and *Table 14.2.2.3.3* respectively for the three endpoints). Proportion of responders/remitters/sustained

responders will also be displayed, using all observed MADRS values irrespective of any new treatment for depression started after baseline.

11.2.2.2 Sensitivity Analysis

The responder and remitter endpoints will be analysed using the same analysis model described in Section 11.2.1.1 (GLMM), but with no explicit imputation, so that a standard GLMM regression model will be fit to derive OR and their 95% CIs (*Table 14.2.2.1.5* and *Table 14.2.2.2.5* respectively for the two endpoints).

11.2.2.3 Supplementary Analysis

The same multiple imputed dataset as for the main analysis will be used for the supplementary analysis, where the responder and remitter endpoints will be analysed with a by-visit logistic regression model using the same covariates and SAS code as described in Section 11.2.1.1 and displaying similar quantities after pooling results across imputations using the approach described in Section 11.2.2.1 (*Table 14.2.2.1.6* and *Table 14.2.2.2.6*).

The sustained responder endpoint will be analysed using a logistic regression based on all observed data and displaying associated OR with 95% CIs and p-values (*Table 12.2.2.3.4*).

11.3 Other Secondary Endpoints

Time to event measures will be analysed for each treatment group for the following parameters:

- Time to the first of any of the following depressive events from baseline to week 12:
  - Initiation of any antidepressive treatment (first new antidepressant treatment in time period only if started for worsening of symptoms or lack of benefit from study treatment and not when initiated to prevent relapses and maintain a positive response)
  - Hospitalisation due to depression or suicidality
  - Suicide attempt, prevention of an imminent suicide attempt/interrupted attempt/aborted attempt, or completed suicide
  - Active suicidal ideation with specific plan according to the C-SSRS (a score of 'yes' on question 4 or 5)
  - Worsening in the MADRS clinician rated severity scale: ie a 5-point worsening compared to baseline score at any timepoint post-baseline; or a worsening of ≥10 points (providing the highest score is ≥15) across two or more consecutive visits (in this case the first date of the ≥10-point increase will be classed as the event and at the final study visit the ≥10-point worsening will qualify as an event without need for confirmation at a subsequent visit)

- Time to relapse from baseline to week 12 (relapse is defined as participants who responded (defined as ≥ 50% reduction on the MADRS at week 3), a worsening of ≥10 points on the MADRS (providing the highest score is ≥20) across two or more consecutive visits (in this case the first date of the ≥10 points increase will be classed as the event and at the final study visit the ≥10 points worsening will qualify as an event without need for confirmation at a subsequent visit)

Participants who withdraw from the study or die (for a cause other than suicide) before experiencing the event of interest will be censored on the date of their last study visit or on the date of their death. The event-free survival curves will be estimated via the Kaplan-Meier method.

A sample SAS code to perform the Kaplan-Meier analysis is as follows:

```
proc lifetest data = <input> method = km;
  strata trt01pn / diff = control('0') test = logrank;
  time time*censored(1);
run;
```

In addition, a Cox regression model will be fit to both time to event endpoints including treatment and baseline MADRS scores (for time to relapse only) as covariates. From this model, the hazard ratios (HRs) of COMP360 25mg and 10mg vs COMP360 1mg arm will be estimated alongside its 95% CI (derived using the profile likelihood approach) using the sample SAS code below:

```
proc phreg data = <input>;
  class trt01pn (ref = 'COMP360 1 mg');
  model time*censored(1) = base trt01pn;
  hazardratio trt01pn / cl = pl diff = ref;
run;
```

For each time to event endpoint quartiles of the Kaplan-Meier survival curve alongside their 95% CI will be displayed by treatment group, together with the observed minimum and maximum time to event occurrence (as measured in days). The HRs and their 95% CI will also be presented (*Table 14.2.2.4* and *Table 14.2.2.5*, respectively). Also, Kaplan-Meier curves for each time to event endpoint will be presented *(Figure 14.2.2.3* and *Figure 14.2.2.4*, respectively).

The time to the occurrence of each separate event as analysed above will also be listed (*Listing 16.2.6.8.1* and *Listing 16.2.6.8.2*, respectively), and should a participant be censored by the time of their study completion this censoring time will be displayed instead.

11.4 Exploratory Efficacy

11.4.1 COVID-19

To better evaluate the impact of COVID-19 on the primary endpoint an exploratory analysis will be performed where participants that had been dosed before the announcement of the first UK lockdown and the first site re-opening after it (29th April 2020) will be compared to those that were dosed after this date.

Descriptive summaries by treatment arm will be provided for the MADRS score and its change from baseline, with 95% CIs added if the sample size in each treatment arm is enough to provide interpretable results (*Table 14.2.1.11*).

11.4.2 Additional Responders Analysis

An additional analysis will be performed to better assess the response to treatment using either data observed prior to the start of a new treatment for depression (in line with the Hypothetical Strategy estimand for the primary endpoint) or data as observed, irrespective of the start of a new treatment for depression (and allowing missing data after study discontinuation), as part of the Treatment Policy estimand.

For this analysis the following summaries will be presented:

- Number and frequency of MADRS responders (as defined in Section 9.3.1.2) that are not MADRS remitters (so a post-baseline MADRS score > 10)
- Number and frequency of MADRS partial responders (as defined in Section 9.3.1.5)

Summaries for each treatment group and post-baseline timepoint will be displayed (*Table 14.2.2.1.7.1* for the Composite Strategy estimand and *Table 14.2.2.1.7.2* for the Treatment Policy estimand).

11.4.3 Correlation Analysis

Exploratory analyses will be performed to investigate the correlation of the following variables with the primary endpoint:

- PANAS at baseline (Negative Affect and Positive Affect scores separately)
  5D-ASC on day 1 (all five dimensions separately)
- STAR-C at baseline
- STAR-P at baseline
- EBI at day 2
- C-reactive protein (CRP) at screening

The correlation with the primary endpoint (change from baseline to week 3 and 12 in the total MADRS score) will be analysed using an ANCOVA model with treatment as factor, the measurement of the variable at the timepoint listed above as a covariate and their interaction to evaluate the presence of unequal slopes. Type III mean squares alongside F-

tests and p-values for each model effect will be displayed (*Table 14.2.1.12*). The code for the ANCOVA model described in Section 11.1.1.3 can be adapted as follows for this analysis:

```
ods output type3 = anova;
proc mixed data = <input-dataset> method = type3;
  by avisitn;
  class trt01pn;
  model chg = trt01pn <predictor> trt01pn*<predictor>;
run;
```

For continuous predictors, a scatterplot will also be created, with the relevant predictor variable on the x-axis and the primary endpoint on the y-axis, the points color-coded based on the treatment received by the participant, the treatment-specific regression lines being displayed with the same colours as the symbol markers and an overall Pearson's correlation coefficient, obtained using Fisher's z transformation, added in the plot area (*Figure 14.2.1.11.1* and *Figure 14.2.1.11.2* for week 3 and 12, respectively, one scatterplot per page). The correlation will be derived using the following example SAS code:

```
ods output FisherPearsonCorr = rho;
proc corr data = <input-dataset> fisher;
  var chg;
  with <list-of-predictor-variables>;
run;
```

The variable *correst* from the ODS output dataset will be used in the reporting.

Three more variables will be added as additional pages to the scatterplot but will not be analysed using the ANCOVA model: change from baseline to day 2 in PANAS, change from baseline to day 2 and week 3 in CRP (the PANAS has two dimensions, positive and negative affect, so there will be 4 additional scatterplots added). The ANCOVA models will not be fit for these extra-variables because they were collected after the study drug dose and this could cause difficulties in interpreting model results. An example SAS code for one plot is as follows:

```
proc template;
  define statgraph ScatCurveLegend;
  begingraph;
  entrytitle halign = left "Predictor: <predictor-name>";
  layout overlay / xaxisopts = (label = "<predictor-name>")
                   yaxisopts = (label = "Change from Baseline to Week 3
in Total MADRS Score");
    scatterplot x = <predictor> y = chg / group = trt01pn name = "scatter"
markerattrs = (symbol = CircleFilled);
      regressionplot x = base y = chg / group = trt01pn name =
"regression";
      textplot x = x y = y text = text / textattrs = (size = 12);
    mergedlegend "scatter" "regression" / border = true;
  endlayout;
  endgraph;
  end;
run;
```

```
proc sgrender data = <input-dataset> template = ScatCurveLegend;
run;
```

The above code works on the following assumptions:

- The input dataset has, at a minimum, the *trt01pn*, *chg* (change from baseline in MADRS) and the <*predictor*> variable populated for all records (any missing data will be excluded from the plot). Only change from baseline to week 3 in MADRS is available.
- The additional variables *x*, *y* and *text* will only be populated for one record in the dataset and will contain the coordinates and the text to display Pearson's correlation coefficient in the following form: 'Pearson's rho = 0.xxxx'.

All the above analyses will be performed using the FAS and observed data (ie no imputation).

11.4.4 Exploratory Endpoints

The following exploratory endpoints will be analysed in similar way as the primary efficacy endpoints, ie using a MMRM with no prior data imputation (see Section 11.1.1.1) and displaying estimated LS means alongside their 95% CIs as line plots for each subgroup:

- Participant EQ-5D-3L total score change from baseline to week 3 and week 12 (*Table 14.2.3.1*, *Figure 14.2.3.1*, *Listing 16.2.6.2.1*)
- EQ Visual Analogue Scale change from baseline to week 3 and 12 (*Table 14.2.3.2*, *Figure 14.2.3.2*, *Listing 16.2.6.2.2*)
- SDS total score change from baseline to week 3 and week 12 (*Table 14.2.3.3*, *Figure 14.2.3.3*, *Listing 16.2.6.3*)
- GAD-7 total score change from baseline to week 3 and week 12 (*Table 14.2.3.4*, *Figure 14.2.3.4*, *Listing 16.2.6.5*)
- QIDS-SR-16 total score change from baseline to every study visit (*Table 14.2.3.5*, *Figure 14.2.3.5*, *Listing 16.2.6.6*)
- WSAS total score change from baseline to week 3 and week 12 (*Table 14.2.3.6*, *Figure 14.2.3.6*, *Listing 16.2.6.7*)
- PANAS total score change from baseline to day 2 and week 3 (*Table 14.2.3.7*, *Figure 14.2.3.7*, *Listing 16.2.9.10*)
- DSST total score change from baseline to day 2, week 3, and week 12 (*Table 14.2.3.8*, *Figure 14.2.3.8*, *Listing 16.2.6.4*)

All these endpoints will be analysed for the FAS and listed as well (see item list above for tables and listings numbers). The Caregiver EQ-5D-3L score, on the other hand, will only be listed alongside the participant score since it is not a mandatory assessment and is provisional of optional caregiver consent.

The DSST will also be summarised by categories of correct symbols at baseline and at each post-baseline timepoint as described in Section 9.3.5 (*Table 14.2.3.9*).

The following endpoints will also be descriptively summarised by treatment group and listed:

- EBI total score at day 2 (*Table 14.2.3.10*, *Listing 16.2.9.11*)
- 5D-ASC (each dimension and subscale, see Section 9.3.14) at day 1 (*Table 14.2.3.11*, *Listing 16.2.9.12*)

11.5 Subgroup Analysis: Efficacy

The primary and key secondary endpoints will be descriptively (and, for the primary endpoint only, also graphically via profile plots for the change from baseline in MADRS, with means and 95% CIs) summarised over time by the subgroups below based on the FAS, to explore the heterogeneity of the treatment effect across strata. Subgroups are defined as follows:

- Country (Canada, Czech Republic, Denmark, Germany, Ireland, Netherlands, Portugal, Spain, United Kingdom, United States)
- Gender (Male, Female)
- Age (18-34, 35-64, 65+)
- Number of failed pharmacological treatments for the current episode of depression (level of treatment resistance) (2, 3, 4 treatments)
- Depression severity baseline level (HAM-D-17: moderate (18-23), severe (≥24))

The table and figure (where applicable) number for each endpoint and subgroup is displayed below (when both a table and a figure are needed their numbers are separated by a '/'):

	Change in MADRS	**Response at Week 3**	**Remission at Week 3**	**Sustained Response at Week 12**
Country	*14.2.1.13/14.2.1.12*	*14.2.2.1.8*	*14.2.2.2.7*	*14.2.2.3.5*
Gender	*14.2.1.14/14.2.1.13*	*14.2.2.1.9*	*14.2.2.2.8*	*14.2.2.3.6*
Age	*14.2.1.15/14.2.1.14*	*14.2.2.1.10*	*14.2.2.2.9*	*14.2.2.3.7*
# of Failed Pharmacological Treatments	*14.2.1.16/14.2.1.15*	*14.2.2.1.11*	*14.2.2.2.10*	*14.2.2.3.8*
Depression Severity at Baseline	*14.2.1.17/14.2.1.17*	*14.2.2.1.12*	*14.2.2.2.11*	*14.2.2.3.9*

A specific category within a subgroup will only be plotted in the associated figure if there are at least 10 participants in each treatment arm for that specific subgroup.

All summaries, both in the tables and the figures, will follow the Hypothetical Strategy estimand, and will only include data collected prior to the start of any new treatment for depression after baseline.

12 SAFETY

The Safety Analysis Set will be used for all safety-related evaluations, unless otherwise specified. The summaries and analyses will be presented for each treatment group and overall.

12.1 Adverse Events (AEs)

An overall summary of AEs, including total number of TEAEs and treatment-emergent serious adverse events (TESAEs), ADRs and SARs, severe TEAEs, AESIs, TEAEs and TESAEs and ADRs leading to study discontinuation, and TEAEs leading to death will be presented (*Table 14.3.1.1*).

The following summary tables will be summarised by SOC and PT and sorted by alphabetical order for SOC and by total decreasing frequency for PT, including both the number of participants with a given event and the number of events.

- Summary of TEAEs (*Table 14.3.1.2*)
- Summary of non-serious TEAEs ≥ 5% (PT for any treatment group ≥ 5%) (*Table 14.3.1.3*)
- Summary of non-Serious TEAEs by worst severity (mild/moderate/severe) (*Table 14.3.1.4*)
- Summary of TEAEs by strongest relationship to study drug (related/not related) (*Table 14.3.1.5*)
- Summary of TEAEs by time of onset (see Section 9.4.1 for categories) (*Table 14.3.1.6*)
- Summary of TEAEs by duration (see Section 9.4.1 for categories) (*Table 14.3.1.7*)
- Summary of TEAEs by time of onset (see Section 9.4.1 for categories) and duration (see Section 9.4.1 for categories) (*Table 14.3.1.8*)
- Summary of ADRs by time of onset (see Section 9.4.1 for categories) and duration (see Section 9.4.1 for categories) (*Table 14.3.1.9*)
- Summary of duration of TEAEs (*Table 14.3.1.10*)
- Summary of ADRs (*Table 14.3.1.11*)
- Summary of TESAEs (*Table 14.3.1.12*)
- Summary of SARs (*Table 14.3.1.13*)
- Summary of AESIs (*Table 14.3.1.14*)
- Summary of TEAEs leading to study withdrawal (*Table 14.3.1.15*)
- Summary of ADRs leading to study withdrawal (*Table 14.3.1.16*)
- Summary of COVID-19 related TEAEs (*Table 14.3.1.17*)

For each of the summaries done at the participant level, multiple occurrences of the same event within a participant will be counted once in the summaries by SOC and PT; multiple occurrences of the same event within a participant will be counted once in the maximum intensity category (severe > moderate > mild) and/or maximum study drug relationship category (related > not related). If intensity or relationship is found to be missing, the most severe occurrence will be imputed for that particular summary.

These summaries will also present the number of events that occurred: summaries by worst severity or strongest relationship will only report the events whose category equals the maximum category for a given participant. As an example, should a participant have 2 mild and 3 severe AEs, only these latter will be reported in the 'Severe' line of the table for the overall and appropriate SOC/PT summaries, whereas the 2 mild events will not be considered; similarly, if a participant has only mild events all of them will be reported in the appropriate 'Mild' lines of the display.

AEs whose end date is not available because the data is genuinely missing will be considered as ongoing for analysis purposes.

All AEs for each participant, including multiple occurrences of the same event, will be presented in full in a comprehensive listing including participant number, treatment, reported term, SOC, PT, date/time and study day when AE starts/stops, duration (days), relationship to treatment, severity, action taken, outcome, seriousness, whether concomitant medication started due to TEAE and whether the TEAE leads to study discontinuation (*Listing 16.2.7.1*). Listings will also be provided for participants who experience TESAEs (*Listing 16.2.7.2*), ADRs (*Listing 16.2.7.3*), AESIs (*Listing 16.2.7.4*), lead to death (*Listing 16.2.7.5*) and are associated to COVID-19 (*Listing 16.2.7.6*).

12.2 Laboratory Tests

The full list of parameters and their units (where applicable) are reported in the table below:

Category	**Parameter (Unit)**
Haematology	Basophils (10^9/L) Basophils/Leukocytes (%) Eosinophils (10^9/L) Eosinophils/Leukocytes (%) Mean Corpuscular Haemoglobin Concentration (g/L) Mean Corpuscular Haemoglobin (pg) Mean Corpuscular Volume (fL) Erythrocytes (10^{12}/L)) Haematocrit (%) Haemoglobin (g/L) Leukocytes (10^9/L) Lymphocytes (10^9/L) Lymphocytes /Leukocytes (%) Monocytes (10^9/L) Monocytes /Leukocytes (%)

Category	**Parameter (Unit)**
	Neutrophils (10^9/L) Neutrophils/Leukocytes (%) Platelets (10^9/L)
Chemistry	Alanine Aminotransferase (ALT) (U/L) Albumin (g/L) Alkaline Phosphatase (ALP) (U/L) Amylase (U/L) Aspartate Aminotransferase (AST) (U/L) Bicarbonate (mmol/L) Bilirubin (direct, indirect, and total) (mg/dL) Calcium (mg/dL) Chloride (mmol/L) Creatine Kinase (U/L) Creatinine (mg/dL) Gamma Glutamil Transferase (U/L) Glucose (mg/dL) Lactate Dehydrogenase (U/L) Lipase (U/L) Magnesium (mG/dL) Phosphate (mg/dL) Potassium (mmol/L) Protein-total (g/L) Sodium (mmol/L) Urate (mg/dL) Urea Nitrogen (mg/dL)
Urinalysis	Bilirubin Glucose Ketones Leukocytes Nitrite Occult Blood Protein Specific Gravity Urobilinogen pH

Descriptive statistics of the observed values and change from baseline (continuous data) will be presented for each haematology (*Table 14.3.4.2*) and chemistry (*Table 14.3.4.3*) parameter by visit. Shift tables will display numbers of participants with normal range shifts at baseline versus post-treatment day 2 and week 3 (*Table 14.3.4.4* and *Table 14.3.4.5* for haematology and chemistry, respectively).

Listings for haematology, chemistry and urinalysis data (*Listing 16.2.8.1*, *Listing 16.2.8.2* and *Listing 16.2.8.3*) will also flag values that are outside normal reference ranges or markedly abnormal findings. Laboratory abnormalities will also be listed separately (*Listing 16.2.8.4*).

12.3 Vital Signs

The full list of parameters and their units (where applicable) are reported below:

- Systolic blood pressure (mmHg)
- Diastolic blood pressure (mmHg)
- Pulse rate (bpm)
- Respiration Rate (breaths / min)
- Body temperature (°C)

Descriptive statistics of the observed values and change from baseline will be presented for each parameter by visit (*Table 14.3.5.1*). The number and percentage of participants with values outside clinically important limits, as described in Section 9.4.3 will be summarised by visit (*Table 14.3.5.2*).

All vital signs data will be listed, which will also include flagged values that are outside normal reference ranges or markedly abnormal findings (*Listing 16.2.8.8*). Clinically important vital signs values will also be listed separately (*Listing 16.2.8.9*).

12.4 Electrocardiogram Data

The full list of parameters and their units (where applicable) are reported below:

- Heart rate (bpm)
- PR interval (msec)
- QRS interval (msec)
- QT interval (msec)
- QTcF interval (msec)
- QTcB interval (msec)

Descriptive statistics of the observed values and change from baseline will be presented for each parameter by visit (*Table 14.3.6.1*). Shift tables in relation to the overall interpretation (Normal, Abnormal Not Clinically Significant (NCS), and Abnormal Clinically Significant (CS)) at baseline versus post-treatment day 2 will be presented (*Table 14.3.6.2*). The number and percentage of participants meeting the intervals as described in Section 9.4.4 will be summarised by visit (*Table 14.3.6.3*) and listed (*Listing 16.2.8.11*).

All ECG data will be listed (*Listing 16.2.8.10*).

12.5 C-SSRS

C-SSRS data will be summarised by visit (*Table 14.3.6.4*) and listed (*Listing 16.2.9.13*). Conversion to the FDA-CASA 2012 code and categories described in Section 9.4.7 will be listed jointly for both the C-SSRS and S-STS (*Listing 16.2.9.15*).

12.6 S-STS

S-STS data will be listed only (including the conversion to the FDA-CASA 2012 code and categories as described in Section 9.4.7) (*Listing 16.2.9.14*).

13 BIOMARKERS

13.1 C-Reactive Protein (CRP)

CRP data and the changes from baseline will be summarised, along with the change from baseline being evaluated with a MMRM analysis (model is similar to primary endpoint analysis model as described in Section 11.1.1.1) using the Safety Analysis Set (*Table 14.3.4.1*). Least Square (LS) means and 95% CIs will be derived from the model for each treatment and visit. LS mean differences for COMP360 (25 mg or 10 mg) versus COMP360 1 mg will also be provided with 95% CIs for each visit. CRP values will also be listed (*Listing 16.2.8.5*).

14 CHANGES FROM THE PROTOCOL-PLANNED ANALYSES

Below is a list of changes from the protocol-planned analyses:

- Section 7 (Analysis Sets): The definition of the Safety Analysis Set was updated to include all randomised participants that received study drug. Accordingly, the FAS definition was updated to include randomised participants that received the drug and had at least one post-dose efficacy assessment, thus incorporating the protocol specified modified Intent-to-Treat (mITT) population definition (which has thus been removed).
- Section 11.2 (Key Secondary Efficacy): The analysis of the key secondary endpoints has been changed from the originally planned Cochran-Mantel-Haenszel (CMH) procedure to either a GLMM (response and remission analysis at week 3) or a logistic regression (sustained response at week 12). This was done to increase statistical power by adjusting for more covariates than a stratified CMH would allow and to account for the assumed missing data mechanism.
- Section 11.5 (Subgroup Analysis: Efficacy): The analysis of the country effect described in Section 11.3.2 of the protocol was simplified to only be a summary of by-country descriptive statistics of the MADRS score over time due to the sparseness of data and the difficulty of interpreting the associated results.

15 DATA NOT PRESENTED

15.1 Mindstrong data

Data collected via the Mindstrong app is not in the scope of this SAP and thus details related to its handling and reporting will not be provided here.

16 REFERENCES

1. SAS Institute Inc., Cary, NC, 27513, USA
2. Carhart-Harris RL, Bolstridge M, Rucker J, et al. Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *Lancet*. 2016; 3(7): 619-627
3. Muller MJ, Himmerich H, Kienzle B, Szegedi A. Differentiating moderate and severe depression using the Montgomery-Asberg depression rating scale (MADRS). *J Affect Disord* 2003 (77): 255-260
4. https://euroqol.org/eq-5d-instruments/eq-5d-3l-about/valuation/
5. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997; 35(11): 1095-1108.
6. Ustun B, Adler La, Rudin C, Faraone SV, Spencer TJ, Berglund P, Gruber MJ, Kessler RC. The World Health Organization Adult Attention-Deficit/Hyperactivity Disorder Self-Report Screening Scale for DSM-5. *JAMA Psychiatry* 2017; 74(5): 520-526
7. Studerus E, Gamma A, Vollenweider FX. Psychometric evaluation of the altered states of consciousness rating scale (OAV). *PLoS One*. 2010;5: e12412.
8. Russell JA, Carroll JM. On the bipolarity of positive and negative affect. *Psychol Bull.* 1999; 125:3–30.
9. Liechti ME, Dolder PC, Schmid Y. Alterations of consciousness and mystical-type experiences after acute LSD in humans. *Psychopharmacology (Berl).* 2017; 234(9): 1499–1510.
10. Sheehan DV, Alphs LD, Mao L, et al. Comparative validation of the SSTS, the ISST Plus, and the C–SSRS for assessing the suicidal thinking and behavior FDA 2012 suicidality categories. *Innov Clin Neurosci* 2014; 11(9-10): 32-46.
11. Rubin, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, Inc.
12. National Research Council (US) Panel on Handling Missing Data in Clinical Trials. The Prevention and Treatment of Missing Data in Clinical Trials. Washington (DC): National Academies Press (US); 2010. PANEL ON HANDLING MISSING DATA IN CLINICAL TRIALS. Available from: https://www.ncbi.nlm.nih.gov/books/NBK209899/

17 PROGRAMMING AND DATA PRESENTATION CONVENTIONS

17.1 Treatment Labelling

COMP360 25 mg	COMP360 10 mg	COMP360 1 mg	Overall

17.2 Visit Labelling

To ensure consistency with Clinical Data Interchange Standards Consortium (CDISC) standards, the day of drug dosing, defined as day 0 in the Protocol, will be mapped to day 1. All subsequent visits will then be consistently re-mapped (and henceforth referred to in this SAP), as follows:

Protocol Visit	**Analysis Visit**	**Output Label**
Screening - Visit 1	Screening - Visit 1	Screening
Screening Period - Visit 1a/1b/1c/…	Screening Period - Visit 1a/1b/1c/…	Screening a/ Screening b/Screening c/…
Screening Telephone Visit	Phone Contact Visit	Phone Contact Visit
Baseline (day -1) - Visit 2	Baseline (day -1) - Visit 2	Baseline or Day -1
Day 0 - Visit 3	Day 1 - Visit 3	Baseline or Day 1[a]
Day 1 - Visit 4	Day 2 - Visit 4	Day 2
Week 1/Day 7 - Visit 5	Week 1/Day 8 - Visit 5	Week 1 or Day 8
Week 2/Day 14 - Visit 6	Week 2/Day 15 - Visit 6	Week 2 or Day 15
Week 3/Day 21 - Visit 7	Week 3/ Day 22 - Visit 7	Week 3 or Day 22
Week 4/Day 28 - Visit 7a	**Week 4/Day 29 - Visit 7a**	**Week 4 or Day 29**
Week 5/Day 35 - Visit 7b	**Week 4/Day 36 - Visit 7b**	**Week 5 or Day 36**
Week 6/Day 42 - Visit 8	Week 6/Day 43 - Visit 8	Week 6 or Day 43
Week 7/Day 49 - Visit 8a	**Week 4/Day 50 - Visit 8a**	**Week 7 or Day 50**
Week 8/Day 56 - Visit 8b	**Week 4/Day 57 - Visit 8b**	**Week 8 or Day 57**
Week 9/Day 63 - Visit 9	Week 9/Day 64 - Visit 9	Week 9 or Day 64
Week 10/Day 70 - Visit 9a	**Week 10/Day 71 - Visit 9a**	**Week 10 or Day 71**
Week 11/Day 77 - Visit 9b	**Week 11/Day 78 - Visit 9b**	**Week 11 or Day 78**
Week 12/Day 84 (ET) - Visit 10	Week 12/Day 85 (ET) - Visit 10	Week 12 or Day 85

[a] Vital signs are collected prior to study drug dosing on day 1 and as such their baseline occurs on day 1 rather than on day -1, as is the case for other assessments.

Visits highlighted in bold only occur in specific countries: Germany (Final Protocol version 4.0, 12 November 2019), Portugal (Final Protocol version 4.0, 13 February 2020) and Czech Republic (Final Protocol version 4.3, 26 February 2020). C-SSRS (not for Portugal), concomitant medications and AEs will be collected at these extra visits.

The handling of all assessments occurring out of protocol-specified visit windows or at an unplanned timepoint will be agreed at the blinded data review meeting and documented in the meeting minutes.

17.3 Baseline Definitions

Parameter	**Baseline**
Efficacy	
All	Day -1
Safety	
Clinical laboratory	Screening
Vital Signs	Day 1 Pre-dose
ECGs	Screening
C-SSRS	Day -1
Biomarker	
CRP	Screening

17.4 Study Day, Duration and Time Derivation

Study day, defined as the number of days from the dose of study drug, will be derived as follows:

- date of event – date of dose of study drug + 1, for events on or after dose
- date of event – date of dose of study drug, for events before dose

Duration of an AE (in days) will be calculated using the following formula:

[(stop datetime of the AE-start datetime of the AE)/(60×60×24)]

In cases where the time component of either the start or the stop date is missing the following approach will be taken:

- If the AE start time is missing and the AE occurred after day 1, the duration will be calculated assuming the event at started at 00:00:00 AM of that day, otherwise if the AE occurred on day 1 the duration will be derived using the datetime of dosing as starting point

- If the AE stop time is missing the duration will be calculated assuming the event stopped at 23:59:59 (11:59:59 PM) of that day

For all time to event analyses, participants not reporting the specified endpoint will be censored at the time that the participants were last known not to have experienced the endpoint. The last visit is a censoring event, and events that occur after the last visit will

be disregarded in the statistical analysis. For all endpoints not encompassing death, all deaths will be treated as censoring events. In complex cases where the censoring time of the participant is uncertain, the case will be reviewed by the Worldwide Statistician and a COMPASS statistician, and a censoring time will be assigned before database lock.

17.5 Decimal Places

Decimal places (dps) for derived data described in Section 9 will be determined by the scale of measurement unless otherwise stated. No dps will be displayed if the smallest calculated value is ≥ 100; 1 dp will be displayed when the smallest value is within the interval (10, 100), with 10 being inclusive; 2 dps will be displayed when the smallest value is within (1, 10), with 1 being inclusive; and so on for even smaller scales of measurement.

Derived data where it is known in advance the result will be an integer for example study day, number of days and total scores (for rating scales) will be presented with zero decimal places.

Means, medians, and percentiles will be displayed to one more decimal place than the data, dispersion statistics (eg. SD) will have two more dps, and the minimum and maximum will be displayed to the same number of dps as reported in the raw data. Percentages will be displayed with one dp. No percentage will be displayed for zero-frequency scenarios.

P-values will be quoted to 3 dps. P-values < 0.001 will be presented as <0.001. Where a P-value is less than 0.05 attention will be drawn to this fact using the "+" annotation.

18 TABLES, FIGURES AND LISTINGS

The following tables includes details of the tables, figures and listings to be included within each section of the electronic Common Technical Document (eCTD). The eCTD section is shown in bold. The following validation methods maybe used:

- Independent programming of numbers and manual review of format (IP)
- Independent programming by statistician of numbers and manual review of format (Stat IP)
- Manual review (MR)
- Code review (CR)

Table Number	Table Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.1	**Demographic Data**				
14.1.1	**Disposition**				
14.1.1.1	Participant Disposition by Country - All Screened Participants	IP			X
14.1.1.2	Participant Disposition, Completions and Early Terminations with Reasons – Randomised Analysis Set	IP	X		X
14.1.1.3	Screen Failures with Reasons – All Screened Participants	IP			
14.1.1.4	Reasons for Exclusion from the Per-Protocol Analysis Set – Full Analysis Set	IP			
14.1.1.5	Visits Impacted Due to COVID-19 – Safety Analysis Set	IP			
14.1.1.6	Important Protocol Deviations - Safety Analysis Set	IP			X
14.1.1.7	COVID-19 Related Protocol Deviations – Safety Analysis Set	IP		14.1.1.6	
14.1.2	**Demographics**				
14.1.2	Demographics - Safety Analysis Set	IP	X		X
14.1.3	**Baseline Characteristics**				
14.1.3.1	Medical History – Safety Analysis Set	IP			X
14.1.3.2	Prior Medications – Safety Analysis Set	IP			X
14.1.3.3	Concomitant Medications – Safety Analysis Set	IP		14.1.3.2	X
14.1.3.4	Psychoanaleptic and Psycholeptic-Related Concomitant Medications – Safety Analysis Set	IP	X	14.1.3.2	
14.1.4.1	Mini International Neuropsychiatric Interview (Version 7.0.2) – Safety Analysis Set	IP			X
14.1.4.2	Other Baseline Assessments – Safety Analysis Set	IP			X
14.2	**Efficacy Data**				
14.2.1	**Primary Efficacy Endpoint**				
14.2.1.1	Primary Endpoint – MMRM for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (MNAR + MAR Imputation, Primary Analysis – Hypothetical Strategy Estimand) - Full Analysis Set	Stat IP	X		
14.2.1.2	Primary Endpoint – MMRM for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (MNAR + MAR	Stat IP	X	14.2.1.1	

Table Number	Table Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
	Imputation, Primary Analysis – Hypothetical Strategy Estimand) – Per-Protocol Analysis Set				
14.2.1.3	Primary Endpoint – Summary of Percentage Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score – Full Analysis Set	IP			
14.2.1.4	Primary Endpoint - MMRM for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (Week 3 and 12 Tipping-Point Analysis – Hypothetical Strategy Estimand) - Full Analysis Set	Stat IP	X		
14.2.1.5	Primary Endpoint – ANCOVA for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score by Visit (MNAR + MAR Imputation, Supplementary Analysis – Hypothetical Strategy Estimand) - Full Analysis Set	Stat IP	X		
14.2.1.6	Primary Endpoint - MMRM for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (MNAR + MAR Imputation, Main Analysis – Treatment Policy Estimand) - Full Analysis Set	Stat IP	X	14.2.1.1	
14.2.1.7	Primary Endpoint - MMRM for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (Week 3 and 12 Tipping-Point Analysis – Treatment Policy Estimand) - Full Analysis Set	Stat IP	X	14.2.1.4	
14.2.1.8	Primary Endpoint - MMRM for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (Sensitivity Analysis – Treatment Policy Estimand) - Full Analysis Set	Stat IP	X		
14.2.1.9	Primary Endpoint – ANCOVA for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score by Visit (MNAR + MAR Imputation, Supplementary Analysis – Treatment Policy Estimand) - Full Analysis Set	Stat IP	X	14.2.1.5	
14.2.1.10	Summary of Missing Montgomery-Asberg Depression Rating Scale (MADRS) Total Scores and New Treatment for Depression – Full Analysis Set	IP	X		

Table Number	Table Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.2.1.11	Primary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Total Score by Drug Dosing Time in Relation to COVID-19 Lockdown – Full Analysis Set	IP			
14.2.1.12	Primary Endpoint – ANCOVA for Correlation of Change from Baseline to Week 3 and 12 in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score with Potential Response Predictors – Full Analysis Set	Stat IP			
14.2.1.13	Primary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Total Score and Change from Baseline by Country – Full Analysis Set	IP			
14.2.1.14	Primary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Total Score and Change from Baseline by Gender – Full Analysis Set	IP		14.2.1.13	
14.2.1.15	Primary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Total Score and Change from Baseline by Age Group – Full Analysis Set	IP		14.2.1.13	
14.2.1.16	Primary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Total Score and Change from Baseline by Number of Failed Pharmacological Treatments for the Current Depressive Episode – Full Analysis Set	IP		14.2.1.13	
14.2.1.17	Primary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Total Score and Change from Baseline by Depression Severity at Baseline – Full Analysis Set	IP		14.2.1.13	
14.2.2	**Secondary Efficacy Endpoints**				
14.2.2.1.1	Secondary Endpoint - GLMM for Montgomery-Asberg Depression Rating Scale (MADRS) Responders (Non-Responder + MAR Imputation, Main Analysis – Composite Strategy Estimand) - Full Analysis Set	Stat IP	X		
14.2.2.1.2	Secondary Endpoint - GLMM for Montgomery-Asberg Depression Rating Scale (MADRS) Responders (Non-Responder + MAR Imputation, Main Analysis – Composite Strategy Estimand) – Per-Protocol Analysis Set	Stat IP		14.2.2.1.1	

Table Number	Table Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.2.2.1.3	Secondary Endpoint – Logistic Regression for Montgomery-Asberg Depression Rating Scale (MADRS) Responders by Visit (Non-Responder + MAR Imputation, Supplementary Analysis – Composite Strategy Estimand) - Full Analysis Set	Stat IP			
14.2.2.1.4	Secondary Endpoint - GLMM for Montgomery-Asberg Depression Rating Scale (MADRS) Responders (Non-Responder + MAR Imputation, Main Analysis – Treatment Policy Estimand) - Full Analysis Set	Stat IP		14.2.2.1.1	
14.2.2.1.5	Secondary Endpoint - GLMM for Montgomery-Asberg Depression Rating Scale (MADRS) Responders (Sensitivity Analysis – Treatment Policy Estimand) - Full Analysis Set	Stat IP			
14.2.2.1.6	Secondary Endpoint – Logistic Regression for Montgomery-Asberg Depression Rating Scale (MADRS) Response by Visit (Non-Responder + MAR Imputation, Supplementary Analysis– Treatment Policy Estimand) - Full Analysis Set	Stat IP		14.2.2.1.3	
14.2.2.1.7.1	Secondary Endpoint – Response Analysis – Full Analysis Set	IP			
14.2.2.1.7.2	Secondary Endpoint – Response Analysis (Observed Cases, Treatment Policy Estimand) – Full Analysis Set	IP		14.2.2.1.7.1	
14.2.2.1.8	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Responders up to Week 12 by Country – Full Analysis Set	IP			
14.2.2.1.9	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Responders up to Week 12 by Gender – Full Analysis Set	IP		14.2.2.1.8	
14.2.2.1.10	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Responders up to Week 12 by Age Group – Full Analysis Set	IP		14.2.2.1.8	
14.2.2.1.11	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Responders up to Week 12 by Number of Failed Pharmacological Treatments for the Current Depressive Episode – Full Analysis Set	IP		14.2.2.1.8	

Table Number	Table Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.2.2.1.12	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Responders up to Week 12 by Depression Severity at Baseline – Full Analysis Set	IP		14.2.2.1.8	
14.2.2.2.1	Secondary Endpoint - GLMM for Montgomery-Asberg Depression Rating Scale (MADRS) Remitters (Non-Responder + MAR Imputation, Main Analysis – Composite Strategy Estimand) - Full Analysis Set	Stat IP	X	14.2.2.1.1	
14.2.2.2.2	Secondary Endpoint - GLMM for Montgomery-Asberg Depression Rating Scale (MADRS) Remitters (Non-Responder +MAR Imputation, Main Analysis – Composite Strategy Estimand) – Per-Protocol Analysis Set	Stat IP		14.2.2.1.1	
14.2.2.2.3	Secondary Endpoint – Logistic Regression for Montgomery-Asberg Depression Rating Scale (MADRS) Remitters by Visit (Non-Responder + MAR Imputation, Supplementary Analysis – Composite Strategy Estimand) - Full Analysis Set	Stat IP		14.2.2.1.3	
14.2.2.2.4	Secondary Endpoint - GLMM for Montgomery-Asberg Depression Rating Scale (MADRS) Remitters (Non-Responder + MAR Imputation, Main Analysis – Treatment Policy Estimand) - Full Analysis Set	Stat IP		14.2.2.1.1	
14.2.2.2.5	Secondary Endpoint - GLMM for Montgomery-Asberg Depression Rating Scale (MADRS) Remitters (Sensitivity Analysis – Treatment Policy Estimand) - Full Analysis Set	Stat IP		14.2.2.1.5	
14.2.2.2.6	Secondary Endpoint – Logistic Regression for Montgomery-Asberg Depression Rating Scale (MADRS) Remitters by Visit (Non-Responder + MAR Imputation, Supplementary Analysis– Treatment Policy Estimand) - Full Analysis Set	Stat IP		14.2.2.1.3	
14.2.2.2.7	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Remitters up to Week 12 by Country (Composite Strategy Estimand) – Full Analysis Set	IP		14.2.2.1.8	
14.2.2.2.8	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Remitters up to Week 12 by Gender (Composite Strategy Estimand) – Full Analysis Set	IP		14.2.2.1.8	

Table Number	Table Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.2.2.2.9	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Remitters up to Week 12 by Age Group (Composite Strategy Estimand) – Full Analysis Set	IP		14.2.2.1.8	
14.2.2.2.10	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Remitters up to Week 12 by Number of Failed Pharmacological Treatments for the Current Depressive Episode (Composite Strategy Estimand) – Full Analysis Set	IP		14.2.2.1.8	
14.2.2.2.11	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Remitters up to Week 12 by Depression Severity at Baseline (Composite Strategy Estimand) – Full Analysis Set	IP		14.2.2.1.8	
14.2.2.3.1.1	Secondary Endpoint – Logistic Regression for Montgomery-Asberg Depression Rating Scale (MADRS) Sustained Responders at Week 12 (Non-Responder + MAR Imputation, Main Analysis – Composite Strategy Estimand) - Full Analysis Set	Stat IP	X		
14.2.2.3.1.2	Secondary Endpoint – Logistic Regression for Montgomery-Asberg Depression Rating Scale (MADRS) Sustained Responders (Relaxed Definition) at Week 12 (Non-Responder + MAR Imputation, Main Analysis – Composite Strategy Estimand) - Full Analysis Set	Stat IP	X	14.2.2.3.1.1	
14.2.2.3.2	Secondary Endpoint – Logistic Regression for Montgomery-Asberg Depression Rating Scale (MADRS Sustained Responders at Week 12 (Non-Responder + MAR Imputation, Main Analysis – Composite Strategy Estimand) – Per-Protocol Analysis Set	Stat IP		14.2.2.3.1.1	
14.2.2.3.3	Secondary Endpoint - Logistic Regression for Montgomery-Asberg Depression Rating Scale (MADRS) Sustained Responders at Week 12 (Non-Responder + MAR Imputation, Main Analysis – Treatment Policy Estimand) - Full Analysis Set	Stat IP		14.2.2.3.1.1	
14.2.2.3.4	Secondary Endpoint – Logistic Regression for Montgomery-Asberg Depression Rating Scale (MADRS) Sustained Responders at Week 12 (Observed Cases, Supplementary Analysis – Treatment Policy Estimand) - Full Analysis Set	Stat IP			

Table Number	Table Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.2.2.3.5	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Sustained Responders at Week 12 by Country – Full Analysis Set	IP			
14.2.2.3.6	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Sustained Responders at Week 12 by Gender – Full Analysis Set	IP		14.2.2.3.5	
14.2.2.3.7	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Sustained Responders at Week 12 by Age Group – Full Analysis Set	IP		14.2.2.3.5	
14.2.2.3.8	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Sustained Responders at Week 12 by Number of Failed Pharmacological Treatments for the Current Depressive Episode – Full Analysis Set	IP		14.2.2.3.5	
14.2.2.3.9	Secondary Endpoint – Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Sustained Responders at Week 12 by Depression Severity at Baseline – Full Analysis Set	IP		14.2.2.3.5	
14.2.2.4	Secondary Endpoint - Analysis of Time to Depressive Event (days) - Full Analysis Set	Stat IP			
14.2.2.5	Secondary Endpoint - Analysis of Time to Relapse (days) - Full Analysis Set	Stat IP			
14.2.3	**Exploratory Efficacy Endpoints**				
14.2.3.1	Exploratory Endpoint – MMRM for Change from Baseline in the Participant EQ-5D-3L Total Score - Full Analysis Set	Stat IP			
14.2.3.2	Exploratory Endpoint – MMRM for Change from Baseline in the EuroQoL Visual Analogue Scale (EQ-VAS) – Full Analysis Set	Stat IP		14.2.3.1	
14.2.3.3	Exploratory Endpoint – MMRM for Change from Baseline in SDS Total Score - Full Analysis Set	Stat IP		14.2.3.1	
14.2.3.4	Exploratory Endpoint – MMRM for Change from Baseline in GAD-7 Total Score - Full Analysis Set	Stat IP		14.2.3.1	
14.2.3.5	Exploratory Endpoint – MMRM for Change from Baseline in QIDS-SR-16 Total Score - Full Analysis Set	Stat IP		14.2.3.1	

Table Number	Table Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.2.3.6	Exploratory Endpoint – MMRM for Change from Baseline in WSAS Total Score - Full Analysis Set	Stat IP		14.2.3.1	
14.2.3.7	Exploratory Endpoint – MMRM for Change from Baseline in PANAS Positive and Negative Affect Total Scores - Full Analysis Set	Stat IP			
14.2.3.8	Exploratory Endpoint – MMRM for Change from Baseline in DSST Total Score – Full Analysis Set	Stat IP		14.2.3.1	
14.2.3.9	Exploratory Endpoint – Summary of Categories of DSST Total Score – Full Analysis Set	IP		14.2.3.1	
14.2.3.10	Exploratory Endpoint – Summary of EBI Total Score – Full Analysis Set	IP			
14.2.3.11	Exploratory Endpoint - Summary of 5D-ASC Questionnaire Values – Full Analysis Set	IP			
14.3	**Safety Data**				
14.3.1	**Displays of Adverse Events**				
14.3.1.1	Summary of Treatment-Emergent Adverse Event (TEAEs) – Safety Analysis Set	IP	X		X
14.3.1.2	MedDRA Summary of Treatment-Emergent Adverse Events (TEAEs) by Primary System Organ Class and Preferred Term – Safety Analysis Set	IP	X		X
14.3.1.3	MedDRA Summary of Non-Serious Treatment-Emergent Adverse Events (TEAEs) with ≥ 5% Incidence by Primary System Organ Class and Preferred Term – Safety Analysis Set	IP			
14.3.1.4	MedDRA Summary of Treatment-Emergent Adverse Events (TEAEs) by Primary System Organ Class, Preferred Term and Worst Severity – Safety Analysis Set	IP			X
14.3.1.5	MedDRA Summary of Treatment-Emergent Adverse Events (TEAEs) by Primary System Organ Class, Preferred Term and Strongest Relationship to Study Drug – Safety Analysis Set	IP			X
14.3.1.6	MedDRA Summary of Treatment-Emergent Adverse Events (TEAEs) by Primary System Organ Class, Preferred Term and Time of Onset – Safety Analysis Set	IP			

Table Number	Table Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.3.1.7	MedDRA Summary of Treatment-Emergent Adverse Events (TEAEs) by Primary System Organ Class, Preferred Term and Duration of Adverse Event – Safety Analysis Set	IP			
14.3.1.8	MedDRA Summary of Treatment-Emergent Adverse Events (TEAEs) by Primary System Organ Class, Preferred Term, Time of Onset and Duration of Adverse Event – Safety Analysis Set	IP			
14.3.1.9	MedDRA Summary of Adverse Drug Reactions (ADRs) by Primary System Organ Class, Preferred Term, Time of Onset and Duration of Adverse Drug Reaction – Safety Analysis Set	IP		14.3.1.8	
14.3.1.10	MedDRA Summary of Duration of Treatment-Emergent Adverse Events (TEAEs) by Primary System Organ Class and Preferred Term – Safety Analysis Set	IP			
14.3.1.11	MedDRA Summary of Adverse Drug Reactions (ADRs) by Primary System Organ Class and Preferred Term – Safety Analysis Set	IP		14.3.1.2	
14.3.1.12	MedDRA Summary of Treatment-Emergent Serious Adverse Events (TESAEs) by Primary System Organ Class and Preferred Term – Safety Analysis Set	IP	X	14.3.1.2	X
14.3.1.13	MedDRA Summary of Serious Adverse Reactions (SARs) by Primary System Organ Class and Preferred Term – Safety Analysis Set	IP		14.3.1.2	
14.3.1.14	MedDRA Summary of Treatment-Emergent Adverse Events of Special Interest (AESIs) by AESI Term and Preferred Term – Safety Analysis Set	IP			X
14.3.1.15	MedDRA Summary of Treatment-Emergent Adverse Events (TEAEs) Leading to Study Withdrawal by Primary System Organ Class and Preferred Term – Safety Analysis Set	IP		14.3.1.2	
14.3.1.16	MedDRA Summary of Adverse Drug Reactions (ADRs) Leading to Study Withdrawal by Primary System Organ Class and Preferred Term – Safety Analysis Set	IP		14.3.1.2	
14.3.1.17	MedDRA Summary of COVID-19 Related Treatment-Emergent Adverse Events (TEAEs) by Primary System Organ Class and Preferred Term – Safety Analysis Set	IP		14.3.1.2	
14.3.2	**Listings of Deaths, Other Serious and Significant Adverse Events**				

Table Number	Table Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.3.4	**Laboratory Values**				
14.3.4.1	MMRM for Change from Baseline in C-Reactive Protein (CRP) - Safety Analysis Set	Stat IP		14.2.3.1	
14.3.4.2	Summary Statistics of Observed Values and Change from Baseline in Haematology – Safety Analysis Set	IP			X
14.3.4.3	Summary Statistics of Observed Values and Change from Baseline in Chemistry – Safety Analysis Set	IP		14.3.4.2	X
14.3.4.4	Normal Range Shifts from Baseline in Haematology Values – Safety Analysis Set	IP			X
14.3.4.5	Normal Range Shifts from Baseline in Chemistry Values – Safety Analysis Set	IP		14.3.4.4	X
14.3.5	**Vital Signs**				
14.3.5.1	Summary Statistics of Observed Values and Change from Baseline in Vital Signs by Visit – Safety Analysis Set	IP			X
14.3.5.2	Overall Summary of Clinically Important Vital Signs Values – Safety Analysis Set	IP			
14.3.6	**Other Safety**				
14.3.6.1	Summary Statistics of Observed Values and Changes from Baseline to Day 2 in ECG Variables – Safety Analysis Set	IP		14.3.4.2	X
14.3.6.2	ECG Clinical Interpretation – Shift from Baseline to Day 2– Safety Analysis Set	IP			X
14.3.6.3	Summary of QTc Intervals Categories – Safety Analysis Set	IP			
14.3.6.4	Summary of C-SSRS by Visit and Treatment Group – Safety Analysis Set	IP			X

Figure Number	Figure Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.2	**Efficacy Data**				
14.2.1	**Primary Efficacy Endpoint**				
14.2.1.1	LS Means Profile for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (MNAR + MAR Imputation, Primary Analysis -- Hypothetical Strategy Estimand) – Full Analysis Set	Stat IP	X		
14.2.1.2	LS Means Profile for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (MNAR + MAR Imputation, Primary Analysis -- Hypothetical Strategy Estimand) – Per-Protocol Analysis Set	Stat IP		14.2.1.1	
14.2.1.3	Differences in LS Means for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (MNAR + MAR Imputation, Primary Analysis -- Hypothetical Strategy Estimand) – Full Analysis Set	Stat IP			
14.2.1.4	Differences in LS Means for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (MNAR + MAR Imputation, Primary Analysis -- Hypothetical Strategy Estimand) – Per-Protocol Analysis Set	Stat IP		14.2.1.3	
14.2.1.5	Means Profile for Percentage Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (Hypothetical Strategy Estimand) – Full Analysis Set	IP			
14.2.1.6	Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (Tipping-Point Analysis at Week 3 and 12 - Hypothetical Strategy Estimand) – Full Analysis Set	Stat IP			
14.2.1.7	LS Means Profile for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (MNAR + MAR Imputation, Main Analysis - Treatment Policy Estimand) – Full Analysis Set	Stat IP		14.2.1.1	

Figure Number	Figure Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.2.1.8	Differences in LS Means for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (MNAR + MAR Imputation, Main Analysis, Treatment Policy Estimand) – Full Analysis Set	Stat IP		14.2.1.3	
14.2.1.9	Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (Tipping-Point Analysis at Week 3 and 12 – Treatment Policy Estimand) – Full Analysis Set	Stat IP		14.2.1.6	
14.2.1.10	LS Means Profile for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score (Sensitivity Analysis, Treatment Policy Estimand) – Full Analysis Set	Stat IP		14.2.1.1	
14.2.1.11.1	Scatterplot of Change from Baseline to Week 3 in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score vs Potential Response Predictors – Full Analysis Set	Stat IP			
14.2.1.11.2	Scatterplot of Change from Baseline to Week 12 in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score vs Potential Response Predictors – Full Analysis Set	Stat IP		14.2.1.11.1	
14.2.1.12	Means Profile for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score by Country (Hypothetical Strategy Estimand) – Full Analysis Set	IP			
14.2.1.13	Means Profile for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score by Gender (Hypothetical Strategy Estimand) – Full Analysis Set	IP		14.2.1.13	
14.2.1.14	Means Profile for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score by Age Group (Hypothetical Strategy Estimand) – Full Analysis Set	IP		14.2.1.13	
14.2.1.15	Means Profile for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score by Number of Failed Pharmacological Treatments for the Current Depressive Episode (Hypothetical Strategy Estimand) – Full Analysis Set	IP		14.2.1.13	

Figure Number	Figure Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.2.1.16	Means Profile for Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score by Depression Severity at Baseline (Hypothetical Strategy Estimand) – Full Analysis Set	IP		14.2.1.13	
14.2.2	**Secondary Efficacy Endpoints**				
14.2.2.1.1	Odds ratios of Montgomery-Asberg Depression Rating Scale (MADRS) Responders (Non-Responder + MAR Imputation, Main Analysis – Composite Strategy Estimand) – Full Analysis Set	Stat IP			
14.2.2.1.2	Odds ratios of Montgomery-Asberg Depression Rating Scale (MADRS) Responders (Non-Responder + MAR Imputation, Main Analysis – Composite Strategy Estimand) – Per-Protocol Analysis Set	Stat IP		14.2.2.1.1	
14.2.2.1.3	Proportion of Montgomery-Asberg Depression Rating Scale (MADRS) Responders up to Week 12 (Composite Strategy Estimand) – Full Analysis Set	Stat IP			
14.2.2.1.4	Proportion of Montgomery-Asberg Depression Rating Scale (MADRS) Responders up to Week 12 (Composite Strategy Estimand) – Per-Protocol Analysis Set	Stat IP		14.2.2.1.3	
14.2.2.2.1	Odds ratios of Montgomery-Asberg Depression Rating Scale (MADRS) Remitters (Non-Responder + MAR Imputation, Main Analysis – Composite Strategy Estimand) – Full Analysis Set	Stat IP		14.2.2.1.1	
14.2.2.2.2	Odds ratios of Montgomery-Asberg Depression Rating Scale (MADRS) Remitters (Non-Responder + MAR Imputation, Main Analysis – Composite Strategy Estimand) – Per-Protocol Analysis Set	Stat IP		14.2.2.1.1	
14.2.2.2.3	Proportion of Montgomery-Asberg Depression Rating Scale (MADRS) Remitters up to Week 12 (Composite Strategy Estimand) – Full Analysis Set	Stat IP		14.2.2.1.3	
14.2.2.2.4	Proportion of Montgomery-Asberg Depression Rating Scale (MADRS) Remitters up to Week 12 (Composite Strategy Estimand) – Per-Protocol Analysis Set	Stat IP		14.2.2.1.3	
14.2.2.3	Kaplan-Meier Plot of Time to Occurrence of First Depressive Event - Full Analysis Set	Stat IP			
14.2.2.4	Kaplan-Meier Plot of Time to Relapse - Full Analysis Set	Stat IP		14.2.2.3	

Figure Number	Figure Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
14.2.2.5	Primary Endpoint - Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score - Forest Plot of Results	Stat IP	X		
14.2.2.6.1	Key Secondary Endpoint - Montgomery-Asberg Depression Rating Scale (MADRS) Response - Forest Plot of Results	Stat IP		14.2.2.5	
14.2.2.6.2	Key Secondary Endpoint - Montgomery-Asberg Depression Rating Scale (MADRS) Remission - Forest Plot of Results	Stat IP		14.2.2.5	
14.2.2.6.3	Key Secondary Endpoint - Montgomery-Asberg Depression Rating Scale (MADRS) Sustained Response - Forest Plot of Results	Stat IP		14.2.2.5	
14.2.3	**Exploratory Efficacy Endpoints**				
14.2.3.1	LS Means Profile for Change from Baseline up to Week 12 in Participant EQ-5D-3L Total Score – Full Analysis Set	Stat IP			
14.2.3.2	LS Means Profile for Change from Baseline up to Week 12 in EuroQoL - Visual Analogue Scale (EQ-VAS) – Full Analysis Set	Stat IP		14.2.3.1	
14.2.3.3	LS Means Profile for Change from Baseline up to Week 12 in SDS Total Score – Full Analysis Set	Stat IP		14.2.3.1	
14.2.3.4	LS Means Profile for Change from Baseline up to Week 12 in GAD-7 Total Score – Full Analysis Set	Stat IP		14.2.3.1	
14.2.3.5	LS Means Profile for Change from Baseline up to Week 12 in QIDS-SR-16 Total Score – Full Analysis Set	Stat IP		14.2.3.1	
14.2.3.6	LS Means Profile for Change from Baseline up to Week 12 in WSAS Total Score – Full Analysis Set	Stat IP		14.2.3.1	
14.2.3.7	LS Means Profile for Change from Baseline up to Week 12 in PANAS Positive and Negative Affect Total Scores – Full Analysis Set	Stat IP			
14.2.3.8	LS Means Profile for Change from Baseline up to Week 12 in DSST Total Score – Full Analysis Set	Stat IP		14.2.3.1	

Listing Number	Listing Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
16.2	**Participant Data Listings**				
16.2.1	**Discontinued Participants**				
16.2.1.1	Details of Study Discontinuations – Safety Analysis Set	IP	X		X
16.2.1.2	Details of Study Completers – Safety Analysis Set	IP			
16.2.1.3	Visits Impacted Due to COVID-19 – Safety Analysis Set	IP			
16.2.2	**Protocol Deviations**				
16.2.2.1	Protocol Deviations – Safety Analysis Set	IP			X
16.2.2.2	COVID-19 Related Protocol Deviations – Safety Analysis Set	IP		16.2.2.1	
16.2.3	**Participants Excluded from The Efficacy Analyses**				
16.2.3	Analysis Sets – Randomised Analysis Set	IP			
16.2.4	**Demographic Data**				
16.2.4.1	Demographics – Safety Analysis Set	IP			X
16.2.4.2	Medical History – Safety Analysis Set	IP			X
16.2.4.3	Inclusion Criteria – All Screened Participants	IP			
16.2.4.4	Exclusion Criteria – All Screened Participants	IP		16.2.4.3	
16.2.4.5	Prior Medications – Safety Analysis Set	IP			X
16.2.4.6	Concomitant Medications – Safety Analysis Set	IP			X
16.2.4.7	Psychoanaleptic and Psycholeptic-Related Concomitant Medications – Safety Analysis Set	IP	X		
16.2.5	**Compliance And/Or Drug Concentration Data**				
16.2.5	Study Medication Administration – Safety Analysis Set	IP			X
16.2.6	**Individual Efficacy Response Data**				
16.2.6.1	Montgomery-Asberg Depression Rating Scale (MADRS) Score - Full Analysis Set	IP	X		
16.2.6.2.1	EQ-5D-3L Descriptive System - Full Analysis Set	IP			
16.2.6.2.2	EQ Visual Analogue Scale (VAS) - Full Analysis Set	IP			
16.2.6.3	Sheehan Disability Scale (SDS) - Full Analysis Set	IP			
16.2.6.4	The Digit Symbol Substitution Test (DSST) - Full Analysis Set	IP			

Listing Number	Listing Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
16.2.6.5	Generalised Anxiety Disorder Scale (GAD-7) – Full Analysis Set	IP			
16.2.6.6	Quick Inventory of Depressive Symptomatology – Self Report – 16 items (QIDS SR-16) - Full Analysis Set	IP			
16.2.6.7	Work and Social Adjustment Scale (WSAS) - Full Analysis Set	IP			
16.2.6.8.1	Time to Depressive Event – Full Analysis Set	IP			
16.2.6.8.2	Time to Relapse – Full Analysis Set	IP			
16.2.7	**Adverse Event Listings**				
16.2.7.1	Treatment-Emergent Adverse Events (TEAEs) – Safety Analysis Set	IP	X		X
16.2.7.2	Treatment-Emergent Serious Adverse Events (TESAEs) – Safety Analysis Set	IP			X
16.2.7.3	Adverse Drug Reactions (ADRs) – Safety Analysis Set	IP		16.2.7.1	
16.2.7.4	Treatment-Emergent Adverse Events of Special Interest (AESIs) – Safety Analysis Set	IP		16.2.7.1	X
16.2.7.5	Deaths – Safety Analysis Set	IP			X
16.2.7.6	COVID-19 Related Treatment-Emergent Adverse Events (TEAEs) – Safety Analysis Set	IP		16.2.7.1	
16.2.8	**Individual Laboratory Measurements and Other Safety**				
16.2.8.1	Haematology – Safety Analysis Set	IP			X
16.2.8.2	Chemistry – Safety Analysis Set	IP		16.2.8.1	X
16.2.8.3	Urinalysis – Safety Analysis Set	IP		16.2.8.1	X
16.2.8.4	Clinically Significant Laboratory Values – Safety Analysis Set	IP			X
16.2.8.5	C-Reactive Protein – Safety Analysis Set	IP		16.2.8.1	
16.2.8.6	Urine Drug Screen – Safety Analysis Set	IP			X
16.2.8.7	Pregnancy Test – Safety Analysis Set	IP			X
16.2.8.8	Vital Signs Data – Safety Analysis Set	IP			X
16.2.8.9	Clinically Important Vital Signs Values – Safety Analysis Set	IP		16.2.8.4	X
16.2.8.10	ECG Data – Safety Analysis Set	IP			X
16.2.8.11	Clinically Important ECG Values – Safety Analysis Set	IP			X
16.2.9	**Other Study Instruments**				
16.2.9.1	Major Depressive Episode – Full Analysis Set	IP			

Listing Number	Listing Title	Validation Method	Topline Results	Shell Number (if repeat)	DSMB
16.2.9.2	Mini International Neuropsychiatric Interview (MINI, Version 7.0.2) - Full Analysis Set	IP			X
16.2.9.3	Hamilton Depression Rating Scale - 17-Item (HAM-D-17) - Full Analysis Set	IP			X
16.2.9.4	The Massachusetts General Hospital Antidepressant Treatment Response Questionnaire (MGH-ATRQ) - Full Analysis Set	IP			X
16.2.9.5	McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD) - Full Analysis Set	IP			X
16.2.9.6	Adult Self-Report Scale (ASRS) - Full Analysis Set	IP			X
16.2.9.7	Scale to Assess Therapeutic Relationships – Patient Version (STAR-P) - Full Analysis Set	IP			
16.2.9.8	Scale to Assess Therapeutic Relationships – Clinician Version (STAR-C) - Full Analysis Set	IP		16.2.9.7	
16.2.9.9	5-HT2A Polymorphism – Full Analysis Set	IP			
16.2.9.10	The Positive and Negative Affect Schedule (PANAS) – Full Analysis Set	IP			
16.2.9.11	Emotional Breakthrough Inventory (EBI) - Full Analysis Set	IP			
16.2.9.12	Five Dimensions Altered States of Consciousness Questionnaire (5D-ASC) – Full Analysis Set	IP			
16.2.9.13	Columbia Suicidal Severity Rating Scale (C-SSRS) – Safety Analysis Set	IP			X
16.2.9.14	Sheehan Suicidality Tracking Scale (S-STS) – Safety Analysis Set	IP			X
16.2.9.15	C-SSRS and S-STS Conversion to FDA-CASA 2012 Codes and Categories – Safety Analysis Set	IP			

19 APPENDIX

19.1 Appendix I – Changes to COMP 001 Protocol

COMP 001 PROTOCOL

COMPARATIVE TABLE

08-March-2021

COMP 001 PROTOCOL Objective and Endpoint wording versus STATISTICAL ANALYSIS PLAN (SAP) and CLINICAL STUDY REPORT (CSR) Objective and Endpoint wording

This document will detail the minor changes between the COMP 001 protocol and the final versions of the SAP and CSR, with regards to the Objective and Endpoint wording. Note the SAP and CSR are not yet finalised during the time of writing this document as Database Lock (DBL) has not occurred. The protocol versions are outlined below.

- Version 4.0 (22 Jul 2019)
- Version 4.1 (USA) (13 May 2020)
- Version 4.0 (Denmark) (14 Aug 2019)
- Version 4.0 (Germany) (12 Nov 2019)
- Version 4.0 (Portugal) (13 Feb 2020)
- Version 4.3 (Czech Republic) (26 Feb 2020)
- Version 4.1 (United Kingdom) (03 Mar 2020)
- Version 4.1 (USA - Emory) (13 May 2020)
- Version 4.2 (USA – Site 212) (05 Nov 20)

General Changes:

- Replacing psilocybin with COMP360 as this is COMPASS Pathways' proprietary formulation name of synthetic psilocybin.
- Replacing day 1 with day 2 as this complies with the Clinical Data Interchange Standards Consortium (CDISC) requirements of there not being a day 0 in a clinical study. So, reference to day 0 and day 1 in the protocol is being replaced with day 1 and day 2.

- Replacing capitalised words with non-capitalised words for better sentence flow.
- Replacing patients with participants to follow consistency with COMPASS Pathways' choice of wording with regards to referencing subjects.
- Replacing fulfills with fulfils to comply with English (UK) spelling.

Specific Changes:

- Secondary Endpoints
    - Deleted the wording '(clinical judgement, supported by the QIDS-SR-16)' from the time to event measures endpoint. This was an error in the protocol. Further detail on defining this endpoint will be provided in the SAP.
- Exploratory Objectives
    - Deleted the sentence 'This assessment is not mandatory' from the participant EQ-5D-3L objective and added this sentence to the caregiver EQ-5D-3L objective. This was an error in the protocol.
    - Amended the last exploratory objective to reflect the correlation of the markers mentioned with the primary outcome only and not secondary outcomes. This was an error in the protocol. In addition the '2a receptor polymorphism test' will not be assessed for correlation with the primary outcome as it is now out of the scope of the CSR and will be reported separately.
    - Removed 'and predictor of response durability' from the exploratory objective regarding the change from baseline in the Work and Social Adjustment Scale (WSAS). Although this measure has been found to be a predictor of longer-term symptomatic improvement with certain treatments, it is not known if this applies consistently across treatments or replicates across studies.

Summary

Protocol Section Concerned with Wording	**SAP and CSR Amended Wording**
Section 2 STUDY OBJECTIVES The main purpose of this study is to allow COMPASS to determine the optimal dose of psilocybin, either 10 mg or 25 mg. The intent of the	Section 2 STUDY OBJECTIVES The main purpose of this study is to allow COMPASS to determine the optimal dose of **COMP360**, either 10 mg or 25 mg. The intent of the

Protocol Section Concerned with Wording	**SAP and CSR Amended Wording**
primary efficacy analysis is to demonstrate superiority of at least one optimal therapeutic dose of psilocybin (10 mg or 25 mg) versus the 1 mg psilocybin via the following objectives.	primary efficacy analysis is to demonstrate superiority of at least one optimal therapeutic dose of **COMP360** (10 mg or 25 mg) versus the 1 mg **COMP360** via the following objectives.
Section 2.1 Primary (Primary Objective) The primary objective of this study is to evaluate the efficacy of psilocybin (25 mg or 10 mg) compared to 1 mg, administered under supportive conditions to adult participants with TRD, in improving depressive symptoms, as assessed by the change in the Montgomery-Asberg Depression Rating Scale (MADRS) total score from Baseline. Baseline is defined as the assessment score obtained on Day -1. The primary timepoint is Week 3; this variable will be analysed for the change from Baseline to Day 1, and Weeks 1, 3, 6, 9, and 12.	Section 2.1 Primary (Primary Objective) The primary objective of this study is to evaluate the efficacy of **COMP360** (25 mg or 10 mg) compared to 1 mg, administered under supportive conditions to adult participants with TRD, in improving depressive symptoms, as assessed by the change in the Montgomery-Asberg Depression Rating Scale (MADRS) total score from **b**aseline. Baseline is defined as the assessment score obtained on **d**ay -1. The primary timepoint is **w**eek 3; this variable will be analysed for the change from **b**aseline to **d**ay **2**, and **w**eeks 1, 3, 6, 9, and 12.

Protocol Section Concerned with Wording	SAP and CSR Amended Wording

Section 2.2 Secondary (Secondary Objectives)
The secondary objectives are:
• To assess the efficacy of psilocybin compared to 1 mg psilocybin on:
  • Proportion of participants with response defined as a ≥ 50% decrease in MADRS total score from Baseline to Week 3. This will also be assessed at Day 1 and at Weeks 1, 6, 9, and 12.
  • The proportion of participants who have a sustained response at Week 12. Sustained response is defined as the proportion of patients fulfilling response criteria at any visit up to and including Week 3, that also fulfills response criteria at all subsequent visits up to and including Week 12. Response is defined as ≥ 50% decrease in MADRS total score from Baseline.
• To evaluate the safety and tolerability of psilocybin in participants with TRD based on AEs, changes in vital signs, and suicidal ideation/behaviour (measured using the Columbia-Suicide Severity Rating Scale [C-SSRS]) score at all visits.
	Section 2.2 Secondary (Secondary Objectives)
	The secondary objectives are:
	• To assess the efficacy of **COMP360** compared to 1 mg **COMP360** on:
	  • Proportion of participants with response defined as a ≥ 50% decrease in MADRS total score from **b**aseline to **w**eek 3. This will also be assessed at **d**ay **2** and at **w**eeks 1, 6, 9, and 12.
	  • The proportion of participants who have a sustained response at **w**eek 12. Sustained response is defined as the proportion of **participants** fulfilling response criteria at any visit up to and including **w**eek 3, that also **fulfils** response criteria at all subsequent visits up to and including **w**eek 12. Response is defined as ≥ 50% decrease in MADRS total score from **b**aseline.
	• To evaluate the safety and tolerability of **COMP360** in participants with TRD based on AEs, changes in vital signs, and suicidal ideation/behaviour (measured using the Columbia-Suicide Severity Rating Scale [C-SSRS]) score at all visits.

Section 2.3 Exploratory (Exploratory Objectives)
The exploratory objectives are:
• To evaluate the effects of psilocybin on quality of life and wellbeing, functioning and associated disability, cognitive function, and anxiety compared to 1 mg psilocybin on:
  • Quality of life in participant EuroQoL (EQ) 5 dimension 3 level scale (EQ-5D-3L) score change from Baseline to
	Section 2.3 Exploratory (Exploratory Objectives)
	The exploratory objectives are:
	• To evaluate the effects of **COMP360** on quality of life and wellbeing, functioning and associated disability, cognitive function, and anxiety compared to 1 mg **COMP360** on:
	  • Quality of life in participant EuroQoL (EQ) 5 dimension 3 level scale (EQ-5D-3L) score change from **b**aseline to **w**eek 3. This will also be assessed at **w**eek 12.

Protocol Section Concerned with Wording	SAP and CSR Amended Wording

Week 3. This will also be assessed at Week 12. **This assessment is not mandatory.**
• Quality of life in caregiver EQ-5D-3L score change from Baseline to Week 3. This will also be assessed at Week 12.
• Functioning and associated disability in the Sheehan Disability Scale (SDS) score change from Baseline to Week 3. This will be also assessed at Week 12.
• Cognitive function as measured by the Digit Symbol Substitution Test (DSST) score change from Baseline to Week 3. This will also be assessed at Day 1 and Week 12.
• Level of anxiety as measured using the change in Generalised Anxiety Disorder 7 item Scale (GAD-7) total score change from Baseline to Week 3. This will also be assessed at Week 12.
• Participant determined level of depression as measured using the change in Quick Inventory of Depressive Symptomatology Self-Rated (QIDS-SR-16) total score from Baseline to Week 3. This will also be assessed at Screening, Day 1, and Weeks 1, 2, 6, 9, and 12.
• Psychosocial functioning **and predictor of response durability** as measured using the change in Work and Social Adjustment Scale (WSAS) from Baseline to Week 3. This will also be assessed at Week 12.
• To evaluate the impact of different psilocybin doses on real life functional activity estimated from passive data streams collected on a mobile app on participants' mobile phones. The data collected from the participant's phone will include:
  • Number of and time of phone calls/e-mails/texts (content will not be collected)
  • Gestures used (taps, swipes, other)
	• Quality of life in caregiver EQ-5D-3L score change from **b**aseline to **w**eek 3. This will also be assessed at **w**eek 12. **This assessment is not mandatory.**
	• Functioning and associated disability in the Sheehan Disability Scale (SDS) score change from **b**aseline to **w**eek 3. This will be also assessed at **w**eek 12.
	• Cognitive function as measured by the Digit Symbol Substitution Test (DSST) score change from **b**aseline to **w**eek 3. This will also be assessed at **d**ay **2** and **w**eek 12.
	• Level of anxiety as measured using the change in Generalised Anxiety Disorder 7 item **s**cale (GAD-7) total score change from **b**aseline to **w**eek 3. This will also be assessed at **w**eek 12.
	• Participant determined level of depression as measured using the change in Quick Inventory of Depressive Symptomatology Self-Rated (QIDS-SR-16) total score from **b**aseline to **w**eek 3. This will also be assessed at **s**creening, **d**ay **2**, and **w**eeks 1, 2, 6, 9, and 12.
	• Psychosocial functioning as measured using the change in Work and Social Adjustment Scale (WSAS) from **b**aseline to **w**eek 3. This will also be assessed at **w**eek 12.
	• To evaluate the impact of different **COMP360** doses on real life functional activity estimated from passive data streams collected on a mobile app on participants' mobile phones. The data collected from the participant's phone will include:
	  • Number of and time of phone calls/e-mails/texts (content will not be collected)
	  • Gestures used (taps, swipes, other)
	  • Gyroscope (orientation) of the phone (the way the phone is pointing)

Protocol Section Concerned with Wording	**SAP and CSR Amended Wording**
• Gyroscope (orientation) of the phone (the way the phone is pointing) • Acceleration of the phone (sudden movements of the phone) • Keystroke patterns with characters redacted • Location information from the GPS • The app also maintains a histogram of daily words that the participant types on their phone. These words will be stripped from their context and syntax, thus preventing the content of any particular message from being deciphered. • Positive and Negative Affect Schedule (PANAS), Five Dimension Altered States of Consciousness Questionnaire (5D-ASC)**, 2a receptor polymorphism test** and Scale to Assess Therapeutic Relationship (Clinician and Patient version, STAR-C and STAR-P, respectively) will be assessed for correlation with the primary **and secondary** outcome**s** as possible predictors of response.	• Acceleration of the phone (sudden movements of the phone) • Keystroke patterns with characters redacted • Location information from the GPS • The app also maintains a histogram of daily words that the participant types on their phone. These words will be stripped from their context and syntax, thus preventing the content of any particular message from being deciphered. • Positive and Negative Affect Schedule (PANAS), Five Dimension Altered States of Consciousness Questionnaire (5D-ASC) and Scale to Assess Therapeutic Relationship (Clinician and Patient version, STAR-C and STAR-P, respectively) will be assessed for correlation with the primary outcome as possible predictors of response.
Section 3.1 Primary (Primary Endpoint) The primary endpoint is the change in MADRS total score from Baseline (Day -1) to 3 weeks post psilocybin.	Section 3.1 Primary (Primary Endpoint) The primary endpoint is the change in MADRS total score from **b**aseline (**d**ay -1) to 3 weeks post **COMP360**.
Section 3.2 Secondary (Secondary Endpoints) The secondary endpoints are: • The proportion of participants with a response (defined as a ≥ 50% improvement in MADRS total score from Baseline) at Week 3 after the psilocybin session. • The proportion of participants with remission (defined as a MADRS total score ≤ 10) at Week 3 post psilocybin.	Section 3.2 Secondary (Secondary Endpoints) The secondary endpoints are: • The proportion of participants with a response (defined as a ≥ 50% improvement in MADRS total score from baseline) at **w**eek 3 after the COMP360 session. • The proportion of participants with remission (defined as a MADRS total score ≤ 10) at **w**eek 3 post COMP360.

Protocol Section Concerned with Wording	**SAP and CSR Amended Wording**
• The proportion of participants who have a sustained response at Week 12. Sustained response is defined as the proportion of patients fulfilling response criteria at any visit up to and including Week 3, that also fulfills response criteria at all subsequent visits up to and including Week 12. Response is defined as ≥ 50% decrease in MADRS total score from Baseline. • Time to event measures: restart antidepressant medication for any reason, restart medication for continuing depressive symptoms, and relapse from a previously recovered state (clinical judgement, supported by the QIDS-SR-16). Participants who withdraw from the study will be censored from the time to event analysis.	• The proportion of participants who have a sustained response at **w**eek 12. Sustained response is defined as the proportion of participants fulfilling response criteria at any visit up to and including **w**eek 3, that also fulfils response criteria at all subsequent visits up to and including **w**eek 12. Response is defined as ≥ 50% decrease in MADRS total score from **b**aseline. • Time to event measures: restart antidepressant medication for any reason, restart medication for continuing depressive symptoms, and relapse from a previously recovered state. Participants who withdraw from the study will be censored from the time to event analysis.

Section 3.3 Exploratory (Exploratory Endpoints)
The exploratory endpoints are:
• Change from Baseline in the following:
  • Participant EQ-5D-3L at Week 3
  • Caregiver EQ-5D-3L at Week 3 (this assessment is not mandatory)
  • SDS at Week 3
  • DSST at Week 3
  • GAD-7 at Week 3
  • QIDS-SR-16 at Week 3
  • WSAS at Week 3
  • Measures derived from smart phone usage via the Mindstrong app
	Section 3.3 Exploratory (Exploratory Endpoints)
	The exploratory endpoints are:
	• Change from **b**aseline in the following:
	  • Participant EQ-5D-3L at **w**eek 3
	  • Caregiver EQ-5D-3L at **w**eek 3 (this assessment is not mandatory)
	  • SDS at **w**eek 3
	  • DSST at **w**eek 3
	  • GAD-7 at **w**eek 3
	  • QIDS-SR-16 at **w**eek 3
	  • WSAS at **w**eek 3
	  • Measures derived from smart phone usage via the Mindstrong app

19.2 Appendix II – Reference of Section Changes from SAP Version 1

VERSION 2 SAP	**VERSION 1 SAP**
1 INTRODUCTION	1 INTRODUCTION
2 STUDY OBJECTIVES	2 STUDY OBJECTIVES
3 STUDY ENDPOINTS	3 ENDPOINTS
4 ESTIMAND FRAMEWORK	NA
5 STUDY DESIGN	NA
5.1 Study Design	NA
5.2 Study Schematic	NA
5.3 Schedule of Assessments	NA
5.4 Hypotheses and Treatment Comparisons	6.12.1 Primary Endpoint
5.5 Multiplicity	6.12.8 Multiplicity
5.6 Sample Size Considerations	4 SAMPLE SIZE
5.7 Randomisation	5 RANDOMISATION
6 PLANNED ANALYSES	6 PLANNED ANALYSES
6.1 Final Analysis	NA
6.2 Interim Analysis	7 INTERIM ANALYSES
6.3 Data Safety Monitoring Board (DSMB)	8 DATA SAFETY MONITORING BOARD ANALYSIS
7 ANALYSIS SETS	6.1 Analysis Sets
8 GENERAL CONSIDERATIONS	6.4 Conventions 6.2.10 Unscheduled Visits
9 DATA DERIVATIONS	NA
9.1 General	NA
9.1.1 Change from Baseline	6.2.9 Change from Baseline
9.1.2 Missing and Partial Dates	6.2.3 Conventions for Missing and Partial Dates
9.1.3 Treatment Exposure and Compliance	6.2.4 Exposure to Study Drug
9.1.4 Inexact Values	6.2.5 Inexact Values
9.2 Study Participants	NA
9.2.1 Protocol Deviations	6.6 Protocol Deviations
9.2.2 Age Categories	NA
9.2.3 Body Mass Index (BMI)	NA
9.2.4 Length of Current Depressive Episode	NA
9.2.5 Missing Major Depressive Disorder (MDD) Episode Start Dates	NA
9.2.6 Prior and Concomitant Medications	6.9 Prior and Concomitant Medications
9.3 Efficacy Assessments	NA
9.3.1 Montgomery-Asberg Depression Rating Scale Score (MADRS)	6.3.2 Montgomery-Asberg Depression Rating Scale Score
9.3.1.1 Total Score	6.3.2 Montgomery-Asberg Depression Rating Scale Score
9.3.1.2 Responder Status	6.12.4 Analysis of Secondary Efficacy Endpoints
9.3.1.3 Remitter Status	6.12.4 Analysis of Secondary Efficacy Endpoints
9.3.1.4 Sustained Response	6.12.4 Analysis of Secondary Efficacy Endpoints 6.3.1 Sustained Response
9.3.1.5 Partial responder Status	NA
9.3.1.6 Severity Categories	NA
9.3.2 Quick Inventory of Depressive Symptomatology (QIDS-SR-16)	6.3.3 Quick Inventory of Depressive Symptomatology
9.3.3 European Quality of Life 5-dimension 3-level Scale (EQ-5D-3L)	6.3.7 European Quality of Life 5-dimension 3-level Dimension 3-Level Scale
9.3.4 Sheehan Disability Scale (SDS)	6.3.4 Sheehan Disability Scale
9.3.5 Digit Symbol Substitution Test (DSST)	6.3.6 Digit Symbol Substitution Test
9.3.6 Generalised Anxiety Disorder Scale (GAD-7)	6.3.5 Generalized Anxiety Disorder scale

VERSION 2 SAP	**VERSION 1 SAP**
9.3.7 Work and Social Adjustment Scale (WSAS)	6.3.8 Work and Social Adjustment Scale
9.3.8 Mini International Neuropsychiatric Interview	6.3.9 MINI International Neuropsychiatric Interview
9.3.9 Hamilton Depression Rating Scale (HAM-D-17)	6.3.10 Hamilton Depression Rating Scale
9.3.10 The Massachusetts General Hospital Antidepressant Treatment Response Questionnaire (MGH-ATRQ)	6.3.11 The Massachusetts General Hospital Antidepressant Treatment Response Questionnaire
9.3.11 McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD)	6.3.12 McLean Screening Instrument for Borderline Personality Disorder
9.3.12 Adult Self-Report Scale (ASRS)	6.3.13 Adult Self-Report Scale
9.3.13 Scale to Assess Therapeutic Relationships – Patient (STAR-P) and Clinician (STAR-C) Version	6.3.14 Scale to Assess Therapeutic Relationships – Patient and Clinician Version
9.3.14 Five Dimensional Altered States of Consciousness Questionnaire (5D-ASC)	6.3.15 Five Dimension Altered States of Consciousness Questionnaire
9.3.15 Emotional Breakthrough Inventory (EBI)	6.3.16 Emotional Breakthrough Inventory
9.3.16 The Positive and Negative Affect Schedule (PANAS)	NA
9.4 Safety	NA
9.4.1 Adverse Events	6.13.1 Adverse Events
9.4.2 Missing / Partial Start / Stop Dates of AEs and Concomitant Medications	6.2.3.1 Missing/Partial Start/Stop Date of Adverse Events and Concomitant Medications
9.4.3 Vital Signs	NA
9.4.3.1 Vital Signs Ranges of Clinical Importance	NA
9.4.4 ECG Categorical Intervals	6.13.4 Electrocardiogram Data
9.4.5 Columbia-Suicide Severity Rating Scale (C-SSRS)	6.2.6 Columbia-Suicide Severity Rating Scale
9.4.6 Sheehan Suicidality Tracking Scale (S-STS)	6.2.7 Sheehan Suicidality Tracking Scale
9.4.7 Algorithm to Map Sheehan Suicidality Tracking Scale and Columbia-Suicide Severity Rating Scale	6.2.8 Algorithm to Map Sheehan Suicidality Tracking Scale and Columbia-Suicide Severity Rating Scale
10 STUDY PARTICIPANTS	NA
10.1 Disposition of Participants	6.5 Subject Disposition
10.2 Protocol Deviations	6.6 Protocol Deviations
10.3 Demographic and Baseline Characteristics	6.7 Baseline Comparability
10.4 Medical History	6.8 Medical History
10.5 Prior and Concomitant Medications	6.9 Prior and Concomitant Medications
10.6 Treatment Exposure	6.10 Exposure to Study Drug
11 EFFICACY	6.12 Efficacy Analyses
11.1 Primary Efficacy	6.12.2 Primary Efficacy Analyses 6.12.3 Sensitivity Analysis
11.1.1 Hypothetical Strategy Estimand (Primary)	NA
11.1.1.1 Main Analysis	NA
11.1.1.2 Sensitivity Analysis	NA
11.1.1.3 Supplementary Analysis	NA
11.1.2 Treatment Policy Estimand (Secondary)	NA
11.1.2.1 Main Analysis	NA
11.1.2.2 Supplementary Analysis	NA
11.1.3 Missing Data	NA
11.2 Key Secondary Efficacy	6.12.4 Analysis of Secondary Efficacy Endpoints
11.2.1 Composite Strategy Estimand (Primary)	NA
11.2.1.1 Main Analysis	NA
11.2.1.2 Sensitivity Analysis	NA
11.2.1.3 Supplementary Analysis	NA
11.2.2 Treatment Policy Estimand (Secondary)	NA
11.2.2.1 Main Analysis	NA
11.2.2.2 Sensitivity Analysis	NA
11.2.2.3 Supplementary Analysis	NA
11.3 Other Secondary Endpoints	6.12.5 Time to event analysis
11.4 Exploratory Efficacy	NA
11.4.1 COVID-19	NA

VERSION 2 SAP	**VERSION 1 SAP**
11.4.2 Additional Responders Analysis	NA
11.4.3Correlation Analysis	NA
11.4.4 Exploratory Endpoints	6.12.7 Analysis of Exploratory Endpoints
11.5 Subgroup Analysis: Efficacy	6.12.6 Subgroup Analysis
12 SAFETY	6.13 Safety Analyses
12.1 Adverse Events (AEs)	6.13.1 Adverse Events
12.2 Laboratory Tests	6.13.2 Laboratory Data
12.3 Vital Signs	6.13.3 Vital Signs
12.4 Electrocardiogram Data	6.13.4 Electrocardiogram Data
12.5 C-SSRS	NA
12.6 S-STS	NA
13 BIOMARKERS	NA
13.1 C-Reactive Protein (CRP)	NA
14 CHANGES FROM THE PROTOCOL-PLANNED ANALYSES	9 CHANGES TO PLANNED PROTOCOL ANALYSIS
15 DATA NOT PRESENTED	NA
15.1 Mindstrong Data	NA
16 REFERENCES	10 REFERENCES
17 PROGRAMMING AND DATA PRESENTATION CONVENTIONS	NA
17.1 Treatment Labelling	6.4 Conventions
17.2 Visit Labelling	NA
17.3 Baseline Definitions	6.2.1 Baseline
17.4 Study Day, Duration and Time Derivation	6.2.2 Duration/Study Day
17.5 Decimal Places	6.4.1 Decimal Places
18 TABLES, FIGURES AND LISTINGS	11 LIST OF TABLES, FIGURES AND LISTINGS
19 APPENDIX	NA
19.1 Appendix I – Changes to COMP 001 Protocol	NA
19.2 Appendix II – Reference of Section Changes from SAP Version 1	NA
NA	6.2.3.2 Missing Dates of Study Drug Dosing