Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

Statistical Analysis Plan

Protocol Title:

A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Assess the Efficacy and Safety of REL-1017 as Adjunctive Treatment of Major Depressive Disorder

(The RELIANCE-II Study)

Protocol Number:

REL-1017-302
Version 9.0, Amendment 8.0, 11Nov2024

Prepared For:

Relmada Therapeutics, Inc. (Relmada)

Prepared By:

Ulker Aydemir, Ph.D.
Director, Biostatistical Consulting
Syneos Health, Inc. (Syneos Health)

Notice of Confidential and Proprietary Information:

The information contained in this document is confidential belonging to Relmada. Acceptance of this document constitutes agreement by the recipient that no information contained herein will be published or disclosed without prior written authorization from an official of Relmada. However, this document may be disclosed to appropriate Institutional Review Board and Ethics Committees or duly authorized representatives of a national regulatory authority under the condition that they are requested to keep it confidential. In the event of an actual or suspected breach of this obligation, Relmada should be notified promptly.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

Approval Signature Page

CRO Approval

Signed by:
Kerry Hafner
821EB6AF4211477...

11/11/2024

Kerry Hafner, Ph.D.
SVP, Biostatistics
Syneos Health, Inc.

Date

Sponsor Approval

DocuSigned by:
Paul Greene
9C35547D6F0742E...

11/11/2024

Paul Greene, Ph.D.
SVP, Clinical Operations
Relmada Therapeutics, Inc.

Date

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

SAP Revision History

Version Number	Date	Document Owner
1.0	08-Mar-2021	Ben Vaughn, M.S. – Rho, Inc.
2.0	23-May-2022	David Bushnell, M.S. – Cytel, Inc.
3.0	10-Nov-2022	David Bushnell, M.S. – Cytel, Inc.
4.0	08-Nov-2024	Ulker Aydemir, Ph.D. – Syneos Health

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

TABLE OF CONTENTS

Page

APPROVAL SIGNATURE PAGE **2**

SAP REVISION HISTORY **3**

TABLE OF CONTENTS **4**

LIST OF TABLES **7**

1. GLOSSARY OF ABBREVIATIONS AND ACRONYMS **8**

2. PURPOSE **11**

2.1 Responsibilities **11**

2.2 Timing of Analyses **11**

3. STUDY OBJECTIVES **12**

3.1 Primary Efficacy Objective **12**

3.2 Secondary Efficacy Objectives **12**

3.3 Exploratory Objectives **12**

3.4 Safety and Tolerability Objectives **12**

3.5 Pharmacokinetic Objectives **12**

4. STUDY DESIGN **13**

4.1 Brief Description **13**

4.2 Subject Selection **13**

4.3 Inclusion/Exclusion Criteria **14**

4.4 Re-screening **14**

4.5 Study Withdrawal, Removal, and Replacement of Subjects **14**

4.6 Determination of Sample Size **14**

4.7 Treatment Assignment and Blinding **15**

4.8 Study Procedures and Flowchart **15**

5. ENDPOINTS **16**

5.1 Primary Efficacy Endpoint **16**

5.2 Secondary Efficacy Endpoints **16**

5.3 Exploratory Endpoints **17**

5.4 Pharmacokinetic and Pharmacodynamic Endpoints **18**

5.5 Biomarker, Pharmacogenomic (DNA), and Expression (RNA) Evaluations **19**

5.6 Safety Endpoints **19**

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

6. ANALYSIS SETS 20

6.1 Screened / Randomized / Enrolled Set 20

6.2 Safety Set (SS) 20

6.3 Full Analysis Set (FAS) 20

6.4 Per-protocol Set (PPS) 20

6.5 Pharmacokinetic Set (PKS) 20

7. PROTOCOL DEVIATIONS 21

8. ESTIMANDS FOR PRIMARY AND SECONDARY EFFICACY ANALYSES 22

8.1 Main Estimand for Primary Efficacy Analysis 22

8.2 Supportive Efficacy Estimand for Primary Efficacy Analysis 22

8.3 Secondary Efficacy Endpoints and Estimands 23

9. GENERAL ASPECTS FOR STATISTICAL ANALYSIS 25

9.1 General Methods 25

9.2 Missing Data 25

9.3 Visit Windows and Definitions 27

9.4 Programming Considerations and Quality Control 28

10. DEMOGRAPHIC, OTHER BASELINE CHARACTERISTICS AND MEDICATION 29

10.1 Subject Disposition and Withdrawals 29

10.2 Demographic Characteristics 29

10.3 Baseline Disease Characteristics 29

10.4 Baseline Assessments 30

10.5 Medical and Psychiatric History 30

10.6 Prior and Concomitant Medications 30

11. ANALYSIS OF EFFICACY 32

11.1 Primary Analysis of Efficacy 32

11.2 Primary Efficacy Sensitivity and Supplemental Analyses 33

11.3 Analysis of Secondary Efficacy Endpoints 34

11.4 Analyses of Subgroups 35

11.5 Exploratory Efficacy Endpoint Analyses 35

11.6 Pharmacokinetics 36

12. SAFETY ANALYSES 37

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

12.1 Extent of Exposure **37**
12.2 Treatment Compliance **37**
12.3 Adverse Events / Adverse Drug Reactions **37**
12.4 Laboratory Evaluations **38**
12.5 Vital Signs **38**
12.6 ECG **38**
12.7 Physical Examination **39**
12.8 C-SSRS **39**
12.9 CADSS and MADDERS **39**
13. INTERIM ANALYSES **40**
14. REFERENCES **41**
15. APPENDIX – STATISTICAL METHODS DETAILS **42**
15.1 Weighted Test-Statistic and Updated Point and Confidence Estimates **42**
15.2 Multiple Imputation Approach **43**
15.2.1 Example Code for Primary Analysis *44*
15.2.2 Example Code for MNAR (Missing Not at Random) *44*
15.2.3 Example Code for MAR (Missing at Random) *44*
15.3 Tipping Point Analysis **45**

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

LIST OF TABLES

Page

Table 1. Strategies for Handling of Intercurrent Events and Missing Data.....23

Table 2. Estimands for Secondary Endpoints.....24

Table 3. Evaluation Intervals for Efficacy Analyses.....27

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

1. GLOSSARY OF ABBREVIATIONS AND ACRONYMS

Abbreviation	Definition
ADT	Antidepressant therapy
AE	Adverse event
ALT	Alanine aminotransferase
ASEX	Arizona Sexual Experience Scale
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
ATRQ	Antidepressant Treatment Response Questionnaire
BDNF	Brain-derived neurotrophic factor
BMI	Body mass index
BP	Blood pressure (systolic and diastolic)
CADSS	Clinician-Administered Dissociative States Scale
CGI-I	Clinical Global Impression of Improvement
CGI-S	Clinical Global Impression of Severity
CHW	Cui-Hung-Wang
CI	Confidence interval
C_{max}	Maximum observed plasma concentration
CNS	Central nervous system
COWS	Clinical Opiate Withdrawal Scale
COVID-19	Novel coronavirus 2019
CRO	Contract research organization
C_{ss}	Steady-state concentration
C-SSRS	Columbia Suicide Severity Rating Scale
CTCAE	Common terminology for adverse events
CTNI	Clinical Trials Network and Institute
DMC	Data Monitoring Committee
DNA	Deoxyribonucleic acid
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
DSST	Digit Symbol Substitution Test
ECG	Electrocardiogram
ECT	Electroconvulsive therapy
EDC	Electronic data capture
EOS	End of study
EOT	End of treatment
FAS	Full analysis set
FDA	United States Food and Drug Administration
GCP	Good Clinical Practice
HAM-A	Hamilton Anxiety Rating Scale

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

HAMD17	Hamilton Depression Rating Scale
HR	Heart rate
IA	Interim analysis
IB	Investigator's Brochure
ICF	Informed consent form
ICH	International Council for Harmonization
ID	Identification
IEC	Independent Ethics Committee
IRB	Institutional Review Board
IRT	Interactive Response Technology
ITT	Intent-to-treat
LS	Least square means
MAD	Multiple ascending dose
MADDERS	Misuse, Abuse, and Diversion Drug Event Reporting System
MADRS10	Montgomery-Åsberg Depression Rating Scale 10 item-questionnaire
MAR	Missing at Random
MCAR	Missing Completely at Random
MCMC	Markov Chain Monte Carlo
MDD	Major depressive disorder
MDE	Major depressive episode
MedDRA	Medical Dictionary for Regulatory Activities
MGH	Massachusetts General Hospital
MI	Myocardial infarction
MMRM	Mixed-effect model repeated measure
MNAR	Missing not at random
NIH	National Institutes of Health
PCRS	Placebo-Control Reminder Script
PCS	Physical Component scale
PD	Pharmacodynamic(s)
PDQ-D-5	Perceived Deficits Questionnaire - Depression
PK	Pharmacokinetic(s)
PO	By mouth
POC	Proof of concept
PPS	Per Protocol Set
PRN	Pro re nata (when necessary)
PROMISTM -SD	Subject Reported Outcomes Measurement Information System Sleep Disturbance
PT	Preferred Term

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

QT	Measure of the time between the start of the Q wave and the end of the T wave in the heart's electrical cycle
QTc	Corrected QT
QTcF	QT interval with Fridericia's correction
REL-1017	Esmethadone (dextro- or d-methadone) hydrochloride
Relmada	Relmada Therapeutics, Inc.
RR	Respiratory rate
RNA	Ribonucleic acid
SAFER	State versus Trait; Assessability; Face Validity; Ecological Validity; and Rule of Three Ps (Pervasive, Persistent, and Pathological)
SAP	Statistical Analysis Plan
SCID-5	Structured Clinical Interview for DSM-5 Disorders
SDQ	Symptoms of Depression Questionnaire
SDS	Sheehan Disability Scale
SOC	System organ class
SS	Safety Set
SSR	Sample size re-estimation
TEAE	Treatment-emergent adverse event
ULN	Upper limit of normal

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

2. PURPOSE

The purpose of this statistical analysis plan (SAP) is to ensure that the data listings, summary tables, and figures which will be produced, and the statistical methodologies which will be used, are complete and appropriate to allow valid conclusions regarding the study objectives.

This study is divided into 2 study periods: A 30-day screening phase and a 28-day randomized double-blind treatment phase. A 14-day safety follow-up period after treatment with study medication is also allowed. The focus of this SAP is on the REL-1017 versus placebo treatment comparisons in the double-blind portion of the study.

2.1 Responsibilities

Syneos Health Biostatistics will perform the programming and statistical analyses for this study, and is responsible for the quality control of all tables, figures, and listings for the interim and final analyses.

Separate analysis plans and brief reports will be prepared by expert providers for pharmacokinetic and ECG data.

2.2 Timing of Analyses

The final analyses of safety, efficacy, quality of life and pharmacokinetics data is planned to be conducted after (1) all subjects either terminate their study participation early or complete the final study visit, (2) all data are cleaned per monitoring and data management plans, and (3) all documentation is collected to formally lock the study database.

A planned (unblinded) interim analysis will be conducted when approximately 75% of subjects have completed their Day 28 assessments or terminate early. The purpose of the interim analysis is to estimate conditional power to assess futility or potentially increase the sample size for this trial.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

3. STUDY OBJECTIVES

3.1 Primary Efficacy Objective

The primary efficacy objective of the study is to evaluate the therapeutic efficacy of REL-1017 as an adjunctive treatment versus placebo with antidepressant therapy (ADT) through Day 28 in the Montgomery-Åsberg Depression Rating Scale (MADRS10) total score.

3.2 Secondary Efficacy Objectives

The secondary efficacy objectives of the study are to evaluate the therapeutic efficacy of REL-1017 as an adjunctive treatment versus placebo with the following secondary efficacy measurements:

- MADRS10 response rate at Day 28 (improvement ≥50% compared to Baseline total MADRS10 score)
- MADRS10 remission rate at Day 28 (total MADRS10 score ≤10)
- MADRS10 score at Day 7
- Clinical Global Impression of Severity (CGI-S) score at Day 28

The primary and all secondary efficacy objectives will also be evaluated for the sub-population of severe MDD subjects, defined as those having a baseline MADRS10 total score ≥ 35.

3.3 Exploratory Objectives

Exploratory objectives will evaluate the therapeutic effect of REL-1017 compared to placebo after 28 days treatment in subjects with inadequate response to ongoing ADT on measures of depression, anxiety, sleep, sexual function, cognitive function, select biomarkers, and quality of life.

3.4 Safety and Tolerability Objectives

The safety and tolerability objectives are to:

- Evaluate safety and tolerability of REL-1017
- Evaluate the effect of REL-1017 on electrocardiographic parameters

3.5 Pharmacokinetic Objectives

The pharmacokinetic objective is to characterize the pharmacokinetics (PK) of REL-1017 and its potential metabolites.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

4. STUDY DESIGN

4.1 Brief Description

The REL-1017-302 study is an outpatient, 2-arm, Phase 3, multicenter, randomized, double-blind, placebo-controlled study to assess the efficacy, safety and pharmacokinetics of once-daily oral dosing with 25 mg REL-1017 for MDD in subjects with inadequate response to ADT. Subjects who experience an inadequate therapeutic response to their current ADT (i.e., a marketed ADT treatment course at the approved dosing regimen) will be stratified according to their MADRS10 total score at Baseline (<35, ≥ 35) and then be randomized 1:1 to receive either REL-1017 or placebo for the 28-day double-blind period.

The overall duration of the study, including screening and the double-blind period, will be approximately 58 days per subject. If a subject discontinues from the study prior to Day 28/End of Treatment, the subject will be asked to complete an End of Study visit 14 days after the last dose of study drug.

Approximately 280 subjects (140 per treatment group) will be independently qualified by the investigator and an independent MGH-CTNI-certified clinician to participate in the study. An unblinded interim analysis for sample size re-estimation (IA-SSR) based on the calculation of the conditional power (Mehta and Pocock) and adopting the CHW paradigm (Cui, Hung and Wang) could result in an increase of the total sample size of up to 460 subjects.

Subjects in the 25 mg REL-1017 and matching placebo treatment arms will take the investigational product once-daily by mouth during the double-blind period according to the following schedule:

- Day 1 – loading dose: Three tablets (all 25 mg REL-1017 or all placebo) administered at the clinical study site
- Day 2 to Day 28 – maintenance dose: One tablet (25 mg REL-1017 or placebo) administered at home

Subjects will be asked to stay on a stable dosing regimen of their current ADT throughout the screening and double-blind periods. Subjects should take the investigational product at the same time of day as her/his first-line ADT.

4.2 Subject Selection

After signing an ICF, a screening number will be assigned to the subject and an outpatient Screening Visit will be conducted during which the subject will undergo an interview and medical assessments to determine eligibility for study participation. Subjects who meet all the screening entry criteria will have the Baseline visit scheduled.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

Subjects who are deemed eligible by the Principal Investigator and confirmed by the Medical Monitor will undergo the SAFER/ATRQ interview by clinicians from the Massachusetts General Hospital Clinical Trials Network and Institute (MGH-CTNI).

4.3 Inclusion/Exclusion Criteria

Inclusion Criteria are listed in Section 6.1 and the Exclusion Criteria are listed in Section 6.2 of the study protocol.

4.4 Re-screening

Individuals who sign the ICF to participate in the study but who do not subsequently meet all the requirements as outlined in the inclusion and exclusion criteria and therefore are not randomized (screen failures), may be rescreened once, at the discretion of the Investigator, with the exception of subjects presenting with alcohol or drug abuse; subjects excluded because of QTc prolongation unless they were receiving a QT-prolonging drug which has subsequently been discontinued; or those who had hypokalemia or hypomagnesemia which has been corrected for a minimum of 14 days.

4.5 Study Withdrawal, Removal, and Replacement of Subjects

Subjects who are enrolled and randomized but discontinue before the planned end of study will not be replaced.

Subjects may be discontinued from the study for any reason. Subjects who are discontinued from study drug treatment prior to study completion at Day 28 will be followed up until Day 42 to complete the required assessments as indicated in the Schedule of Assessments (see Protocol Table 2).

4.6 Determination of Sample Size

Approximately 280 subjects (140 per treatment group) are required to demonstrate efficacy for the primary endpoint in this study, the change from Baseline to Day 28 in the MADRS10 total score.

The sample size was selected to detect a difference between means of the REL-1017 and placebo groups equal to Δ=3.8 points with a common standard deviation of $\sigma = 11$ (Cohen's d: $\Delta / \sigma = 0.35$). The difference in treatment group means for this study is anticipated to be larger than the prior REL-1017-301 study due to changes made in this trial's screening criteria to strengthen the selection of subjects with a true MDD diagnosis. The planned standard deviation for the primary endpoint is approximately the same as the one observed in REL-1017-301.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

An unblinded interim analysis is planned to take place when approximately 75% of the total number of subjects have completed their Day 28 visit or have prematurely dropped out of the study. The interim analysis will include sample size re-estimation based on conditional power (CP) with a 20% futility boundary, i.e., a CP below 20% indicates the study is futile. Based on these assumptions, a two-sample, two-sided t-test achieves at least 80% power with 140 subjects per group (N=280 total) and a type 1 error rate of 5% (α=0.05). The number of additional subjects derived from the IA sample size re-estimation procedure will range from no additional subjects (N=280) to a maximum of 180 additional subjects (N=460). Sample size calculations and design of the interim analysis were performed with the software package East® v6.5.

An unblinded statistical analysis team, that is firewalled from the project team and Relmada staff will provide the IA outputs to the DMC members. To enhance data and trial conduct integrity, the details of the CP computations and sample size re-estimation are not included in this SAP but are instead described in the DMC Charter.

4.7 Treatment Assignment and Blinding

Eligible subjects will be centrally randomized in 1:1 ratio to receive either 25 mg REL-1017 or matching placebo. Subjects randomized after implementing amended Protocol Version 8 (or later) will be stratified into one of two strata according to their Baseline MADRS10 total score:

- Baseline MADRS10 total score <35 (moderate depression)
- Baseline MADRS10 total score ≥ 35 (severe depression)

Blocked randomization will be used (within each stratum, after implementing Protocol Version 8.0) to reduce temporal bias and imbalance in the assignment of subjects to treatment, to increase the likelihood that known and unknown subject attributes are evenly balanced between treatment groups, and to enhance the validity of statistical comparisons between the two treatment groups.

Each potential subject will be assigned a unique number at Screening and for the duration of the Screening period (Subject ID Number). This number will be used to identify the subject throughout the study. Subjects who are randomized will be assigned a unique randomization number. Once any subject number is assigned, it cannot be reassigned to any other subject.

The random treatment assignment code used in the IRT will be prepared by an unblinded randomization statistician who has no other role in the study.

4.8 Study Procedures and Flowchart

The visit schedule is displayed in Table 2 of the REL-1017-302 study protocol.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

5. ENDPOINTS

5.1 Primary Efficacy Endpoint

The primary efficacy endpoint is defined as the absolute change from Baseline to Day 28 of the MADRS10 total score for REL-1017 compared to placebo in subjects with inadequate response to ongoing ADT

- The MADRS10 is a 10-item questionnaire with each item scored between 0 and 6, resulting in a total score between 0 to 60. Enrollment eligibility criteria for this study required subjects to have a MADRS10 total score of at least 24 points. Higher MADRS10 scores indicate increasingly severe depression. The questionnaire is most frequently used in clinical studies to measure efficacy of antidepressants. This scale exhibits construct validity (internal homogeneity) and concurrent validity relative to the HAM-D-17 and the concepts of endogenous and non-endogenous depression. The questionnaire includes questions on the following symptoms: (1) Apparent Sadness; (2) Reported Sadness; (3) Inner Tension; (4) Reduced Sleep; (5) Reduced Appetite; (6) Concentration Difficulties; (7) Lassitude; (8) Inability to Feel; (9) Pessimistic Thoughts; and (10) Suicidal Thoughts.

5.2 Secondary Efficacy Endpoints

- MADRS10 response rate (improvement ≥50% compared with total Baseline score) at Day 28, defined as:

  [(Day 28 MADRS10 total score – Baseline MADRS10 total score) / (Baseline MADRS10 total score) × 100%] ≤ -50%

- MADRS10 remission rate (total score ≤10) at Day 28
- Absolute change from Baseline to Day 7 of the MADRS10 total score
- Absolute change from Baseline to Day 28 of the CGI-S score

Secondary Efficacy Endpoints in the Sub-Population with Baseline MADRS10 ≥ 35:

- Absolute change from Baseline to Day 28 of the MADRS10 total score
- MADRS10 response rate (improvement ≥50% compared with total Baseline MADRS10 score) at Day 28
- MADRS10 remission rate (total score ≤10) at Day 28
- Absolute change from Baseline to MADRS10 total score at Day 7
- Absolute change from Baseline Clinical Global Impression of Severity (CGI-S) score at Day 28

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

5.3 Exploratory Endpoints

The exploratory endpoints are:

- Absolute change from Baseline to Day 28 for the Hamilton Anxiety Rating Scale (HAM-A) total score

  The HAM-A is developed to measure the severity of anxiety symptoms. The scale consists of 14 items, each defined by a series of symptoms, and measures both psychic anxiety (mental agitation and psychological distress) and somatic anxiety (physical complaints related to anxiety). Each of the 14 items are scored on a scale of 0 (not present) to 4 (severe), with a total score range of 0 to 56

- Absolute change from Baseline to Day 28 of the Symptoms of Depression Questionnaire (SDQ) total score

  The SDQ is a 44-item, self-report scale designed to measure the severity of symptoms across several subtypes of depression. The 44 items are rated on a 6-point scale. Each item is rated based on a subject's perception of what is normal for the individual (score=2), what is better than normal (score=1), and what is worse than normal (scores=3 to 6)

- Absolute change from Baseline to Day 28 of each Sheehan Disability Scale (SDS) domain score

  The SDS is a self-reported tool and measures the functional impairment in 3 inter-related domains: Work/school, social and family life. The extent of impairment is rated on a 10-point visual scale in each domain

  Total score is sum of the score from each domain: work/school, social life and family life; if no work/school then impute by: *(social life + family life) / 2*

- Clinical Global Impression of Improvement (CGI-I) score at Day 28

  CGI Improvement (CGI-I) measures the improvement since the start of study treatment of the subjects using a 7-point Likert scale ranging from 1=very much improved to 7= very much worse

- Time course of MADRS10 response rate (defined as a ≥50% decrease from Baseline of MADRS10 total score) at all post-Baseline assessment time points

- Time course in MADRS10 remission rate (defined as absolute MADRS10 score ≤10) at all post-Baseline assessment time-points

- Absolute change from Baseline to Day 28 in MADRS10, SDQ, and HAM-A component scores

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

- MADRS10 total score as a function of the time since first diagnosis of MDD
- Correlation between selected biomarkers with therapeutic efficacy as measured via MADRS10 total score until Day 28
- Time course of the changes from Baseline in MADRS10 total score at all post-Baseline assessment time points
- Absolute change from Baseline of the SDS domain score
- Absolute change from Baseline in Digit Symbol Substitution Test (DSST)

  The DSST is an experimental tool to assess cognitive dysfunction. It measures processing speed and attention. The DSST, a symbol-digit pair exercise, requires subjects to match numbers with symbols within 90 seconds. The DSST will be conducted via paper version only and completed at the site visit. The DSST evaluation may not be performed if subjects are unable to attend the site visit.
- Absolute change from Baseline in Perceived Deficits Questionnaire – Depression (PDQ-D-5) – cognition scale

  The PDQ-D-5, a short version of PDQ-D-20, is a subject-reported measure of cognitive symptoms. The PDQ-D-5 measures 4 domains of cognition, namely, attention/concentration, retrospective memory, prospective memory, and planning/organization. Each item is scored on a 5-point ordinal categorical scale, with 0 (never in the past 7 days) to 4 (very often). The total score ranges from 0 to 20, with higher scores reflecting greater cognitive severity
- Absolute change from Baseline in Arizona Sexual Experience Scale (ASEX)

  ASEX assesses five global aspects of sexual dysfunction which includes (1) sex drive, (2) arousal, (3) penile erection / vaginal lubrication, (4) ability to reach orgasm, and (5) satisfaction from orgasm. Each item is scored on a 6-point Likert scale (1-6). Total score ranges from 5 to 30, with higher scores reflecting more sexual dysfunction. Sexual dysfunction is defined by any one of the following conditions: ASEX total score ≥19, or any one item with a score of ≥5, or any 3 items with a score of ≥4

5.4 Pharmacokinetic and Pharmacodynamic Endpoints

An expert PK/PD vendor will prepare a separate analysis plan and report to characterize the REL-1017 PK profile and investigate PK/PD relationships.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

5.5 Biomarker, Pharmacogenomic (DNA), and Expression (RNA) Evaluations

Assessment of biomarkers and their potential relationship to treatment groups and therapeutic efficacy will be explored. Blood samples will be collected to measure genetic and protein markers including, but not limited to, brain-derived neurotrophic factor (BDNF) allelic variants and plasma levels, as well as exploratory metabolite characterization.

5.6 Safety Endpoints

- Treatment-emergent adverse events (TEAEs)
- ECG with heart rate, PR, QRS and QT (with QTcF)
- Vital signs and weight
- Physical examination
- Clinical laboratory parameters (chemistry, hematology and urinalysis)
- Columbia-Suicide Severity Rating Scale (C-SSRS)
- Clinician-Administered Dissociative States Scale (CADSS)

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

6. ANALYSIS SETS

6.1 Screened / Randomized / Enrolled Set

The Screened/Enrolled Set will include all subjects who signed an informed consent.

The Randomized Set will include all subjects randomized and will be used for the presentation of subjects in all subject listings unless noted otherwise.

6.2 Safety Set (SS)

The SS will include all randomized subjects who received any dose of study drug. The treatment group assignment in this population will be according to the treatment actually received. This population will be used for the analysis of safety.

6.3 Full Analysis Set (FAS)

The FAS will include all subjects with inadequate response to ongoing ADT (primary target population) who are randomized, irrespective of any deviation from the protocol or premature discontinuation. Subjects will be analyzed according to their randomized treatment group. The FAS will be used for the analyses of efficacy estimands and endpoints.

6.4 Per-protocol Set (PPS)

The PPS is a subset of FAS and will include subjects who are treated and do not have any major protocol deviations impacting the efficacy assessments. Per-Protocol membership must be determined prior to database lock and treatment unblinding for the final analysis.

6.5 Pharmacokinetic Set (PKS)

The PKS will include all subjects who received at least one dose of REL-1017 and have at least one measured PK concentration.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

7. PROTOCOL DEVIATIONS

All protocol deviations related to study inclusion or exclusion criteria, conduct of the trial, subject management, dosing, and sampling procedures or subject assessment will be listed. The list of protocol deviations will be reviewed by the Sponsor, the principal investigator and the study statistician, and finalized before database lock. Determination of whether the deviation is major or minor will be determined prior to database lock.

Prohibited and/or Restricted concomitant medications and therapies are detailed in Protocol Section 7.10.

The number (%) of subjects will be presented for protocol deviations and major protocol deviations. Major protocol deviations will be further defined as the number (%) of subjects in each deviation category.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

8. ESTIMANDS FOR PRIMARY AND SECONDARY EFFICACY ANALYSES

8.1 Main Estimand for Primary Efficacy Analysis

The **main estimand for the primary efficacy analysis** is defined with the following attributes:

A. **Treatment condition:** Oral administration of 25 mg REL-1017 or matching placebo

B. **Population:** Randomized subjects as defined by the FAS (see Section 6)

C. **Variable/Endpoint:** Change from Baseline to Day 28 in the MADRS10 total scores

D. **Intercurrent events (ICE):** See **Table *1*** below.

E. **Population-level summary:** Difference between REL-1017 and placebo group mean changes from Baseline to Day 28 in MADRS10 total scores as estimated by MMRM.

The main estimand for the primary efficacy analysis will use a "treatment-policy" strategy, in which the observed data will be used for analysis, i.e., the occurrence of ICEs will be ignored and there will be no values imputed for missing data. Missing data will be handled within the MMRM analyses by a so-called direct likelihood method which is appropriate when it can be assumed that the missing data are MAR. The only exception to the treatment policy strategy is for the (few, if any) cases of death (ICE6) that will utilize a "while-alive strategy"; i.e., the last available MADRS10 value prior to the death will be used for the MMRM analysis.

8.2 Supportive Efficacy Estimand for Primary Efficacy Analysis

A **supportive estimand for the primary efficacy analysis** will be used to evaluate the robustness of the main estimand analysis. The treatment condition, population, variable/endpoint, and handling of missing data for this supportive estimand will be the same as the main estimand. However, for the supportive estimand the ICEs will be handled by use of a "hypothetical strategy", in which the data after the ICE will be replaced by data generated from MI modelling. The MI model will use data from 'similar' subjects defined by baseline depression severity (MADRS10 total score at baseline <35 or ≥35) and by age class (<50 years or ≥50 years). Two different data sources will be used for the different data imputation strategies: Reference-based MNAR-MI modelling will use subjects from the placebo group, and MAR-MI modelling will use subjects from the same treatment group. The MI approach to use will depend on the specific ICE that is observed in the study (see **Table 1)**. Cases of death (ICE6) will have data treated in the same way as for the main estimand. The specific ICEs and strategies for handling the ICEs with the supportive estimand are presented in **Table 1** alongside the main estimand.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

Table 1. Strategies for Handling of Intercurrent Events and Missing Data

	ICE	Strategy for Main Estimand	Strategy for Supportive Estimand
ICE1	Treatment discontinuation due to lack of efficacy	Treatment policy	Hypothetical using reference-based MNAR-MI
ICE2	Treatment discontinuation due to AE related to study medication	Treatment policy	Hypothetical using reference-based MNAR-MI
ICE3	Treatment discontinuation due to AE unrelated to study medication	Treatment policy	Hypothetical using MAR-MI
ICE4	Treatment discontinuation due to other reasons	Treatment policy	Hypothetical using MAR-MI
ICE5	Treatment with prohibited medications classified as major protocol deviations impacting efficacy	Treatment policy	Hypothetical using reference-based MNAR-MI
ICE6	Death	While alive (use last available MADRS10 value)	While alive (use last available MADRS10 value)

The frequency of the types and timings of intercurrent events will be summarized by treatment group with further details provided in listings.

8.3 Secondary Efficacy Endpoints and Estimands

Table 2 below summarizes the estimands for the secondary efficacy endpoints. No adjustments for type I error inflation (due to multiple endpoints) will be made in evaluating the four secondary endpoints.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

Table 2. Estimands for Secondary Endpoints

1. Estimand for MADRS10 Response at Day 28 (Visit 7)	
Variable/Endpoint	Binary endpoint whereby a subject is considered a MADRS10 responder if he/she has an observed MADRS10 total score at Day 28 which is at least 50% less (improved) than Baseline; otherwise, the subject is deemed a non-responder.
ICE	All ICEs except death will be handled with a treatment policy strategy. Death will be handled by the composite strategy, in which the subject is considered a non-responder.
Population-level summary	Difference between the REL-1017 and placebo group proportions of subjects with MADRS10 response at Day 28. A missing Day 28 assessment will be MAR-MI imputed on the MADRS10 scale, dichotomized as a binary endpoint, and then combined by Ruben's rule.
2. Estimand for MADRS10 Remission at Day 28 (Visit 7)	
Variable/Endpoint	Binary endpoint whereby a subject is considered to be in remission if he/she has an observed MADRS10 total score ≤10 points at Day 28; otherwise, the subject is considered to not be in remission.
ICE	All intercurrent events except death will be handled with a treatment policy strategy. Death will be handled by the composite strategy, in which the subject is considered to not be in remission.
Population-level summary	Difference between the REL-1017 and placebo group proportions of subjects with MADRS10 remission at Day 28. A missing Day 28 assessment will be MAR-MI imputed on the MADRS10 scale, dichotomized as a binary endpoint, and then combined by Ruben's rule.
3. Estimand for Change from Baseline in MADRS10 Total Score on Day 7 (Visit 4)	
Variable/Endpoint	Change from Baseline to Day 7 in the MADRS10 total score.
ICE	Same as the main estimand for the primary efficacy endpoint.
Population-level summary	Difference of the REL-1017 and placebo group mean changes from Baseline to Day 7 in the MADRS10 total score.
4. Estimand for Change from Baseline in CGI-S Score on Day 28 (Visit 7)	
Variable/Endpoint	Change from Baseline to Day 28 in CGI-S.
ICE	Same as the main estimand for the primary efficacy endpoint.
Population-level summary	Difference between the REL-1017 and placebo group mean changes from Baseline to Day 28 in the CGI-S score.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

9. GENERAL ASPECTS FOR STATISTICAL ANALYSIS

9.1 General Methods

In general, summaries will display data by treatment group for each scheduled assessment time where appropriate.

Unless stated otherwise, descriptive summaries will include n, mean, standard deviation, median, first and third quartiles, minimum and maximum for continuous variables, and n and percent for categorical variables.

Adverse events and medical history will be coded according to the most recent version of Medical Dictionary for Regulatory Activities (MedDRA). Adverse events will be classified using the NIH Common Terminology Criteria for Adverse Events v5.0 (CTCAE 5.0).

Prior and concomitant medications will be coded using the most recent version of the WHO Drug Dictionary. The actual version of MedDRA and WHO drug used will be noted in the footnote of the respective output.

9.2 Missing Data

The handling of missing data for various analyses of the primary efficacy endpoint and secondary efficacy endpoints are presented in Section 8. Details of the MI algorithm are provided in Section 15.2.

Unique issues not addressed in the above sections are described below.

- Missing baseline assessments of the primary and secondary endpoints will be imputed for analysis. If a subject is missing their baseline value but has non-missing post-baseline data, then single imputation of their baseline value will be performed using the first non-missing post-baseline assessment.
- If a subject has a baseline value and has been randomized, but not treated with any REL-1017 or placebo study medication, then the subjects post-baseline data will be single imputed by their baseline value.
- No missing data imputations will be performed for exploratory endpoints.
- If a individual items/questions are missing from the MADRS10 (or other questionnaires), then the calculation of the total score will be performed using the validated method described in the "Scoring Manual" or publication. In absence of that guidance, the following process will be used:

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

- If a single item/question is missing for a subject, then the total score will be estimated by the proportion of completed items; e.g., if one MADRS10 item is missing and the total score of the 9 available items is 32, then the MADRS10 total score will be estimated as ROUND (32*10 / 9) = ROUND (35.56) = 36, in which rounding is done to the closest whole number.
- If more than one item/question is missing, then the total score will be assumed to be missing and will be imputed as described in the estimand Section 8; i.e., (1) Non-monotone missing data will be imputed using SAS PROC MI with the Markov Chain Monte Carlo (MCMC) statement to create datasets with monotone missing data patterns, and (2) The monotone missing values will then be multiply imputed using SAS PROC MI with the MONOTONE statement.

All safety listings will have missing or incomplete dates left as they have been recorded. However, for calculation and sorting based on dates and for consideration in summary tables, the following method will be used:

For partial or missing AE start dates the following imputation will be applied:

1. If year is not missing and is after the year of first dose:
   a. If month is missing, then month will be imputed as January.
   b. If day is missing, then day will be imputed as the first of the month.
2. If year is not missing and is the same as the year of the first dose:
   a. If month is missing, then impute the month as the month of the first dose date.
   b. If day is missing, and the month is the same as the month of the first dose date, then impute day as the day of the first dose date.
   c. If day is missing but month is after the month of first dose date, then impute day as the first day of the month.
3. If year is missing, then impute the year as the year of the first dose date:
   a. If month is missing, then impute the month as the month of the first dose date.
   b. If day is missing, then impute the day as the day of the first dose date.
4. If the start date is completely missing, but the AE is either ongoing or the stop date is after the first dose date then impute the start date as the first dose date.
5. For any cases involving the rules above, if the AE end date is before the AE start date, then do not impute the AE start date and assume that the AE is treatment emergent for the purpose of the analysis. Further, if the AE stop date occurs prior

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

to the first dose date, do not impute the AE start date, and assume that the AE is not treatment emergent.

No imputations will be applied to AE stop dates.

A missing/incomplete date of medical history or disease diagnosis will be assumed to have occurred before any study treatment.

Missing TEAE relationship will be imputed by 'related' for tabulation.

9.3 Visit Windows and Definitions

Any analysis summarized or modelled by visit will use analysis visits assigned by study day ranges. All safety or efficacy assessments (scheduled, unscheduled or early termination (ET)) will be mapped to an analysis visit as defined in **Table 3** below. In cases where a subject has multiple assessments windowed to the same analysis visit, the assessment closest to the protocol specific target day will be selected for analysis. End of Study/Follow-up will not be windowed.

First dose date and time will be assigned based on the Loading Dose Date and Time from the "Study Drug Administration" CRF page. If the loading dose information is not available, the first IRT dispense date following randomization can be assigned.

A subject's **Baseline** value is defined as the last non-missing value prior to the start date and time of study drug administration.

Study days prior to treatment will be calculated as: date of event/collection – first dose date of the study. Study days after treatment will be calculated as: date of event/collection – first dose date of the study + 1. (Day 1 is defined as the first date of study drug administration, i.e., the first dose date).

Table 3. Evaluation Intervals for Efficacy Analyses

Scheduled Visit	Protocol-Specified Target Day	Interval for Analysis Visit
Visit 2	Day 1 (Baseline)	Last non-missing value prior to, or shortly after, start of study drug at clinical site
Visit 4	Day 7	Day 2 to 10
Visit 5	Day 14	Day 11 to 17
Visit 6	Day 21	Day 18 to 24
Visit 7	Day 28	Day 25 to 31

Note: The Screening visit will be performed up to 30 days prior to the Baseline visit.

Actual dates and times of blood draws will be used for PK analyses rather than scheduled days/visits and times.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

9.4 Programming Considerations and Quality Control

All tables, figures, data listings (TFLs), and statistical analyses will be generated using SAS® for Windows, Release 9.4 (SAS® Institute Inc., Cary, NC, USA) or equivalent R programming.

Computer programs will be developed to produce output such as SDTM and ADaM datasets, summary tables, data listings, figures or statistical analyses. Quality control procedures are performed for all SAS programs and output as documented in SOPs and procedures. Quality control is defined here as the operational techniques and activities undertaken to verify that the programs produce the output by: (1) Checking the logic/algorithms of analysis specifications, (2) Independent double-programming of analysis datasets and output programs, and (3) Statistical review of outputs and independent confirmation of endpoints.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

10. DEMOGRAPHIC, OTHER BASELINE CHARACTERISTICS AND MEDICATION

10.1 Subject Disposition and Withdrawals

Summary tables reflecting the number and percentage of subjects for the following categories will be presented by treatment group and overall:

- Screened/enrolled subjects (only absolute number for overall)
- FAS, SS, PPS, PKS
- Subjects who complete the double-blind study treatment of 28 days
- Subjects who early terminate study treatment together with reasons
- Subjects who early terminate the study together with reasons

Subject disposition will be listed for all randomized subjects. Additionally, Screen Failure subjects will be listed with reason for failure.

A listing of subjects with attended visits will be provided. Additionally, a frequency table of number and percentage of subjects who completes each visit will be presented by treatment group.

10.2 Demographic Characteristics

Demographics characteristics (age, sex, race, ethnicity, body weight, height, body temperature, and BMI) will be summarized using descriptive statistics for the FAS and the Safety population. No formal statistical comparison between the groups will be performed. Demographics, height, and BMI will be recorded at Screening only. Weight will be recorded at Day 1 and at Day 28. Demographics and baseline characteristics will be summarized by subgroup, additionally.

Age (years) will be calculated based on date of signed informed consent as: (date of informed consent - date of birth) / 365.25 and truncated to complete years.

BMI (kg/m^2) will be calculated as: Weight (kg) / $[\text{Height (cm)}/100]^2$.

10.3 Baseline Disease Characteristics

Baseline Disease characteristics (time since first diagnosis of MDD, Age MDD began to impact function, Number of Lifetime Depressive Episodes, Number of Depressive Episodes within the Past 5 years, Duration of Current Major Depressive Episode, antidepressant tachyphylaxis-tolerance (Yes, No), treatment-resistant depression (Yes, No) and prior participation in clinical trials (Yes, No) will be summarized using descriptive statistics for the FAS and the Safety population. No formal statistical comparison between the groups will be performed.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

Time Since First Diagnosis of MDD (Years) will be calculated based on the first diagnosis date in the medical history CRF page: (Date of informed consent - date of diagnosis) / 365.25. This will be categorized by <5 years or >=5 years.

Duration of Current Major Depressive Episode (Years) will be calculated from Medical and Psychiatric History Log CRF page where MedDRA Preferred Term is "Major depression"; as (first dose date - diagnosis date)/365.25.

Age MDD Begin to Impact Functioning, Number of Lifetime Depressive Episodes and Number of Depressive Episodes within the Past 5 years is collected by CRF.

Antidepressant tachyphylaxis-tolerance, treatment-resistant depression, and prior participation in clinical trials will be identified by the client.

10.4 Baseline Assessments

The baseline assessments of all efficacy endpoints will be summarized using descriptive statistics for the FAS and the Safety population. No formal statistical comparison between the groups will be performed.

Assessments at screening of the Safer Interview, of Structured Clinical Interview for DSM-5 (SCID-5), of Antidepressant Treatment Response Questionnaire (ATRQ) and of Hamilton Depression Rating Scale (HAM-D-17) will be listed only.

10.5 Medical and Psychiatric History

Medical history and psychiatric history will be coded into the most recent version of Medical Dictionary for Regulatory Activities (MedDRA®). Medical and psychiatric history will be listed by treatment arm. No summaries will be provided.

10.6 Prior and Concomitant Medications

Prior medication is defined as any medication which has a start date prior to the date of first dose of study drug administration.

Concomitant medication is defined as any medication which has an end date or is continuing after the date of first dose of study drug administration or has a missing end date.

Prior and concomitant medications will be listed by treatment arm and subject. Medication history (prior medications) will be listed at Screening, and concomitant medications will be listed at every visit from Day 1 through EOS (Day 42) and summarized by treatment arm.

Listings will include the therapeutic (i.e., the second level of Anatomical Therapeutic Chemical [ATC] classification system code, that is, corresponding to the first 3 figures), preferred name

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

(the fourth level of the ATC code, that is, corresponding to the first 5 figures) and verbatim text. The listings will be sorted by treatment group, subject id, chronological start date, stop date, therapeutic class, preferred name and verbatim name for the FAS.

A frequency table of the number and percentage of subjects will be provided for prior medications and concomitant medications by therapeutic class and preferred name for each treatment group and overall for the FAS.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

11. ANALYSIS OF EFFICACY

11.1 Primary Analysis of Efficacy

The primary analysis of efficacy will be performed using an MMRM and main estimand. The MMRM will include the four post-baseline (Visits 4-7) MADRS10 total scores for each subject as (correlated) dependent outcomes. Treatment, visit, the interaction of treatment and visit, and the baseline total MADRS10 score will be independent variables in the MMRM. The main estimand approach to handling of missing MADRS10 total scores and occurrences of ICEs is described in Section 8.1.

An unstructured variance-covariance matrix will be used in the MMRM. The denominator degrees of freedom will be estimated according to the Kenward-Roger method. In the unlikely event that the MMRM estimation does not converge, alternative specifications of the variance-covariance matrix structure will be adopted as appropriate. The order of variance-covariance structural specification will be: (1) unstructured, (2) first-order ante dependence, and (3) first-order heterogeneous autoregressive with the first to result in computational convergence reported. If the unstructured variance-covariance matrix does not result in convergence of the model and a structured variance-covariance matrix is used, then an empirical (sandwich) estimator will be used to estimate the standard error for the fixed effect estimates (using the EMPIRICAL statement in PROC MIXED).

The results of the primary efficacy analysis using the MMRM will be presented for the main estimand. The treatment group LSMeans, LSMean Differences, standardized effect sizes, p-values from the 2-sided tests, and 95% C.I.s will be presented to evaluate efficacy. The standardized effect size is calculated by dividing the LSMean difference by the corresponding standard deviation. If the final sample size is increased after the interim analysis, then the final primary efficacy analysis will be conducted using the CHW and Lawrence and Hung methods as described in Section 15.1.

The MMRM results for Day 28 (Visit 7) will be considered as the primary efficacy comparison of REL-1017 versus placebo. The same analyses described above will also be repeated for the supportive estimand of the primary efficacy analysis as described in Section 8.2.

Analogous results for the other post-baseline visits (Day 7 MADRS10 total score change from Baseline is a secondary endpoint) will also be summarized for each treatment group in tabular format, and the LSMeans ± SD for each visit and treatment group will be plotted over time to assess onset and durability of response.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

11.2 Primary Efficacy Sensitivity and Supplemental Analyses

The following primary efficacy sensitivity analyses will be performed to assess alternative assumptions to the ones provided in the main estimand. Details of the MI algorithms are provided in Section 15.2.

1. MI-MAR for MMRM: As a first sensitivity evaluation of the primary efficacy analysis, the MMRM analysis will be repeated after MAR-MI of the missing values of MADRS10 total score, i.e. missing data will be multiply imputed using data of similar subjects from the same treatment group. Similar subjects will be defined by depression severity (MADRS10 total score at baseline <35 or ≥35) and by age class (≤50 years/>50 years).

2. MI-MNAR for MMRM: As a second sensitivity evaluation of the primary efficacy analysis, the MMRM analysis will be repeated after MNAR-MI of the missing values of MADRS10 total score, i.e. missing data will be multiple imputed using data of similar subjects from the placebo group. Similar subjects will be defined by depression severity (MADRS10 total score at baseline <35 or ≥35) and by age class (≤50 years/>50 years).

3. A common alternative to MMRM analysis is an ANCOVA analysis at the end of treatment. For this study, the ANCOVA model will include the change from Baseline to Day 28 in the total MADRS10 score as the outcome, and treatment and baseline as independent variables (so-called, parallel slopes ANCOVA). MAR-MI will be used to impute missing Day 28 MADRS10 total scores.

   Prior studies indicated a larger treatment difference with increasing Baseline MADRS10 total scores. To investigate that observation in this trial, the same ANCOVA above will be conducted with the addition of an independent variable for the treatment*baseline interaction (so-called, non-parallel slopes ANCOVA). Treatment differences will vary by baseline level and will be evaluated at the quartiles (Q1, Q2, Q3) of the overall Baseline MADRS10 total scores.

4. A tipping point analysis will be used to investigate the robustness of the primary efficacy analysis results. The MAR-MI MMRM model (main estimand) will utilize an increasing delta shift (towards NMAR) that will be added and the MMRM model will be repeated until the tipping point is reached, i.e., when the significant result (if applicable) changes to non-significant. Details are provided in Section 15.3.

The following descriptive, exploratory and supportive efficacy analyses will be performed:

1. A descriptive summary table of the observed MADRS10 actual total score (i.e., without any manipulation for estimand and missing data), changes from baseline, the raw treatment difference with 95% CIs and the standardized effect size will be presented for each scheduled post-baseline time point by treatment group.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

2. The results of the MMRM (with MAR imputation by direct likelihood), i.e., LSMeans, LSMean Difference and standardized effect size and p-value will be presented for all scheduled post-baseline visits on FAS.
3. The main and the second/supportive estimand analyses will be repeated using the PPS (rather than FAS).

Figures will be presented for observed MADRS10 actual total scores and change from baseline, and for MADRS10 MMRM estimated LSMean change from baseline (results of the MMRM with MAR imputation by direct likelihood): Estimates and corresponding 95% CI will be presented for each scheduled post-baseline time point.

The MADRS10 response rates and remission rates will be displayed as a cumulative response analysis curve for REL-1017 and placebo with the assessment time displayed on the x-axis (days) and the cumulative response rate in % on the y-axis.

11.3 Analysis of Secondary Efficacy Endpoints

The following four secondary endpoints will be analyzed as per the estimands defined in Section 8.3. All analyses will be repeated using the PPS (rather than FAS).

- MADRS10 Response at Day 28.

  The analysis of this secondary efficacy endpoint will be performed using a Chi-square test (2-sided with α=0.05) on the FAS population for the comparison of REL-1017 versus placebo.

  The raw proportion of responders (numerator is number of responders at Day 28 and denominator is the total number of subjects) for each treatment group and the estimated proportion following MI will be presented with Wilson 95% CIs for both. In addition, 95% CIs for treatment differences between REL-1017 treatment group and placebo will be computed.

  Additionally, as sensitivity analysis, subjects with missing Day 28 assessment of MADRS10 will be considered as non-responder (instead of multiple imputation): estimated response proportions for the treatments will be presented with Wilson 95% CIs, and the 95% CIs for treatment differences between the REL1017 and placebo groups.

- MADRS10 Remission at Day 28

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

The analysis of this secondary efficacy endpoint will be performed in the same manner as Response analysis.

Additionally, as sensitivity analysis, subjects with missing Day 28 assessment of MADRS10 will be considered to not be in remission (instead of estimated by MI): estimated response proportions for the treatments will be presented with Wilson 95% CIs, and the 95% CIs for treatment differences between REL1017 and placebo groups.

- Change from Baseline in MADRS10 Total Score on Day 7

  The analysis of this secondary efficacy endpoint will be performed similarly to the primary analysis of efficacy.

- Change from Baseline in CGI-S Score on Day 28

  The analysis of this secondary efficacy endpoint will be performed similarly to the analysis of efficacy.

11.4 Analyses of Subgroups

The primary efficacy analysis with the main estimand and the secondary efficacy analyses will be repeated for the following demographic and baseline defined subgroups: Race (White, non-White), Gender (Male, Female), Age (≤ 50 years or > 50 years) and Baseline Depression Severity (Baseline MADRS10 total score <35 or ≥35). The FAS will be divided according to each defined subgroup to conduct the analyses, e.g., the FAS will be separated by White and non-White subjects to conduct the Race subgroups analyses.

Additional post-hoc analyses of other subgroups may be performed as appropriate.

11.5 Exploratory Efficacy Endpoint Analyses

The continuous endpoint HAM-A and MADRS10 component scores will be analyzed in the same manner as the main estimand analysis. Exploratory endpoints without repeated measures (SDQ, SDS, DSST, PDQ-D-5, ASEX) will be evaluated by an ANCOVA model with MAR MI.

Analysis on the Work/school domain score of SDS will exclude subjects who did not work during the past week due to reasons unrelated to major depressive disorder.

The SDS score will be categorized (SDS>=5 or SDS<5): number and percentage of subjects will be provided for the categories by treatment group for each scheduled time point.

A summary table of the CGI-I score will be presented for each scheduled time point by treatment group. The analysis will be performed using an MMRM model. Denominator degrees of freedom and the covariance structure will be handled in a manner identical to the primary efficacy analysis.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

The CGI-I score will be categorized into improved (CGI-I ≤3) versus no change/worsened (CGI-I ≥4): number and percentage of subjects will be provided for the categories by treatment group for each scheduled time point.

11.6 Pharmacokinetics

PK analyses will be described in a separate PK Analysis Plan and results will be provided in a separate PK Report.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

12. SAFETY ANALYSES

The Safety Set (SS) population will be used for all safety analyses.

All safety assessments will be listed by treatment group, subject ID and visit/timepoint if appropriate. Unscheduled assessments will be flagged and presented in listings. Adverse events will be listed by treatment group, subject ID, start date/time and stop date/time.

12.1 Extent of Exposure

The duration of study drug exposure (in days) and the total number of tablets taken will be summarized by treatment group and overall. The duration of exposure (days) will be calculated as date of last study drug intake – date of first study drug intake +1.

Duration of study drug exposure and the total number of tablets taken each day will be listed.

12.2 Treatment Compliance

The treatment compliance (%) is calculated as (total number of tablets taken / total expected number of tablets taken) x 100%. The treatment compliance will be categorized into <80%, 80% to 100%, >100% to 120%, >120%.

A summary table of the treatment compliance will be provided for each treatment group and overall. The compliance categories will be presented by number and percentage of subjects in each category by treatment group and overall. Treatment compliance and categorized compliance will be listed.

12.3 Adverse Events / Adverse Drug Reactions

All AEs will be summarized by SOC and PT for each treatment group and in total. A treatment-emergent AE (TEAE) is an AE that starts or worsens at any time after initiation of study drug. Overall and SOC/PT summaries will include the number of unique subjects with an event, the percent within respective treatment group and the total occurrences of that AE type.

An overall summary will be provided of subjects reporting AEs, TEAEs, treatment-related (possibly, probably and definitely related) TEAEs, serious TEAEs, treatment-related serious TEAEs, TEAEs leading to study drug withdrawal and TEAEs leading to death, Worst Relationship and TEAEs by maximum severity.

The TEAEs, treatment-related TEAEs, serious TEAEs, treatment-related serious TEAEs leading to study drug withdrawal and TEAEs leading to death will be summarized by SOC and PT for each treatment group and total. Another summary of TEAEs by PT only for each treatment group and total. TEAEs will also be summarized by SOC and PT for maximum intensity and maximum

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

causality for each treatment group and total. TEAEs by SOC and PT will be summarized by gender subgroup and Treatment group.

Summary tables of AEs during and after the treatment with time-adjusted count- and incidence rates will be presented.

12.4 Laboratory Evaluations

Laboratory data will be summarized by the type of laboratory test (clinical chemistry, hematology, urinalysis- continuous data) and scheduled visit (refer to Section 10.5 of the protocol for the List of Proposed Laboratory Assessments). The descriptive statistics will be calculated and presented for each time point by treatment (absolute values and change from baseline). For categorical data of urinalysis, tables of the number and percentage of subjects for each scheduled visit by treatment group will be provided.

Shift tables of the number and percentage of subjects below normal range (low), within normal range (normal) and above normal range (high) at each scheduled post-baseline visit compared to the low/normal/high categorization at baseline will be provided by treatment group. A listing of clinically significant abnormal values will be provided in the tables. Biomarkers and genetic markers will also be summarized by treatment group. Pregnancy test results, drug abuse screen results, breath alcohol data, Hepatitis B, Hepatitis C, HIV assessments, viral screen data, COVID-19 PCR, and the thyroid panel will be listed only.

Additionally, a summary of the number of patients with liver test abnormalities will be provided for the worst value after first dose of study drug including ALT, AST, ALP and total bilirubin as a function of test's upper limit of normal (ULN); Also summarizing ALT or AST (>3xULN or >5xULN) with elevated total bilirubin or fractionation of bilirubin. A listing of all patients with liver test abnormalities will be provided including all laboratory values and corresponding abnormality.

12.5 Vital Signs

Vital signs (body temperature (°C), pulse oximetry, RR, HR, and systolic and diastolic BP measurements), weight, and BMI will be summarized for each time point by treatment group and overall. Summary tables will present actual values and the change from baseline.

12.6 ECG

For each 12-lead ECG variable (average of triplicate assessments, where applicable) the actual value and the change from baseline will be summarized for the scheduled visits by treatment group and overall.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

A shift table of the number and percentage of subjects with normal and abnormal, and clinically significant abnormal values at each scheduled post-baseline visit compared to the normal / abnormal /clinically significant abnormal categorization at baseline will be provided by treatment group and in total for the overall interpretation. A frequency table with number and percentage of subjects with abnormal clinically significant overall interpretation will be presented for each scheduled visit by treatment group, and total. Additionally, a frequency table with number and percentage of subjects with noteworthy QTcF values (>450 msec in subjects) during the study and of subjects noteworthy QTcF changes from baseline (≥ 30 but <60; >60 ms) will be provided by treatment group and total.

Further ECG analyses, including possible evaluation of a potential exposure-response relationship or a population PK/QTc model may be detailed in a separate ECG Statistical Analysis Plan.

12.7 Physical Examination

Physical examination results will be presented in data listings only.

12.8 C-SSRS

The C-SSRS data, i.e., measure of suicidal ideation severity, including suicidal ideation intensity, suicidal behavior, non-suicidal self-injuries behavior and lethality, will be summarized by treatment group and visit; It will include a categorical summary of subjects answering “Yes” to questions of C-SSRS form, individual questions will be grouped by suicidal behavior or suicidal ideation along with actual suicides.

A shift table will be provided to demonstrate changes in C-SSRS categories from baseline to the worst post-baseline category (suicidal behavior being considered worse than suicidal ideation).

12.9 CADSS and MADDERS

The CADSS total score and subscale scores will be summarized by treatment group and visit. Actual values and changes from Baseline will be presented.

The MADDERS Decision Report will be listed by treatment group.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

13. INTERIM ANALYSES

An unblinded interim analysis will be conducted when approximately 210 (75%) of the randomized subjects have completed their Day 28 assessments or prematurely terminated from the study.

The IA will be based on the main estimand of the primary efficacy endpoint to compute conditional power (CP) for purposes of sample size re-estimation. The interim analysis will be conducted by an independent, firewalled, unblinded statistical team, and the analysis outputs will be provided to the Data Monitoring Committee (DMC) members for this study. The details of the CP computations and the pre-specified sample size re-estimation decision rules will be provided in the DMC Charter.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

14. REFERENCES

1. Cui, L, HMJ Hung and SJ Wang. "Modification of sample size in group sequential clinical trials." *Biometrics* 55 (1999): 853-857.
2. Lawrence, J and HMJ Hung. "Estimation and confidence intervals after adjusting the maximum information." *Biometrical Journal* 45 (2003): 143-152.
3. Mehta, CR and SJ Pocock. "Adaptive increase in sample size when interim results are promising: a practical guide with examples." *Statistics in Medicine* 30 (2011): 3267-3284.

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

15. APPENDIX – STATISTICAL METHODS DETAILS

15.1 Weighted Test-Statistic and Updated Point and Confidence Estimates

If the initial sample size is increased after the IA sample size re-estimation, then the following method will be used to compute the inferential statistics for the final analysis.

- The data collected up to the IA and the data collected after the IA will be combined for the final analysis by using the Cui-Hung-Wang (CHW) method (Cui, Hung and Wang) as summarized below
- The point and confidence interval estimates will be calculated using the formula presented by Lawrence and Hung (Lawrence and Hung) as described below
- The CHW method [1] (Cui, Hung and Wang) is described below. The pre-specified weights for the two stages are based on the fraction of the planned number of subjects at the interim analysis among the total number of subjects: $w_1 = \frac{n_1}{N} = \frac{210}{280} = 0.75$ and $w_2 = 1 - w_1 = 0.25$ (the actual/observed sample sizes will be used to compute the weights). In case of sample size increase, the data collected up until the time of the IA and the data collected after the IA will be combined with the CHW weighted test statistic as follows:

  $$Z_{CHW} = \sqrt{w_1} \cdot Z_1 + \sqrt{w_2} \cdot Z^*_{2,new} ,$$

  in which Z_1 is a Wald-type test-statistic computed from data for the n_1 subjects used in the IA, and $Z^*_{2,new}$ is a Wald-type test-statistic computed from data for the m subjects collected after the IA (with $m > N - n_1$). Note that for the final analysis Z_1 will be computed from the <u>final</u> data in the cleaned and locked database, which may be slightly different than the data used for computations at the time of the interim analysis.

  This weighted test-statistic preserves the type-1 error (α) when increasing sample size.

- Updated point estimates and confidence interval estimates

  Lawrence and Hung (Lawrence and Hung) provide a method for obtaining confidence intervals and point estimate in utilizing the maximum information if the sample size is increased after the interim analysis. The updated point estimate (1) and confidence interval estimates (2) of treatment effect per Lawrence and Hung (Lawrence and Hung) are calculated as below:

  $$\hat{\delta} = \frac{t_1\hat{\delta}^{(1)} + \sqrt{1-t_1}\sqrt{t^*-t_1}\hat{\delta}^{(2)}}{t_1 + \sqrt{1-t_1}\sqrt{t^*-t_1}} \quad (1)$$

  $$\hat{\delta} \pm \frac{\hat{\delta}}{Z_{CHW}} Z_{\frac{\alpha}{2}} \quad (2)$$

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

where:

$$t_1 = \frac{total\ number\ of\ subjects\ at\ interim}{total\ number\ of\ subjects\ originally\ planned}$$

$$t^* = \frac{final\ total\ number\ of\ subjects}{total\ number\ of\ subjects\ originally\ planned}$$

$$k = \frac{1}{\sigma^2}\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

$$\hat{\delta}^{(1)} = Least\ square\ mean\ difference\ from\ stage\ 1\ analysis$$

$$\hat{\delta}^{(2)} = Least\ square\ mean\ difference\ from\ stage\ 2\ analysis$$

Note that the term k is not used in computing the point estimate or C.I. but refers to the paper Lawrence and Hung (Lawrence and Hung).

15.2 Multiple Imputation Approach

The described multiple imputation approach (either assuming MAR or MNAR) will be applied to impute missing data due to study termination and will also be used to model data for the hypothetical intercurrent event strategy (see Section 8).

Missing post-baseline values will be imputed via MI assuming MAR or MNAR as appropriate for each scheduled visit (i.e., Days 7, 14, 21, 28).

Fifty imputed datasets will be created to calculate change from baseline. Each of the imputed datasets will be analyzed via the specified efficacy endpoint analysis. The results across the multiple imputed data sets will be combined using SAS Proc MIANALYZE.

If the missing data does not follow a monotone pattern, a sequential approach to imputing the data will be applied. First, a Markov chain Monte Carlo (MCMC) method will be used to produce a monotone missing data pattern using the MCMC impute=monotone option in Proc MI (SAS System). Following the method described in Smith (2017) the non-monotone data imputation will be performed by the categorical treatment variable.

The MAR and MNAR multiple imputations will be performed using monotone linear regression imputation methods which will impute the patient's missing post-baseline change scores at each scheduled visit. The imputation model will consider data of 'similar patients', in which similarity will be defined by categorized age and by categorized baseline MADRS10 total score.

Example code for the MI procedures is given below. The variable TRTp represents assigned treatment arm, BASE represents the baseline observation, _2 through _5 represent observations at Day 7 to Day 28, _IMPUTATION_GROUP_ denotes which group (placebo or active) the

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

model should impute the observation from, ageclass is categorized age as ≤50 years or > 50 years and severe is categorized baseline MADRS10 total score as <35 or ≥ 35.

15.2.1 Example Code for Primary Analysis

```
** Convert AVAL to horizontal structure;

PROC MI DATA=madrs_NIMPUTE=50 OUT=MI_MONOTONE SEED=25140;
  by TRTp;
  VAR BASE _2 _3 _4 _5;
  MCMC CHAIN=single IMPUTE=MONOTONE;
RUN;

PROC SORT DATA=MI_MONOTONE;
  by _IMPUTATION_ _IMPUTATION_GROUP_ usubjid;
run;

The above statement produces a monotone missing pattern in the
data.
```

15.2.2 Example Code for MNAR (Missing Not at Random)

```
** Full regression imputation of the monotone missing data sets
via MNAR using Placebo distribution - that means reference-based
imputation using copy reference **;

PROC MI data=MI_MONOTONE NIMPUTE=1 OUT=MI_COMPLETE1 SEED=25140;
  CLASS TRTp ageclass severe;
  VAR ageclass severe base _2 _3 _4 _5;
  by _IMPUTATION;
  MONOTONE reg(_2 _3 _4 _5 /details);
  MNAR MODEL(_2 _3 _4 _5 / modelobs= (TRTp='PLACEBO'));
RUN;
```

15.2.3 Example Code for MAR (Missing at Random)

```
** Full regression imputation of the monotone missing data sets
**;
PROC MI data=MI_MONOTONE NIMPUTE=1 OUT=MI_COMPLETE1 SEED=25140;
  CLASS TRTp ageclass severe;
  VAR ageclass severe base _2 _3 _4 _5;
  by _IMPUTATION_ TRTp;
  MONOTONE reg(_2 _3 _4 _5 /details);
RUN;
```

Docusign Envelope ID: 2C8E456A-1F31-40C9-8E58-ACC1B491CECD

15.3 Tipping Point Analysis

The tipping-point approach will be conducted to assess the robustness of the analysis with regards to the imputation of missing values. This approach will run MI under MNAR (missing not at random) assumption by searching for a tipping point that reverses a positive study conclusion. This imputation analysis used a specified sequence of shift parameters, which adjusts the imputed values for observations only in the active treatment group (Yang Yuan, 2014).

The following steps will be performed:

1. Achieve a monotone missing data pattern by imputing intermittent missing values by treatment group based on MAR assumption (SAS PROC MI with MCMC). This assumes a multivariate normal distribution over all variables included in the imputation model. The variables to be used in the imputation are treatment, baseline score and score values observed at all visits (i.e., Days 7, 14, 21, 28)

2. Generate 30 datasets based on step 1 (SAS PROC MI)

3. For each of these generated datasets, achieve imputation using MNAR including using a shift parameter δ only for imputed values from REL-1017 treatment group (SAS PROC MI with MNAR ADJUST)

4. Run the MMRM analysis on each of these imputed datasets

5. Combine the estimated coefficients from each dataset (using SAS PROC MIANALYZE) to get a combined estimate (mean and std) and produce the p-value for the combined estimates

6. Repeat steps 1-5 for different shift parameter δ values starting from 0 with increment value of 0.1 until statistical significance is lost